



Suivi automatique de nageurs à partir des séquences vidéo : application à l'analyse de performances

Djamel-Eddine Benarab

► To cite this version:

Djamel-Eddine Benarab. Suivi automatique de nageurs à partir des séquences vidéo : application à l'analyse de performances. Traitement du signal et de l'image [eess.SP]. Université Bretagne Loire, 2016. Français. NNT : . tel-01493439v2

HAL Id: tel-01493439

<https://hal.science/tel-01493439v2>

Submitted on 3 Apr 2017 (v2), last revised 23 May 2017 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



université de bretagne
occidentale

UNIVERSITE
BRETAGNE
LOIRE

THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE

sous le sceau de l'Université Bretagne Loire

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Mention : Traitement du signal et des images

École Doctorale SICMA

présentée par

Djamel-Eddine BENARAB

Préparée Dans l'équipe VISION de l'ISEN Brest
et la société ACTRIS

Suivi automatique de nageurs à partir des séquences vidéo : application à l'analyse de performances

Thèse soutenue le 02 décembre 2016

devant le jury composé de :

Guohai SITU

Professeur, Shanghai Institute of Optics and Fine Mechanics /
Rapporteur

Denis HAMAD

Professeur, Université de Littoral Côte d'Opale / *Rapporteur*

Christian BROSSEAU

Professeur, Université de Bretagne Occidentale / *Examineur*

Ayman ALFALOU

Professeur, ISEN Brest / *Directeur de la thèse*

Thibault NAPOLEON

Enseignant chercheur, ISEN Brest / *Encadrant de la thèse*

Philippe HELLARD

Directeur de recherche, FFN / *Co-encadrant*

Antoine VERNEY

Dirigeant, ACTRIS / *Co-encadrant*

Remerciements

Je souhaite remercier en premier lieu mon directeur de thèse, M. Ayman ALFA-LOU, Professeur et directeur du laboratoire Vision de L'ISEN Brest pour m'avoir choisi sur ce projet et pour m'avoir fait confiance tout au long de ces trois années. Je lui suis également reconnaissant pour le temps conséquent qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. J'ai pris un grand plaisir à travailler avec lui.

J'adresse de chaleureux remerciements à mon co-encadrant de thèse, M. Thibault NAPOLEON, enseignant chercheur à l'ISEN Brest, pour son attention de tout instant sur mes travaux, pour ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de cette thèse. Son énergie et sa confiance ont été des éléments moteurs pour moi. J'ai beaucoup appris à ses côtés et je lui adresse ma gratitude pour tout cela.

Je souhaite également remercier les membres du jury Messieurs, Situ GUOHAI, Denis HAMAD et Christian BROSSEAU, d'avoir accepté d'évaluer mon travail et de l'avoir enrichi par leurs remarques et leurs conseils.

Un grand merci à M. Philippe HELLARD et l'ensemble de l'équipe de recherche à la FFN, pour leur implication dans le projet, notamment sur la partie biomécanique et durant les réunions et les différentes compétitions nationales et internationales où nous avons échangé des idées intéressantes menant à l'enrichissement de ce travail.

Je tiens à remercier M. Antoine VERNEY, co-gérant de la société ACTRIS pour son implication dans le projet, sa disponibilité, sa bienveillance ainsi que pour le temps et l'effort consacrés pour la correction du manuscrit.

Je remercie également l'ensemble des personnels de L'ISEN Brest et particulièrement l'équipe VISION de m'avoir accueilli chaleureusement au sein de leur laboratoire. Plus particulièrement, je remercie mes deux camarades de bureau Dr. Philippe KATZ pour l'invention du mini-golf de bureau et Dr. Angel BLEUNVEN pour tout ce qui est fou, notamment les extrêmes Marathons et les baignades froides à l'océan atlantique.

Je tiens également à remercier mes amis (Mhammed, Yacine, Zinou, Jaâfar, Mossab, Hafidh, Helmi, Rafik, Hamza, Abdou...) pour leur encouragement, leur soutien inconditionnel et leur présence dans ma vie.

Enfin, Je dédie ce travail à MAMAN, MAMAN, MAMAN et PAPA... Quoi que je dise quoi que je fasse, je ne pourrai jamais exprimer ma gratitude, mon admiration et l'amour qui inonde mon cœur... Yemma, Baba je vous aime... J'espère pouvoir vous rendre fiers et montrer le bon exemple à mes deux frères que j'adore Badie et Akram.

Résumé

Dans le but d'améliorer les performances des nageurs professionnels, nous avons développé, en collaboration avec la Fédération Française de Natation, un système automatique de suivi à base des séquences vidéo. Pour ce faire, nous proposons un nouveau système de prise de vue $8K$ adapté au milieu aquatique et permettant un calibrage du bassin. Celui-ci établit le lien entre les coordonnées pixels et métriques permettant, entre autres, d'extraire le couloir concerné pour effectuer les différents traitements.

Afin d'initialiser le suivi, il est nécessaire de localiser le nageur. Pour cela, nous proposons d'utiliser une approche *a contrario* pour détecter le mouvement, puis l'approche *Scaled Composite JTC* pour localiser précisément la tête du nageur. Ensuite, nous implémentons et adaptons les techniques de suivi de la littérature, notamment celles basées sur la corrélation NL-JTC, les histogrammes de couleur, les motifs binaires locaux (LBP) et les histogrammes de gradient orienté (HOG). Suite aux différentes limitations de ces techniques, nous proposons de nouvelles approches optimisées basées principalement sur la fusion de données. Tout d'abord, nous développons l'approche multipiste constituée de plusieurs pistes de suivi, où chacune représente l'une des techniques de suivi citées précédemment. Ensuite, un choix basé sur l'histogramme de couleur est effectué afin de choisir la meilleure décision parmi celles offertes par chaque piste. Cette approche a significativement amélioré les résultats mais ceux-ci restent insuffisants pour l'étude des performances. Dans ce sens, nous proposons une nouvelle approche par fusion dynamique qui consiste à fusionner le plan de corrélation NL-JTC et le plan de scores couleurs dans le but d'extraire une description plus riche de la cible (forme + couleur). Cette approche a montré de très bons résultats dans le cas où la cible à suivre est visible mais reste très sensible aux occultations de celle-ci. Afin de résoudre cette difficulté, nous améliorons l'approche proposée en suivant simultanément la tête et le maillot de bain du nageur. Cette approche multizone permet, grâce à un critère de décision complexe, de retrouver la zone occultée à l'aide de la zone visible.

Enfin, une étude de performances a été menée et les résultats obtenus ont permis de valider ce système. En particulier, nous nous sommes intéressés aux mesures de vitesse cyclique, intra-cyclique et instantanée, afin d'étudier et améliorer les performances des nageurs.

Abstract

In order to improve the performance of professional swimmers, we have developed, in collaboration with the FFN (French Swimming Federation), an automatic tracking approach using video sequences. To do this, we propose a new *8K* shooting system adapted to the aquatic environment and allowing a pool calibration. This establishes the link between pixel and metric coordinates, which allows among others, to extract the concerned lane to carry out the different treatments.

In order to initialize the tracking, it is necessary to locate the swimmer. For this, we propose to use an *a contrario* approach to detect movement, then the *Scaled Composites JTC* approach to precisely locate the swimmer's head. Afterwards, we implement and adapt several tracking techniques well-known in the literature, namely those based on the NL-JTC correlation, color histograms, Local Binary Patterns (LBP) and histograms of oriented gradient (HOG). Given the various limitations of these techniques, we propose new optimized approaches based primarily on data fusion. First, we develop a multitasking approach consists of several tracks, where each track represents one of the tracking techniques mentioned above. Then, a choice based on the color histogram is made to select the best decision among those offered by each track. This approach has significantly improved the results, but they remain insufficient for the performance analysis. Therefore, we propose a new dynamic fusion approach that combines NL-JTC correlation plane and color scores plane in order to generate a richer description of the target (form + color). This approach has shown very good results in the case where the target is visible but it is still sensitive to occlusions. To solve this problem, we improve this proposed approach by tracking simultaneously the head and the swimsuit of the athlete. This multi related targets approach enables, through a complex decision criterion, to find the the occluded zone based on the visible one.

Finally, a performance study is conducted and the results have validated the system. In particular, we were interested in cyclical, intra-cyclical and instantaneous speed measurements, to study and improve the swimmers performance.

Table des matières

Remerciements	iii
Résumé	v
Abstract	vii
Introduction	1
1 Etat de l’art des techniques de suivi	5
1.1 Introduction	6
1.2 Processus général de suivi	7
1.3 Calibrage	8
1.3.1 Modèle en trou d’épingle	8
1.3.2 DLT (<i>Direct Linear Transformation</i>)	9
1.4 Prétraitement	10
1.4.1 Soustraction de fond	11
1.4.2 Détection de peau	14
1.4.3 Détection de lignes	17
1.5 Suivi par corrélation	20
1.5.1 Cl-JTC (Joint Transform Correlator Classique)	21

TABLE DES MATIÈRES

1.5.2	<i>NZ-JTC</i> (JTC Non-Zéro)	23
1.5.3	<i>B-JTC</i> (JTC Binaire)	23
1.5.4	<i>FA-JTC</i> (<i>Fringe-adjusted JTC</i>)	24
1.5.5	NL-JTC (JTC Non-Linéaire)	25
1.5.6	Critère de décision basé sur le <i>PCE</i>	25
1.5.7	Discussion	26
1.6	Suivi par des techniques basées sur les histogrammes	27
1.6.1	Histogramme de couleurs	27
1.6.2	Spatigramme	28
1.6.3	HOG	28
1.6.4	LBP	28
1.6.5	Discussion	30
1.7	Fusion de descripteurs	30
1.7.1	Pourquoi la fusion	30
1.7.2	Niveaux de fusion	31
1.7.3	Normalisation	32
1.7.4	Types de fusion	33
1.8	Prédiction de mouvements	34
1.8.1	Filtre de Kalman	34
1.8.2	Filtre particulaire	35
1.8.3	Mean-shift	35
1.8.4	Discussion	36
1.9	Systèmes de suivi	36
1.9.1	Suivi de personnes	37

1.9.2	Suivi de visages	38
1.9.3	Suivi de véhicules	39
1.9.4	Discussion	40
1.10	Applications relatives à la natation	41
1.10.1	Analyse des performances des nageurs à partir de vidéos	41
1.10.2	Projet Poséidon pour la détection de noyade	41
1.11	Conclusion	43
2	Prétraitements adaptés aux vidéos du milieu aquatique : Analyse, choix et validation	45
2.1	Introduction	46
2.2	Spécification du milieu	46
2.3	Segmentation du bassin	48
2.3.1	Etudes basées sur les espaces couleur	48
2.3.2	Etude basée sur la détection des lignes	52
2.3.3	Discussion	53
2.4	Calibrage	53
2.4.1	Objectifs du calibrage	54
2.4.2	DLT pour le calibrage du bassin	55
2.4.3	Correction de perspective	56
2.5	Localisation de la région d'intérêt	56
2.5.1	Extraction du couloir	57
2.5.2	Différence d'images	59
2.5.3	Décomposition en blocs	60
2.5.4	Seuillage et classification	61

TABLE DES MATIÈRES

2.5.5	Elimination des faux blocs	61
2.5.6	Détection de mouvement du nageur	62
2.6	Initialisation du suivi par l'approche <i>Scaled Composite JTC</i>	63
2.6.1	Génération de la base de données	63
2.6.2	Application de la technique NL-JTC	65
2.6.3	Référence composite re-dimensionnée	66
2.6.4	Pré-localisation de la région d'intérêt	67
2.6.5	Décision basée sur PCE	68
2.7	Conclusion	69
3	Système de suivi de nageurs : Analyse, choix et validation	71
3.1	Introduction	71
3.2	Processus général de suivi de nageurs	72
3.3	Suivi par corrélation	72
3.3.1	NL-JTC pour le suivi des nageurs	73
3.3.2	Choix du coefficient de non-linéarité	73
3.3.3	Limitations de NL-JTC pour le suivi	74
3.3.4	Etude et optimisation du plan de corrélation	77
3.4	Suivi par des techniques basées sur les histogrammes	83
3.4.1	Processus général	83
3.4.2	Histogramme de couleur pour le suivi des nageurs	84
3.4.3	LBP pour le suivi des nageurs	87
3.4.4	HOG pour le suivi des nageurs	89
3.4.5	Avantages et limitations des techniques basées sur les histogrammes	89

3.5	Conclusion	91
4	Système de suivi de nageurs : Optimisation	93
4.1	Introduction	94
4.2	Prédiction de la trajectoire	95
4.2.1	Localisation du couloir de nage	95
4.2.2	Direction de nage	96
4.2.3	Vitesse moyenne du nageur	96
4.2.4	Position référentielle validée	97
4.2.5	Discussion	97
4.3	Approche de suivi multipiste	98
4.3.1	Principe du suivi multipiste	98
4.3.2	Approche de suivi multipiste uni-descripteur	99
4.3.3	Approche de suivi multipiste multidescripteur	100
4.3.4	Discussion	101
4.4	Approche de suivi par fusion dynamique	102
4.4.1	Inspiré du principe de débruitage	102
4.4.2	Principe de la fusion dynamique	103
4.4.3	Génération des plans de scores	103
4.4.4	Normalisation des scores	104
4.4.5	Fusion des plans	105
4.4.6	Analyse du plan de fusion : précision et robustesse	105
4.4.7	Facteur de confiance	106
4.4.8	Expérimentation sur des images synthétiques	106

TABLE DES MATIÈRES

4.4.9	Discussion	109
4.5	Approche de suivi multizone	109
4.5.1	Inspiré par l'idée du multipiste	110
4.5.2	Processus du suivi multizone	111
4.5.3	Détections potentielles	112
4.5.4	Critère de décision complexe	112
4.5.5	Discussion	117
4.6	Conclusion	118
5	Expérimentations et résultats	119
5.1	Introduction	119
5.2	Protocole expérimental	120
5.2.1	Choix du système de prise de vue	120
5.2.2	Système de prise de vue $8K$	125
5.2.3	Base de données	126
5.3	Résultats de suivi	127
5.3.1	Validation de la vérité terrain	128
5.3.2	Evaluation des prétraitements	133
5.3.3	Evaluation de l'approche de suivi multipiste	136
5.3.4	Evaluation de l'approche de suivi à base de la fusion dynamique	140
5.3.5	Evaluation de l'approche de suivi multizone	144
5.3.6	Discussion	146
5.4	Etude de performances de nageurs	147
5.4.1	Introduction à l'étude de performances de nageurs	147

5.4.2	Vitesse cyclique et intra-cyclique	147
5.4.3	Vitesse instantanée	150
5.4.4	Filtrage des courbes de vitesse instantanée	151
5.4.5	Exemples d'application réelle	153
5.5	Conclusion	156
Conclusion		159
A Prototype d'évaluation des nageurs basé sur l'annotation manuelle		163
A.1	Introduction	163
A.2	Présentation de l'interface	163
A.3	Calibrage	165
A.4	Annotation	165
A.5	Analyse	166
B Prototype d'évaluation des nageurs basé sur le suivi automatique		169
B.1	Introduction	169
B.2	Acquisition (<i>Video controller</i>)	169
B.3	Suivi automatique (<i>Tracking</i>)	170
B.4	Statistiques (<i>Statistics</i>)	170
B.5	Performances (<i>Evaluation</i>)	170
Publications		183

Introduction

Contexte

Les applications de suivi des personnes deviennent de plus en plus nombreuses grâce au grand intérêt donné à la recherche et au développement dans ce domaine. Ces systèmes diffèrent selon leurs objectifs, à savoir, la sûreté des lieux publics, le domaine militaire, médical ou encore le sport de haut niveau. Récemment, les recherches se sont orientées vers les milieux sportifs, notamment la natation. En effet, un grand intérêt est donné à l'analyse et à l'amélioration des performances des nageurs amateurs et de hauts-niveaux.

Dans ce cadre, plusieurs études cinématiques (études des mouvements indépendamment des causes qui les produisent) et statistiques ont été faites afin d'évaluer les performances des nageurs durant les entraînements ainsi que dans les compétitions. Les systèmes existants évaluent les nageurs en termes de temps de passages, plongeon, reprise de nage, retournement, cycle de respiration, etc. Ces études permettent aux entraîneurs d'accompagner leurs nageurs en analysant leurs stratégies de nage afin d'améliorer leurs performances.

Parmi ces systèmes, on retrouve les chronomètres qui permettent de mesurer la vitesse moyenne et les temps de passages. Ils sont principalement utilisés pour des évaluations rapides qui ne nécessitent qu'une faible précision. D'autres systèmes se basent sur l'annotation manuelle de la position du nageur dans chaque image de la vidéo pour estimer les différentes mesures. Ceux-ci permettent une analyse précise sans contraintes imposées aux nageurs, mais sont chronophage et nécessitent d'importantes ressources humaines pour l'annotation de chaque image des vidéos de compétitions. D'autres systèmes acquièrent les données à partir de capteurs portés par les nageurs. Ceci permet l'obtention d'analyses d'une très grande précision. En revanche, ces capteurs ne sont pas autorisés dans les compétitions et limitent le mouvement des nageurs dans le cas des entraînements. De ce fait, les systèmes actuels de mesures des performances des nageurs restent parfois insuffisants malgré leur besoin croissant.

Objectifs

Afin de pallier les problèmes des systèmes existants, nous menons des recherches dans le cadre d'un contrat de thèse CIFRE. Celle-ci se fait en collaboration entre le laboratoire VISION de l'ISEN Brest spécialisé en traitement d'image numérique et optique, l'entreprise ACTRIS Brest spécialisée en acquisition et traitement de données et la Fédération Française de Natation (FFN). Ce projet s'inscrit dans le cadre des études biomécaniques menées par la FFN permettant de développer une nouvelle solution pour l'étude des performances des nageurs. En effet, pour aboutir à cet objectif, l'étape la plus importante consiste à localiser et à suivre des nageurs tout au long d'une course. Par conséquent, notre objectif principal est de développer un système de suivi automatique de nageurs qui répond aux contraintes suivantes :

- Système automatique (intervention minimale de l'utilisateur)
- Sans marqueur physique
- Sans capteur porté par le nageur
- Facile à mettre en place
- Adapté aux différentes configurations des bassins

Afin de répondre à ces contraintes, nous proposons une nouvelle solution basée sur la vision par ordinateur. Cette solution est basée sur la mise en place d'un système de prise de vue adapté pour filmer les compétitions. Ensuite, nous proposons un système de suivi automatique de nageurs à partir des ces séquences vidéo prises pendant les entraînements ou les compétitions.

Contributions

Tout d'abord, nous avons commencé par une analyse approfondie des séquences vidéos des compétitions de natation mises à notre disposition par la FFN. En effet, nous nous sommes principalement intéressés aux spécificités du milieu aquatique, notamment les dimensions des bassins, les couloirs de nage, les couleurs des différents éléments, etc. Ceci a mené à proposer des prétraitements adaptés afin de calibrer le bassin, extraire le couloir, restreindre la région d'intérêt et initialiser le suivi de la tête du nageur [1, 2].

Ensuite, après avoir fait une étude de l'état de l'art des systèmes de suivi, nous avons choisi les techniques les plus adéquates afin de les adapter, par la suite, à notre sujet de suivi des nageurs. Nous parlons ici des techniques suivantes : NL-JTC (*Non-linear Joint Transform Correlator*), Histogramme de couleurs, LBP (*Local Binary Patterns*) et HOG (*Histogram of Oriented Gradient*).

Partant du principe que chacune de ces techniques a des avantages et des limitations, et dans le but d'optimiser le suivi, nous avons proposé un système multipiste de suivi de

nageurs [3] où chaque piste consiste en un processus de suivi indépendant basé sur une des techniques citées précédemment. La décision est prise après quelques itérations et la meilleure piste est choisie selon un critère de ressemblance basé sur l'histogramme de couleurs et l'historique des détections. En d'autres termes, l'approche multipiste peut être considérée comme une fusion effectuée au niveau des décisions. Cette approche a significativement optimisé les résultats de suivi. Cependant, le critère de décision utilisé n'est pas optimal et favorise l'information issue de la couleur.

Dans le but d'améliorer ce critère et pour mieux fusionner les données, nous avons proposé de remonter dans les niveaux de fusion en effectuant celle-ci au niveau des plans de scores. Pour cela, nous avons choisi de fusionner les techniques NL-JTC et l'histogramme de couleur, après avoir remarqué une certaine complémentarité entre ces méthodes. En effet, la technique NL-JTC a montré de bons résultats en terme de précision pour la localisation, mais reste sensible à la rotation et à la déformation des contours. D'autre part, les approches basées sur l'histogramme de couleur ont montré une certaine robustesse pour le suivi des nageurs, mis à part dans les cas où une éventuelle confusion de couleur peut être faite avec d'autres objets de la scène, malgré certaines imprécisions de localisation. Ceci nous a mené à proposer une nouvelle approche de suivi appelée "fusion dynamique" [4]. Cette approche a nettement amélioré les résultats du suivi, particulièrement dans le cas où la cible est entièrement ou partiellement visible. En revanche, dans le cas de l'occultation quasi-totale de la cible, le suivi reste défaillant.

Afin de Pallier ce problème, nous avons proposé une approche "multizone" (*multi related targets*) [5] qui permet de suivre deux cibles en parallèle (la tête et le maillot de bain) en utilisant l'approche de suivi par fusion dynamique [4]. Cette approche multizone permet, entre autre, de prendre en compte la distance entre les deux cibles ainsi que la vitesse moyenne de nage afin d'optimiser le suivi. En effet, cette approche optimise principalement deux points, la précision du suivi et la possibilité de retrouver la cible entièrement occultée en faisant référence à l'autre cible lorsque celle-ci est visible. Dans le même contexte, nous proposons de calibrer le bassin afin de faire une prédiction optimisée de la trajectoire de nage, assurer le passage entre les coordonnées métriques réelles dans le bassin et les coordonnées pixels dans l'image et calculer des mesures permettant l'estimation de la vitesse instantanée du nageur afin d'étudier ses performances.

Organisation du manuscrit

Cette thèse comporte sept chapitres :

- Introduction
- Chapitre 1 : Etat de l'art des systèmes de suivi

- Chapitre 2 : Prétraitements adaptés aux vidéos du milieu aquatique - analyse, choix et validation
- Chapitre 3 : Système de suivi de nageurs - analyse, choix et validation
- Chapitre 4 : Système de suivi de nageurs - optimisation
- Chapitre 5 : Expérimentations et résultats
- Conclusion

Tout d'abord, nous commençons par une étude approfondie de l'état de l'art des systèmes de suivi d'objets dans le **Chapitre 1**. Nous abordons cette partie par une présentation générale des différents modules constituant un système de suivi. Puis, nous présentons des applications réelles de suivi rassemblant les différents modules afin d'illustrer différents enchaînements possibles de ces blocs. Nous finissons cette partie par la présentation d'applications de suivi appliqué au milieu aquatique et à la natation.

Dans le **Chapitre 2**, nous proposons des prétraitements adaptés à la préparation des vidéos et à l'initialisation de notre système de suivi. Pour cela, nous étudions les vidéos des courses selon les spécificités du milieu aquatique (dimensions, lignes, couleurs, etc) afin de permettre le calibrage du bassin, l'extraction du couloir et la restriction de la région d'intérêt autour du nageur concerné. Ceci permet de faciliter le suivi mais aussi de l'initialiser en détectant automatiquement la référence initiale de la tête du nageur en se basant sur sa forme définie dans une base d'apprentissage pré-générée.

Nous faisons, dans le **Chapitre 3**, des choix sur les techniques présentées dans l'état de l'art et nous proposons un premier système de suivi optimisé et adapté au suivi des nageurs. Ensuite, nous concluons ce chapitre par une discussion sur les avantages et les limitations de chaque approche.

Partant de la conclusion précédente, nous proposons dans le **Chapitre 4** de nouvelles approches de suivi basées principalement sur la fusion multi-descripteurs nous permettant d'aboutir à une approche robuste de suivi de la tête du nageur. Finalement, nous proposons une approche multizone qui consiste à suivre la tête et le maillot de bain en même temps afin de surmonter les problèmes d'occultation d'une partie de la cible à suivre.

Enfin, dans le **Chapitre 5**, nous présentons le protocole expérimental et le choix du système de prise de vue et nous définissons la base de données utilisée pour les tests. Nous validons ensuite les choix techniques faits dans les chapitres précédents par des tests présentés dans le même ordre que le manuscrit. Enfin, nous finissons par une étude biomécanique sur la vitesse instantanée du nageur calculée en utilisant une centrale inertielle, une annotation manuelle et notre approche de suivi automatique.

Chapitre 1

Etat de l’art des techniques de suivi

Sommaire

1.1	Introduction	6
1.2	Processus général de suivi	7
1.3	Calibrage	8
1.3.1	Modèle en trou d’épingle	8
1.3.2	DLT (<i>Direct Linear Transformation</i>)	9
1.4	Prétraitement	10
1.4.1	Soustraction de fond	11
1.4.2	Détection de peau	14
1.4.3	Détection de lignes	17
1.5	Suivi par corrélation	20
1.5.1	Cl-JTC (Joint Transform Correlator Classique)	21
1.5.2	NZ-JTC (JTC Non-Zéro)	23
1.5.3	B-JTC (JTC Binaire)	23
1.5.4	FA-JTC (<i>Fringe-adjusted JTC</i>)	24
1.5.5	NL-JTC (JTC Non-Linéaire)	25
1.5.6	Critère de décision basé sur le <i>PCE</i>	25
1.5.7	Discussion	26
1.6	Suivi par des techniques basées sur les histogrammes	27
1.6.1	Histogramme de couleurs	27
1.6.2	Spatiogramme	28
1.6.3	HOG	28
1.6.4	LBP	28
1.6.5	Discussion	30
1.7	Fusion de descripteurs	30
1.7.1	Pourquoi la fusion	30

1.7.2	Niveaux de fusion	31
1.7.3	Normalisation	32
1.7.4	Types de fusion	33
1.8	Prédiction de mouvements	34
1.8.1	Filtre de Kalman	34
1.8.2	Filtre particulaire	35
1.8.3	Mean-shift	35
1.8.4	Discussion	36
1.9	Systèmes de suivi	36
1.9.1	Suivi de personnes	37
1.9.2	Suivi de visages	38
1.9.3	Suivi de véhicules	39
1.9.4	Discussion	40
1.10	Applications relatives à la natation	41
1.10.1	Analyse des performances des nageurs à partir de vidéos . . .	41
1.10.2	Projet Poséidon pour la détection de noyade	41
1.11	Conclusion	43

1.1 Introduction

Dans le contexte de notre collaboration avec la FFN (Fédération Française de Natation), nous sommes amenés à développer un système de suivi automatique des nageurs à partir de séquences vidéo dans le but d'estimer leurs performances. Pour ce faire, nous sommes confrontés à des contraintes liées aux spécificités des compétitions de natation et du milieu aquatique. Ces difficultés surviennent le plus souvent en raison de :

- Positionnement variable de la caméra dans les gradins.
- Mouvement de la caméra et des différents éléments dans le bassin (eau, lignes de couloirs, nageur, etc).
- Déformation structurelle et changement d'apparition des nageurs.
- Occlusion partielle ou totale du nageur durant la séquence.
- Changement d'éclairage.

Dans le but d'aider à résoudre ces difficultés, certaines spécificités de l'environnement aquatique peuvent être utilisées. En particulier les dimensions du bassin et des couloirs, la couleur des différents éléments dans la scène (eau, lignes de couloirs, peau, etc.) ou encore le profil de nage peuvent apporter une aide précieuse pour le suivi des nageurs.

Dans ce chapitre, nous présentons un aperçu général de l'état de l'art des systèmes de suivi. Pour cela, nous commençons par la présentation du processus général de suivi, puis, nous détaillons chaque étape en tenant en compte des différentes contraintes imposées ainsi que des spécificités liées à la natation et à l'environnement aquatique.

Ensuite, nous présentons quelques systèmes de suivi existants afin de montrer l'enchaînement et le fonctionnement des différents modules. Enfin, quelques travaux en rapport étroit avec le sujet de cette thèse sont présentés, à savoir, le suivi des nageurs et l'évaluation de leurs performances.

1.2 Processus général de suivi

Le suivi d'un objet, ou *object tracking* [6–11] d'un point de vue de la vision par ordinateur, consiste à localiser et suivre une cible dans chaque image d'une séquence vidéo. Afin de réaliser ces deux étapes (localisation et suivi), des caractéristiques de la cible doivent être extraites, dans le domaine spatial ou fréquentiel. Ces caractéristiques doivent permettre de différencier la cible de l'arrière-plan. Contrairement aux méthodes de reconnaissance qui se font sur des images fixes, le suivi s'effectue sur des séquences ou des flux vidéo, parfois en temps réel. Ceci apporte une information temporelle qui permet de prédire le mouvement de la cible dans les images suivantes de la séquence. Par conséquent, il est ainsi possible de restreindre la zone de recherche de la cible afin de faciliter le suivi.

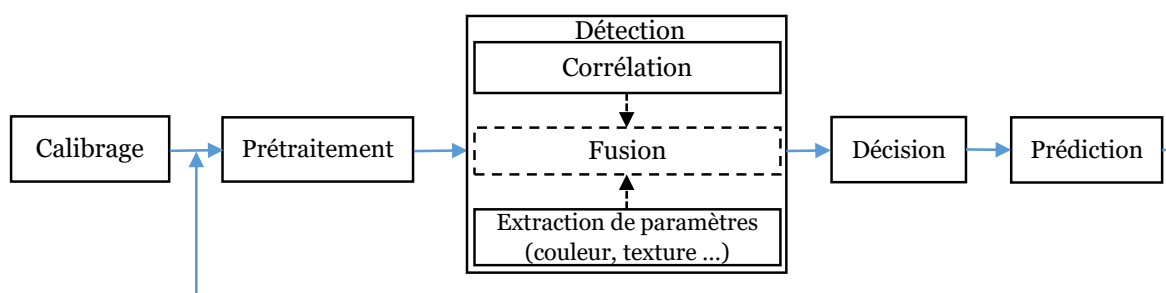


FIGURE 1.1 – Diagramme illustratif du processus général de suivi.

Comme le montre la figure 1.1, les systèmes de suivi se composent généralement des étapes suivantes : calibrage, prétraitement, détection et prédiction. La première étape de calibrage est nécessaire pour corriger les perspectives de la caméra ainsi que pour faire des mesures dans l'image qui permettent l'optimisation du suivi. Les prétraitements, quant à eux, peuvent être appliqués aux données brutes (séquences d'images) pour faciliter et améliorer le suivi. Ainsi, les images prétraitées pourront être analysées pour détecter et localiser la cible.

Pour cela, la littérature propose des méthodes basées sur la corrélation optique [7, 12–15] permettant de détecter la cible en analysant le contour de l'image. Il existe aussi des méthodes numériques qui décrivent d'autres paramètres comme : la couleur [16–18], la texture [19, 20] et le gradient [21]. Nous trouvons également des méthodes

basées sur la fusion de plusieurs descripteurs qui permettent d'enrichir la description de l'objet à suivre [22, 23]. Une fois ces paramètres extraits, des comparaisons sont effectuées entre l'image de référence de l'objet à suivre et la scène, dans le but de prendre une décision et de choisir la meilleure cible. Enfin, une prédiction de la prochaine position de la cible peut être appliquée afin de restreindre la zone de recherche et optimiser le suivi. Par la suite, nous allons détailler les différents modules du système de suivi.

1.3 Calibrage

Afin de concevoir un système de suivi robuste et précis à partir d'une séquence vidéo, il est nécessaire de surmonter les difficultés liées aux caractéristiques optiques de la caméra ainsi qu'à son mouvement et à son positionnement. En effet, les images acquises subissent souvent des déformations causées par les lentilles de la caméra. D'autre part, le mouvement brusque de la caméra peut poser des difficultés pour la localisation des objets dans la scène. Enfin, la position variable de la caméra peut engendrer des changements d'échelle et de l'objet à suivre. Ces différents problèmes peuvent cependant être résolus dans le cas où la relation entre les coordonnées métriques réelles de l'objet et ses coordonnées pixels dans l'image est connue.

Pour résoudre ces difficultés, nous proposons de nous inspirer des travaux réalisés par la FFN dans le cadre de la thèse de Marc Elipot intitulée : "Application des nouvelles techniques de calibration de caméra à l'analyse cinématique en natation et identification des coordinations motrices lors des phases sous-marines de départs chez des nageurs de haut niveau" [24]. Dans la suite nous présentons le principe de calibrage en commençant par introduire le modèle en trou d'épingle, qui permet d'expliquer le fonctionnement de la caméra, puis nous présenterons la technique DLT (*Direct Linear Transformation*) qui permet de calibrer la caméra [25].

1.3.1 Modèle en trou d'épingle

Le modèle en trou d'épingle proposé par Zhang [26] décrit la relation mathématique entre les coordonnées d'un point $3D$ et sa projection sur le plan de l'image d'une caméra. Dans ce modèle basique, l'ouverture de la caméra est représentée par un point, il est à noter que ce modèle ne contient pas de lentilles pour la focalisation de la lumière. Par conséquent, ce modèle n'illustre pas les effets correspondant aux lentilles, notamment, les distorsions géométriques et le flou. Cela signifie que le modèle en trou d'épingle ne peut être utilisé que comme une approximation de premier ordre de la mise en correspondance d'une scène $3D$ en une image $2D$.

La mise en correspondance d'une scène 3D dans le plan $Oxyz$, en une image 2D dans un plan $Óuv$, consiste en une projection de perspective suivie d'une rotation de 180° dans le plan image. La taille relative de l'image résultante des objets projetés dépend de leur distance par rapport au point focal. La taille globale de l'image, quant à elle, est liée à la distance f entre le plan image et le point focal comme le montre la figure 1.2. Ce modèle permet de schématiser le fonctionnement de la caméra et de calculer la relation entre les coordonnées métriques dans le plan réel et leur projection dans le plan d'image, comme nous allons l'expliquer dans la section suivante.

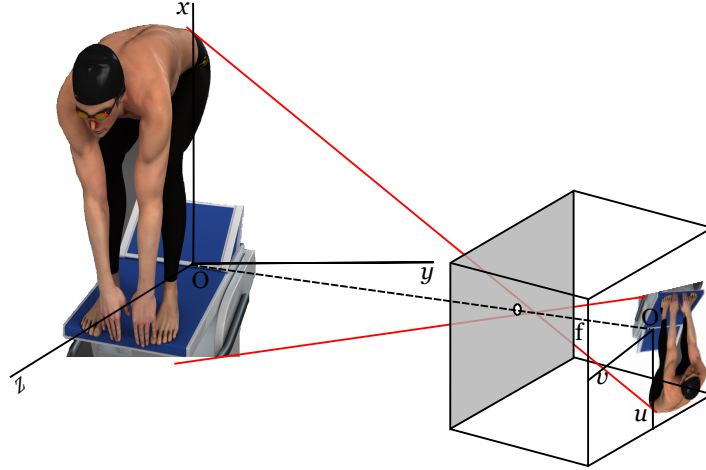


FIGURE 1.2 – Principe de fonctionnement de la caméra avec un modèle de base en trou d'épingle. La lumière issue de la scène 3D passe par le trou d'épingle et projette une image inversée sur le côté opposé de la boîte simulant le capteur d'acquisition.

1.3.2 DLT (*Direct Linear Transformation*)

Afin d'assurer le passage entre les coordonnées métriques réelles et les coordonnées pixels, Abd-El-Aziz et Karara ont proposé la technique de calibrage DLT [25]. Cette technique consiste à calculer les paramètres de calibrage résolvant l'équation suivante :

$$\begin{cases} u = \frac{L_1x + L_2y + L_3z + L_4}{L_9x + L_{10}y + L_{11}z + 1} \\ v = \frac{L_5x + L_6y + L_7z + L_8}{L_9x + L_{10}y + L_{11}z + 1} \end{cases} \quad (1.1)$$

Avec (u, v) représentent les coordonnées pixels d'un point dans l'image, (x, y, z) représentent les coordonnées métriques réelles du même point, comme le montre la figure 1.2. $L_{1..11}$ sont les paramètres de calibrage à calculer.

Cette technique peut être simplifiée dans le cas du calibrage d'un plan 2D dans une image. Dans ce cas, l'équation 1.1 devient :

$$\begin{cases} u = \frac{L_1x + L_2y + L_3}{L_7x + L_8y + 1} \\ v = \frac{L_4x + L_5y + L_6}{L_7x + L_9y + 1} \end{cases} \quad (1.2)$$

Dans ce cas, le calcul des 8 paramètres de calibration $L_{1..8}$ nécessite la connaissance des coordonnées pixels et métriques de quatre points de la scène filmée. En remplaçant (u, v) et (x, y, z) de l'équation 2.2 par ces valeurs, nous obtiendrons un système de 8 équations qui permet de calculer les inconnus $L_{1..8}$.

1.4 Prétraitement

Dans le domaine du traitement d'image, les prétraitements diffèrent d'un système à un autre. En effet, ils dépendent de l'environnement de travail, de la qualité d'image et des objectifs visés par le système. Dans notre cas, le suivi des nageurs s'effectue dans un environnement spécifique et non-contrôlé à cause des différentes difficultés rencontrées comme les éclaboussures, l'éclairage, le mouvement de l'eau, du nageur et de la caméra, etc. Ceci nous a amené à étudier plusieurs techniques de prétraitements afin de faciliter la tâche de localisation des nageurs dans les images successives d'une vidéo.

Dans un premier temps, il est important d'améliorer la qualité de l'image et de la préparer pour la prochaine étape de détection. Pour cela, des techniques classiques de filtrage pourront être appliquées, notamment le filtre Sobel, Median, gaussien, etc [15, 27]. D'autre part, nous allons nous concentrer sur des prétraitements adaptables à l'environnement de la natation. En effet, la piscine peut être considérée comme un environnement homogène qui contient une forte densité de couleur bleu. Pour cette raison, la soustraction de fond [28–32] semble intéressante afin de restreindre la zone de recherche. D'un autre côté, la peau des nageurs contient une forte composante rouge ce qui donne la possibilité d'isoler la peau pour localiser plus précisément le nageur [33–38]. D'autre part, étant donné que les déplacements du nageur sont limités à son couloir, il est possible de prédire sa trajectoire grâce à une localisation des lignes d'eau par des méthodes de détection de lignes [39–42]. La combinaison de ces différentes techniques de prétraitement peut mener à faciliter la localisation des nageurs.

1.4.1 Soustraction de fond

La soustraction de fond est une technique connue en traitement d'image [28–31]. Elle consiste à séparer le premier plan, qui contient souvent la région d'intérêt, de l'arrière-plan (le fond). L'objectif est d'extraire le premier plan pour des traitements ultérieurs : détection, localisation ou reconnaissance d'objet. En d'autres termes, la soustraction de fond est largement utilisée pour détecter des objets en mouvement dans des séquences vidéo. La détection de l'objet en mouvement se fait en effectuant la soustraction d'une image de référence de l'image courante. L'image de référence est souvent appelée image de fond ou modèle de base qu'il faut déterminer. Dans ce qui suit nous allons exposer quelques méthodes de soustraction de fond.

1.4.1.1 Filtre médian temporel

Le modèle de filtre médian temporel [28,31] consiste à considérer la valeur médiane M entre les dernières images comme arrière-plan, comme le montre l'équation suivante :

$$\begin{cases} Si & |I_t - M| < S & I_t \in \text{arrière plan} \\ Sinon & I_t \in \text{premier plan} \end{cases} \quad (1.3)$$

Avec, I_t représente un pixel dans l'image à l'instant t , M est la valeur médiane des pixels I entre l'instant 0 et t et S représente un seuil empirique. Le modèle de filtre médian temporel ne se base pas sur une description statistique rigoureuse. Ceci est expliqué par un manque d'information concernant la variance des intensités des pixels. Cette information est nécessaire pour la détermination automatique d'un seuil pertinent. Pour cette raison, la modélisation par filtre médian temporel n'est pas adaptée à notre cas de suivi de nageur et nous allons investiguer, par la suite, d'autres techniques qui prennent en compte cette information.

1.4.1.2 Eigen backgrounds

Parmi les méthodes proposées pour la modélisation de fond, on trouve la décomposition des images en valeurs propres. Cette méthode se fait en deux étapes : apprentissage et classification [31].

Apprentissage : Une matrice moyenne μ et une matrice de covariance σ sont calculées sur quelques images de n pixels. Ensuite une matrice ϕ de taille $m * p$, dite matrice de l'espace propre, constituée des m vecteurs propres est calculée. Cette matrice servira de modèle pour représenter les parties immobiles de la scène.

Classification : les pixels de chaque image sont projetés sur l'espace propre comme le

montre l'équation suivante :

$$I'_t = \phi(I_t - \mu) \quad (1.4)$$

Afin d'extraire les pixels correspondant aux objets en mouvement dans une image, on les reprojette sur l'espace de l'image comme le montre l'équation suivante :

$$I''_t = \phi^t I'_t + \mu \quad (1.5)$$

En faisant la comparaison entre le pixel I''_t et le pixel I_t les objets en mouvement peuvent être extraits. Cette méthode est efficace dans le cas d'un arrière-plan statique. Cependant, elle n'est pas adaptée à notre cas car l'eau, qui représente notre arrière-plan, est toujours en mouvement.

1.4.1.3 Modèle gaussien

Pour chaque pixel I_t , une moyenne cumulative μ_t et une variance σ_t^2 sont calculées sur l'ensemble des images de la séquence vidéo entre l'instant 0 et t afin d'obtenir un modèle statistique adapté. Ce dernier suit la loi gaussienne (normale) $N(I_t | \mu_t, \sigma_{i,t})$ et il représente l'évolution des valeurs de chaque pixel I_t au cours de la scène [31,32], comme le montre les équations suivantes :

$$\mu_t = \alpha I_t + (1 - \alpha) \mu_{t-1} \quad (1.6)$$

$$\sigma_t^2 = \alpha d^2 + (1 - \alpha) \sigma_{t-1}^2 \quad (1.7)$$

Avec α un facteur de pondération qui permet de mettre à jour les moyennes μ_t et les écarts-types σ_t . Par conséquent, un modèle global de l'arrière-plan de la scène peut être généré en regroupant les moyennes μ_t de tous les pixels dans une matrice. Dans chaque image de la scène, si la différence d calculée selon l'équation 1.8 entre la valeur d'un pixel I_t et la moyenne μ_t est supérieure de k fois son écart type σ_t , ce pixel sera considéré comme premier plan (cible à suivre), sinon il sera considéré comme faisant partie du fond, comme le montre l'équation 1.9.

$$d = |I_t - \mu_t| \quad (1.8)$$

$$\begin{cases} \text{Si } d < k\sigma_t & I_t \in \text{arrière plan} \\ \text{Sinon} & I_t \in \text{premier plan} \end{cases} \quad (1.9)$$

La modélisation par une gaussienne est une méthode pertinente dans le cas d'arrière plan n'ayant qu'une faible variance. En revanche, pour les cas plus complexes, l'eau qui bouge dans la piscine par exemple, cette approche ne propose pas un modèle adéquat et ne peut donc pas être utilisé pour le cas du suivi de nageur..

1.4.1.4 Modélisation par mélange de gaussiennes

Pour pallier le problème de la soustraction de fond pour les arrières-plans complexes, le modèle par mélange de gaussiennes GMM (*Gaussian Mixture Model*) [31,32] se montre plus adéquat à notre étude car il peut s'adapter au cas d'un fond en mouvement. La loi du mélange de gaussiennes peut s'écrire comme le montre l'équation 1.10.

$$P(I_t) = \sum_{i=1}^k \omega_{i,t} N(I_t | \mu_{i,t}, \sigma_{i,t}) \quad (1.10)$$

Avec $N(I_t | \mu_{i,t}, \sigma_{i,t})$ la loi gaussienne paramétrée par $(\mu_{i,t}, \sigma_{i,t})$, $\omega_{i,t}$ le poids donné à la $i^{\text{ème}}$ gaussienne. Tout d'abord, les gaussiennes sont ordonnées selon la valeur de ω/σ . Cette valeur augmente dans le cas où le poids de la gaussienne est important alors que sa variance est faible. Ensuite, une ré-estimation des paramètres du mélange de gaussiennes est nécessaire afin de mettre à jour le modèle de fond. Les pixels qui sont éloignés plus de $n\sigma_{i,t}$ fois de la distribution du modèle de fond sont considérés en mouvement. Ce paramètre n peut être fixé à $n = 2.5$ comme le montre Benezeth et al. [31]. Cette technique présente un modèle adapté pour la description d'un fond en mouvement, mais il présente de faible performance dans le cas de fond complexe en mouvement.

1.4.1.5 Différence d'images

La méthode *Frame difference* [28, 31] a pour but de déterminer la position de la cible à suivre en faisant une soustraction élémentaire entre l'image à l'instant t et $t - 1$. Celle-ci est calculée grâce à l'équation 1.11 :

$$D_t = |I_t - I_{t-1}| \quad (1.11)$$

Avec, I_t représente un pixel dans l'image à l'instant t et D_t représente le pixel du même indice dans l'image de référence à l'instant t . Cette image de différence D va mettre en évidence les pixels qui ont subi un changement entre les deux images. Ceci génère une intensité élevée, visuellement remarquable sur ces pixels. Pour localiser l'objet dans l'image de différence D , un seuil S est choisi empiriquement afin de différencier, les objets en mouvements, du bruit. Cette opération de seuillage est effectuée selon l'équation 1.12 :

$$\begin{cases} \text{Si } D_t(x, y) < S & I_t \in \text{arrière plan} \\ \text{Sinon} & I_t \in \text{premier plan} \end{cases} \quad (1.12)$$

Malgré la sensibilité de cette méthode au changement d'éclairage et au mouvement de la caméra, elle peut être adaptée à notre cas de suivi de nageur, et plus particulièrement, à la détermination de la zone d'intérêt. En effet, elle offre une grande simplicité ainsi qu'une bonne efficacité et ses différentes limitations peuvent être contournées grâce à l'utilisation d'une caméra statique.

1.4.1.6 Discussion

Parmi les techniques de soustraction de fond qui existent dans la littérature, nous avons présenté les méthodes les plus répandues dans la littérature et qui nous semblent adaptables à notre cas de suivi des nageurs. Après cette étude, nous retenons la technique de différence d'images sur laquelle nous ferons des tests approfondis par la suite. Ce choix est motivé par sa simplicité ainsi que par la perspective d'une application sur une zone restreinte caractérisée par un couloir spécifique. D'autant plus que cette extraction est possible en utilisant les résultats du calibrage présenté précédemment.

1.4.2 Détection de peau

La détection de peau est souvent utilisée comme une étape de prétraitement, permettant de restreindre la zone d'intérêt, dans les applications de type détection ou suivi de personnes. Elle offre la possibilité d'extraire, par exemple, une partie du corps humain telle que le visage ou les mains. Elle consiste à détecter les pixels qui correspondent à la peau dans une image donnée. Notre objectif consiste à tirer profit de la particularité de la couleur de la peau qui contient une forte composante rouge par rapport à l'eau, qui elle, contient une forte composante bleue. Cette détection de peau doit permettre dans le cas du nageur de restreindre la zone d'intérêt afin d'optimiser le processus de suivi. Pour ce faire, différentes méthodes ont été proposées dans la littérature et validées pour des applications spécifiques. Dans la suite nous présentons les méthodes les plus répandues dans l'état de l'art.

1.4.2.1 Modélisation par une gaussienne

Le modèle gaussien [33] est un modèle paramétrique qui permet d'estimer la distribution des couleurs de la peau par l'estimation de la fonction de densité de probabilité gaussienne p :

$$p(c|peau) = \frac{1}{2\pi\sqrt{|\sigma_{peau}|}} e^{-\frac{1}{2}(c-\mu_{peau})^T \Sigma_{peau}^{-1}(c-\mu_{peau})} \quad (1.13)$$

Avec, c la variable aléatoire représentant la couleur, μ_{peau} et σ_{peau} qui représentent respectivement la moyenne et la matrice de covariance du modèle gaussien. Ces paramètres sont à estimer lors de la phase d'apprentissage selon les équations suivantes :

$$\mu_{peau} = \frac{1}{N_{peau}} \sum_{c \in C} N_{peau}(c)c \quad (1.14)$$

$$\sigma_{peau} = \frac{1}{N_{peau} - 1} \sum_{c \in C} N_{peau}(c)(c - \mu_{peau})(c - \mu_{peau})^T \quad (1.15)$$

Les pixels ayant une couleur proche de la distribution de ce modèle gaussien sont considérés comme étant de la peau. Cependant, la modélisation par une simple gaussienne n'est pas toujours suffisante pour modéliser la distribution couleur de la peau. En effet, celle-ci est généralement plus complexe qu'une simple distribution gaussienne.

1.4.2.2 Modélisation par un mélange de gaussiennes

La modélisation par un mélange de gaussiennes, comme nous l'avons présenté dans la section précédente, est une combinaison pondérée de plusieurs gaussiennes. Jones et al. [34] ont proposé de l'utiliser afin d'améliorer la représentation de la peau par un modèle de mélange de gaussiennes, comme le montre l'équation suivante :

$$p(c|peau) = \sum_{n=1}^N \omega_n p_n(c|peau) \quad (1.16)$$

Avec, p_n une distribution gaussienne, ω_n le poids de la $n^{ième}$ gaussienne, où la somme des ω est égale à 1, et N le nombre de gaussiennes permettant de modéliser la peau. A titre d'exemple Jones et al. [34] ont fixé la valeur de N à 16 mais peut varier selon l'espace couleur. Pour l'apprentissage d'un modèle de la peau, l'algorithme EM (Expectation Maximisation) [34,35] estime les trois paramètres du modèle de mélange de gaussiennes (μ_n , σ_n et ω_n). Cette méthode a montré des résultats intéressants dans la littérature mais reste trop complexe pour la détection de la peau dans le cas des nageurs et du milieu aquatique.

1.4.2.3 Modélisation par une table de correspondance

Cette méthode consiste à représenter les blocs de la peau par un histogramme de couleurs appelé table de correspondance (lookup table) [36]. Chaque case de cette table contient le nombre de pixels appartenant à un intervalle de couleur fixé. Une étape d'apprentissage est nécessaire pour construire la table de correspondance. Pour chaque image de la base d'apprentissage, la peau est sélectionnée et la case correspondant à l'intervalle de couleurs de chaque pixel est incrémentée. Les valeurs de la table de correspondance sont normalisées, afin d'obtenir des valeurs entre 0 et 1, donnant ainsi la probabilité qu'un intervalle de couleur représente la peau. Afin de détecter cette dernière dans une image cible, la table de correspondance est utilisée pour déterminer la possibilité que la couleur d'un bloc corresponde à la peau. Le critère de décision est basé sur un seuil appris lors d'une phase d'apprentissage. Étant donné que les résultats de cette technique dépendent fortement des données d'apprentissage nous choisissons de ne pas l'utiliser pour notre cas de détection de la peau des nageurs.

1.4.2.4 Modélisation basée sur l'appariement d'histogrammes

Cette méthode est basée sur la comparaison des histogrammes des images [37]. Elle nécessite la sélection d'un bloc de référence de peau, aussi appelé bloc de contrôle. Puis, une comparaison est faite entre le bloc de référence et les blocs, de même taille, dans l'image cible. Le critère permettant de différencier les blocs « peau » des blocs « non peau » est le score $M_{c,i}$ calculé par la méthode d'intersection des histogrammes. Ce score est obtenu comme le montre l'équation suivante :

$$M_{c,I} = \frac{\sum_i^N \min(H_c(i), H_I(i))}{H_c(i)} \quad (1.17)$$

Avec H_c et H_I représentent respectivement les histogrammes du bloc de référence et des blocs de l'image cible. N représente le nombre de *bins* (composantes de l'histogramme). Dans le cas où $M_{c,i}$ dépasse un certain seuil, le bloc est considéré comme peau. Une analyse des composantes connexes permet alors de détecter plus précisément les zones de peau. Cette approche n'est pas adéquate à notre sujet car elle nécessite une intervention manuelle pour déterminer le bloc de contrôle [43].

1.4.2.5 Méthodes de segmentation basées sur la détermination d'un intervalle de couleur

Dans les systèmes qui utilisent la détection de peau comme étape de prétraitement, le critère le plus important est la simplicité et la rapidité de la détection. Afin de satisfaire ces contraintes de temps, la méthode de segmentation par intervalle de couleur a été proposée [38, 44]. Elle consiste à définir empiriquement un intervalle de couleur qui représente la couleur de la peau. Celui-ci est obtenu en sélectionnant, dans une image d'entrée, une région de peau permettant de définir les minimums et maximums de chacune des couleurs : rouge, verte et bleu. L'image cible peut alors être binarisée ("peau" et "non peau") grâce à l'équation suivante :

$$ImB(i) = \begin{cases} 1 & \text{si } R_{min}, V_{min}, B_{min} < I_{R,V,B}(i) < R_{max}, V_{max}, B_{max} \\ 0 & \text{sinon} \end{cases} \quad (1.18)$$

Avec $Im(i)$ et $ImB(i)$ qui représentent respectivement les images avant et après la segmentation et où R , V et B sont les composantes rouge verte et bleu du pixel i . Cette méthode de segmentation peut être appliquée sur l'espace de couleur HSV (*Hue*, *Saturation*, *Value* – Teinte, Saturation, Valeur). On note que la composante H représente la partie la plus importante de l'information de couleur lorsqu'elle est séparée de la saturation et de la brillance. Cette technique est adaptée à notre cas de détection de peau des nageurs de part sa simplicité et ses bonnes performances [38, 44].

1.4.2.6 Autres méthodes de détection de peau

Finalement, la détection de peau a fait l'objet de nombreuses autres études dans la littérature telles que la détection de contour [45], la soustraction du fond [46] ou la détection de mouvement [47]. Les méthodes basées sur la détection de contour sont dédiées aux applications spécifiques où il est possible d'avoir une connaissance préalable de la forme de la partie du corps à détecter, permettant ainsi de construire un modèle géométrique. Les informations acquises par les techniques de soustraction de fond et de détection de mouvement sont souvent insuffisantes car elles ne permettent généralement pas de différencier les objets en mouvement de la peau. Ainsi, même si elles sont rapides elles nécessitent souvent l'ajout d'information afin de localiser la peau dans la zone en mouvement.

1.4.2.7 Discussion

Notre objectif consiste à détecter la peau des nageurs afin de les pré-localiser et de restreindre la zone d'intérêt. Pour cela, nous avons étudié et analysé différentes techniques de détection de peau connues dans la littérature. Parmi ces dernières, nous retenons la méthode de segmentation basée sur la détermination d'un intervalle de couleur pour la suite de ce manuscrit pour sa simplicité et ses performances. En effet, cette méthode a l'avantage de pouvoir segmenter la peau du nageur en analysant les couleurs de la scène dans un espace donné (RGB, HSV, etc).

1.4.3 Détection de lignes

Dans le but de se situer dans le bassin, on peut tirer profit de la particularité de sa forme géométrique et l'utiliser comme repère référentiel. Nous parlons ici des bords du bassin et des lignes délimitant les couloirs. Afin de les détecter, nous nous sommes orientés vers l'étude des méthodes de détection de lignes, dans les images, connues dans la littérature [39–42]. Le plus souvent, la détection des lignes est considérée comme une étape de prétraitement délivrant des informations utiles sur la scène à étudier, dans notre cas la piscine et ses lignes d'eau. Parmi les méthodes les plus connues, on peut citer : les méthodes basées sur le principe de *thinning*, les méthodes basées sur l'Analyse en Composantes Principales (ACP) et la transformée de Hough.

1.4.3.1 *Thinning (skeletonization)*

La technique de *thinning*, souvent appelée *skeletonization*, est une technique de vectorisation largement répandue [48]. Cette technique a pour but l'échantillonnage de l'image pour obtenir un squelette, d'un pixel de largeur, en pré-traitement au processus de détection de ligne. Pour cela, des opérations de morphologie mathématique [49] sont appliquées sur l'image originale d'entrée afin de générer son squelette, comme le montre la figure 1.3. Le squelette de la surface noire du nageur est la représentation la plus synthétique possible tout en gardant une structure topologique identique à l'image originale. Par conséquent, l'analyse devient beaucoup plus facile vis-à-vis de l'image originale.

Abdol-Reza Mansouri et al. [50] ont proposé un paradigme d'hypothèse prédiction-vérification. Ce paradigme consiste à prédire les segments de droites et leur orientation selon la position et l'orientation des pixels du contour. Ensuite, une vérification avec des tests statistiques est effectuée sur la ligne dans le but de la valider ou de la rejeter.

Il est à noter que l'objectif de cette technique est d'obtenir les squelettes de tous les objets dans une image. Cette technique est efficace mais a montré certaines limitations liées à la génération du squelette qui crée parfois de fausses branches au niveau des jonctions [48]. Dans notre cas, l'application de la *skeletonization* sur une image de bassin peut générer des lignes (squelette) au centre des couloirs de nage à la place des lignes séparatrices de ces couloirs. Afin de pallier ce problème nous proposons d'étudier d'autres techniques plus adaptées à notre cas de détection de lignes.



FIGURE 1.3 – Résultat de l'application de la technique *thinning* sur l'image binaire d'un nageur.

1.4.3.2 Analyse en Composantes Principales - ACP

L'analyse en composantes principales consiste à réduire l'espace des données en projetant celles-ci sur les axes représentant au mieux leur distribution. Les composantes principales de l'espace de données sont déduites des vecteurs propres et des valeurs propres de la matrice de dispersion des données. Les vecteurs propres représentent la

direction principale de la répartition des points de données tandis que les valeurs propres représentent leur importance. Récemment, des études ont été faite pour la détection des lignes en utilisant l'analyse en composantes principales. Yun-Seok Lee et al. [40] ont proposé d'appliquer le détecteur de contour Canny [51] afin de générer l'image binaire de contour. Sur cette image de contour, les segments de droites horizontaux et verticaux sont extraits selon leurs formes primitives. Ces derniers sont marqués, et l'analyse en composantes principales est réalisée pour chaque segment marqué. L'utilisation des composants principaux permet de détecter les lignes droites ainsi que leurs orientations [52]. En revanche, la littérature montre des résultats insuffisants pour les détecteurs de lignes basées sur la l'utilisation de l'ACP.

1.4.3.3 Transformée de Hough

La transformée de Hough est une technique de détection de lignes très répandue dans la littérature [53, 54]. Des variantes de cette technique pour la détection d'autres formes telles que les cercles et les ellipses [41] ont aussi été proposées.

Le principe de la transformée de Hough se base sur la représentation polaire des points correspondants aux contours présents dans l'image. Les contours peuvent être extraits en utilisant une technique de seuillage et de détection de contour [27, 51]. L'idée de base de la transformée de Hough est de représenter chaque pixel du contour par les coordonnées polaires (θ, ρ) de toutes les droites qui le traversent. Sachant que θ et ρ sont respectivement l'angle et la norme de chacune de ces droites dans l'espace polaire. En d'autres termes, ρ représente la longueur du segment perpendiculaire à la droite de l'angle θ passant par l'origine, comme le montre la figure 1.4.a. Pour chaque point, une sinusoïde est obtenue à partir des coordonnées polaires des droites hypothétiques passant par ce point, comme illustré dans la figure 1.4.b. L'ensemble des sinusoïdes obtenues représente l'espace de Hough où un croisement d'une multitude de sinusoïdes traduit l'alignement d'un ensemble de points dans l'image originale, comme le montre la figure 1.4. Étant donné que cette méthode de détection de lignes donne de très bons résultats [53, 54] nous proposons de l'utiliser afin de détecter les bords du bassin ainsi que les lignes des couloirs de nage.

1.4.3.4 Discussion

Dans le but de restreindre la zone de recherche et de faciliter la localisation et le suivi des nageurs, nous avons fait une étude qui porte sur la détection des lignes. Parmi les méthodes que nous avons présentées, nous choisissons la technique de la Transformée de Hough pour des tests ultérieurs. Ce choix est motivé par ses bonnes performances et sa large utilisation dans la littérature [53, 54].

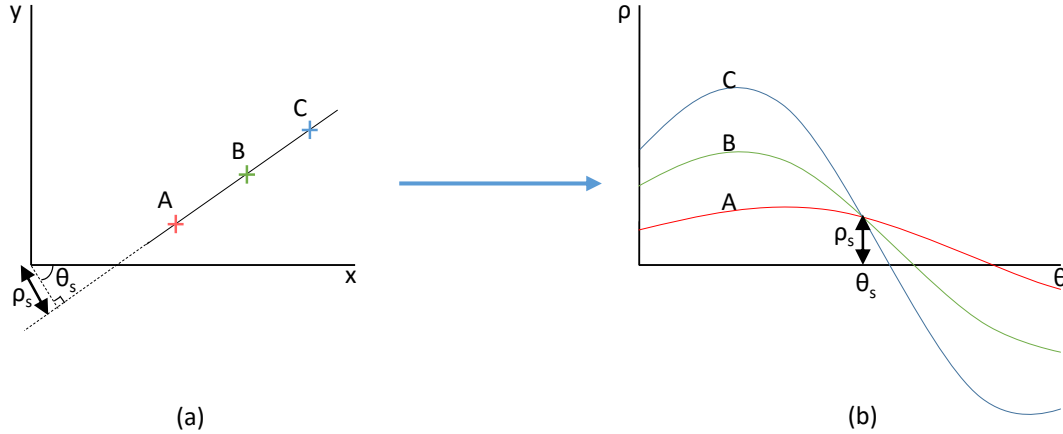


FIGURE 1.4 – Principe de la détection des lignes par la transformée de Hough, les trois point A, B et C (a) sont alignés si leurs courbes dans l'espace de Hough (b) se croisent en un point.

1.5 Suivi par corrélation

Après avoir introduit le calibrage et les prétraitements qui peuvent être utilisé pour faciliter et optimiser le suivi, nous présentons des techniques de suivi connues dans la littérature en commençant par la corrélation. Cette dernière est une technique classique très connue dans le domaine de la détection et de la reconnaissance des formes. Deux architectures différentes sont proposées [15] : celle du VLC (*VanderLugt Correlator*), principalement utilisée pour l'identification, et celle du JTC (*Joint Transform Correlator*) plus robuste pour la détection, la localisation et le suivi d'un objet référence dans une image. Cette dernière a fait ses preuves comme en témoigne les nombreuses publications dans la littérature [7, 12–15]. Dans le cadre d'un système de suivi, l'architecture JTC est privilégiée car plus à même de réaliser les étapes de détection et de localisation en même temps, améliorant de ce fait le temps de calcul. Cependant, cette approche d'origine optique (figure 1.5) est parfois difficile à mettre en place physiquement, particulièrement dans des applications en situation réelle. Pour cette raison, nous proposons ici une implémentation numérique à travers des techniques de traitement d'images.

L'idée de base d'une implémentation toute-optique est d'utiliser deux lentilles, l'une pour faire la transformée de Fourier du plan d'entrée qui contient les images à comparer et l'autre pour faire la transformée de Fourier inverse pour avoir le plan de la corrélation. Entre les deux, un modulateur spatial de lumière (SLM) est placé pour afficher et filtrer le spectre joint, voir figure 1.5. Dans la littérature, différentes méthodes ont été proposées et validées, nous allons présenter, par la suite, les plus répandues et les mieux adaptées à notre cas de suivi de nageurs.

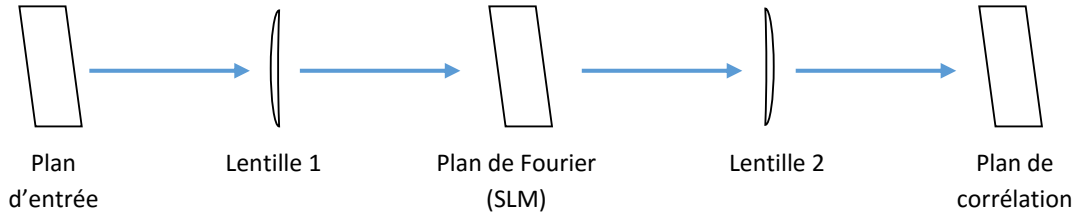


FIGURE 1.5 – Montage optique de la corrélation où la lumière passe par le plan d'entrée puis par une lentille qui le transforme en un plan de Fourier. Un modulateur spatial de lumière (SLM) est placé pour le récupérer et l'afficher avant de le transmettre, à travers une deuxième lentille, au plan de corrélation final.

1.5.1 Cl-JTC (Joint Transform Correlator Classique)

Le principe fondamental de l'approche Cl-JTC (JTC classique) [15] est d'appliquer une transformée de Fourier sur un plan d'entrée constitué d'une image référence et d'une image cible, i.e. scène cible, pour obtenir un spectre joint de ces images. Ensuite, une transformée de Fourier inverse est appliquée sur l'intensité de ce spectre afin d'obtenir le plan de corrélation contenant les informations de similarité référence/cible. La formulation du Cl-JTC (figure 1.6) peut être formulée comme le montre les équations suivantes :

$$f(x, y) = c(x, y) + r(x - d, y - d) \quad (1.19)$$

Où f représente le plan d'entrée constitué de l'image cible $c(x, y)$ et de la référence $r(x - d, y - d)$ placées à une distance d . Le spectre joint F est alors obtenu par la transformée de Fourier appliquée au plan d'entrée :

$$F(u, v) = |C(u, v)|\exp[\phi_c(u, v)] + |R(u, v)|\exp[\phi_r(u, v)]\exp[-j(ud + vd)] \quad (1.20)$$

L'intensité du spectre est alors récupérée par le calcul de son module au carré, comme le montre la formule suivante :

$$\begin{aligned} G_{Cl-JTC}(u, v) &= |F(u, v)|^2 \\ &= |C(u, v)|^2 + |R(u, v)|^2 + |C(u, v)|\exp[\phi_c(u, v)] \\ &\quad |R(u, v)|\exp[\phi_r(u, v) + j(ud + vd)] \\ &\quad |C(u, v)|\exp[-\phi_c(u, v)]|R(u, v)|\exp[\phi_r(u, v) - j(ud + vd)] \end{aligned} \quad (1.21)$$

Finalement, le plan de corrélation est obtenu par application de la transformée de Fourier inverse sur l'intensité du spectre joint G_{Cl-JTC} .

Dans le cas où une similarité existe entre les deux images, le plan de corrélation obtenu va contenir trois pics : un pic d'autocorrélation très intense qui représente la corrélation de chaque image avec elle-même et deux pics d'intercorrélation entre les images référence et cible. La distance entre ces deux pics en fonction de l'axe des X et

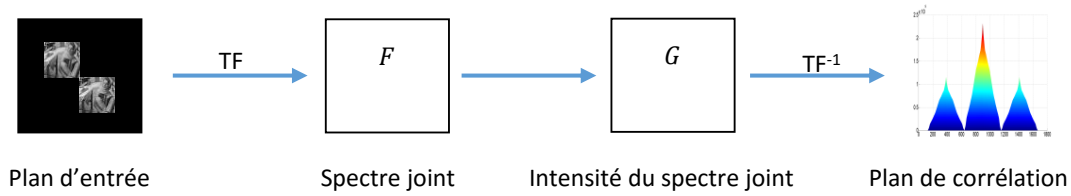


FIGURE 1.6 – *Processus global du JTC Classique.*

l'axe des Y est égale à $2d$ et la droite qui les lie forme un angle α avec l'axe des X , comme le montre la figure 1.7. Dans le cas d'un système de suivi, seule cette dernière information est intéressante car elle représente le degré de ressemblance entre nos deux images ainsi que la position de la référence dans la cible (grâce à la direction et à la norme du segment reliant les deux pics). Différentes vues du plan de corrélation du Cl-JTC sont présentées dans la figure 1.7.

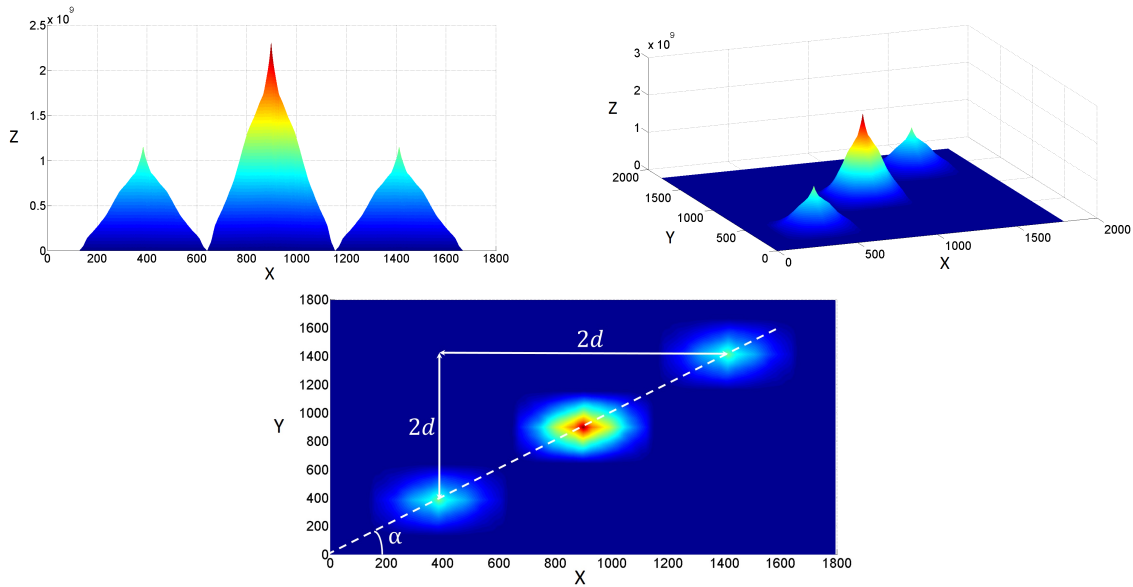


FIGURE 1.7 – *Différentes vues 3D du plan de corrélation issu du Cl-JTC. Le pic central correspond au pic d'autocorrélation qui représente la ressemblance entre chaque image et elle-même. Les deux pics d'intercorrélation permettant de représenter la ressemblance entre l'image de référence et l'image cible.*

Le Cl-JTC représente l'approche la plus classique du JTC. De ce fait, elle montre une certaine faiblesse, à savoir, l'existence d'un pic d'autocorrélation très large et très intense. Ceci peut perturber la détection et la largeur des pics d'intercorrélations et générer des imprécisions pour la localisation. Par conséquent, différentes approches ont été proposées : NZ-JTC, B-JTC, etc.

1.5.2 NZ-JTC (JTC Non-Zéro)

Ce qui est intéressant dans le plan de corrélation est l'information issue des deux pics d'intercorrélation, en revanche, l'information issue du pic d'autocorrélation est inutile. Afin d'éliminer ce pic central, le NZ-JTC a été proposé [12]. Le pic d'autocorrélation est issu de la transformée inverse des termes $|C(u, v)|^2 + |R(u, v)|^2$. Afin de l'éliminer il est possible d'éliminer ces deux termes. L'équation 1.21 devient :

$$G_{NZ-JTC}(u, v) = |F(u, v)|^2 - |C(u, v)|^2 + |R(u, v)|^2 \quad (1.22)$$

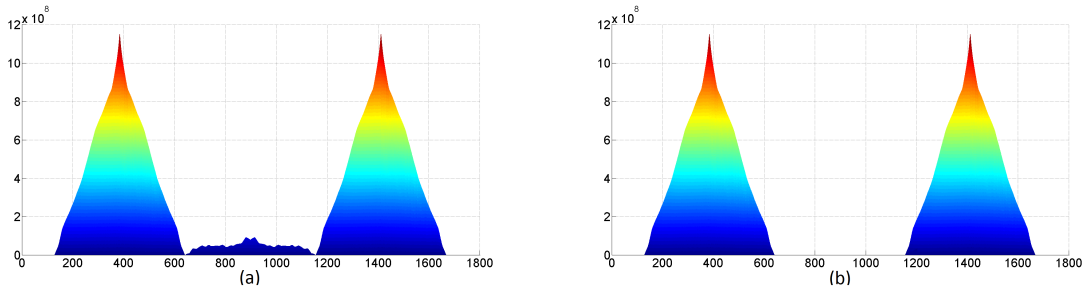


FIGURE 1.8 – Plans de corrélation de l'approche NZ-JTC. (a) Plan de corrélation issu de l'application de la formule 1.22. (b) Plan de corrélation obtenu après le filtrage de la zone d'autocorrélation.

Comme nous pouvons le constater dans la figure 1.8.a, l'élimination du pic d'autocorrélation en utilisant la formule 1.22 laisse des résidus. Cependant, ces derniers peuvent être plus importants que les pics d'intercorrélation dans le cas d'une faible ressemblance entre la référence et la cible. Pour corriger cela, un filtre peut être appliqué dans le plan de corrélation dans le but d'enlever intégralement ce pic d'autocorrélation. Ce filtre consiste en une fenêtre de la taille de la cible contenant des valeurs nulles qui sera appliqué au centre du plan de corrélation. Le résultat est présenté dans la figure 1.8.b. Dans notre cas de suivi de nageur, cette technique n'est pas adéquate à cause de la largeur des pics de corrélation qui peut falsifier les résultats lorsque la cible à suivre présente de petites déformations. Notons que par la suite nous garderons le principe *Non-Zéro* pour toutes les méthodes basées sur la corrélation JTC.

1.5.3 B-JTC (JTC Binaire)

Les pics de corrélation obtenus par le JTC sont relativement larges ce qui diminue la discrimination. Dans le but d'affiner ces pics, une binarisation de l'intensité du spectre joint est effectuée en utilisant un seuil S qui peut être considéré comme la médiane de

l'intensité du spectre joint [55]. Cependant, nous proposons d'appliquer la binarisation sur l'intensité du spectre joint de la technique NZ-JTC directement afin d'éliminer le pic d'autocorrélation. Finalement, le spectre joint binaire est obtenu par la formule suivante :

$$G_{B-JTC}(u, v) = \begin{cases} 1 & \text{si } G_{NZ-JTC}(u, v) > S \\ -1 & \text{sinon} \end{cases} \quad (1.23)$$

Le plan de corrélation obtenu en utilisant l'architecture B-JTC est présenté dans la figure 1.9.

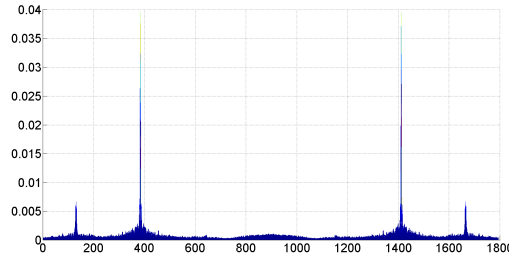


FIGURE 1.9 – Plan de corrélation de l'approche B-JTC.

Cette approche a montré sa précision en terme de localisation et de discrimination. En revanche, elle se montre très sensible envers le bruit et les différences minimales entre la référence et la cible ce qui la rend difficile à utiliser dans le milieu aquatique.

1.5.4 FA-JTC (*Fringe-adjusted JTC*)

Afin d'avoir une architecture discriminante et robuste au bruit, Alam et al. [56] ont proposé d'appliquer un filtre adapté sur le plan de l'intensité du spectre joint. Ce filtre est appelé FAF (*Fringe-Adjusted Filter*) et est décrit par la formule suivante :

$$FAF(u, v) = \frac{B(u, v)}{A(u, v) + |R(u, v)|^2} \quad (1.24)$$

Où A et B peuvent être soit des constantes soit des fonctions. B permet d'obtenir un gain optique supérieur à l'unité. A peut être une constante de petites valeurs afin d'éviter le problème des pôles (dans le cas où la référence contient des pixels nuls) ou une fonction utilisée pour supprimer le bruit dans le cas où ce dernier est connu. Le plan de corrélation de cette technique est représenté dans la figure 1.10. Cette technique a montré des bons résultats, notamment, dans les images bruitées. En revanche, elle reste moins robuste dans le cas des déformations minimes. Pour cette raison, elle n'est pas adéquate à notre cas de suivi de nageurs où la cible (tête du nageur) subit beaucoup de déformations au cours de la séquence vidéo.

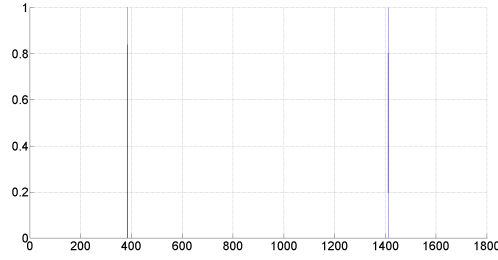


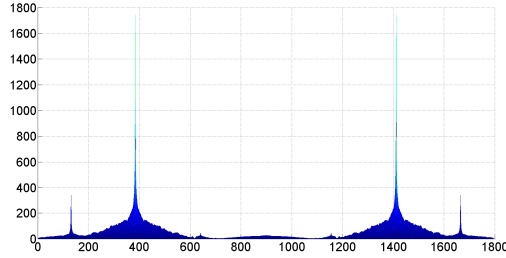
FIGURE 1.10 – Plan de corrélation de l'approche FA-JTC.

1.5.5 NL-JTC (JTC Non-Linéaire)

Finalement, la largeur des pics de corrélation peut être contrôlée grâce à un seuillage non-linéaire de l'intensité du spectre joint [57]. Nous proposons d'appliquer ce seuillage sur l'intensité du spectre joint de la technique NZ-JTC directement afin d'éliminer le pic d'autocorrélation. Le spectre joint de la technique NL-JTC est obtenu par la formule suivante :

$$G_{NL-JTC}(u, v) = \alpha |G_{NZ-JTC}(u, v)|^k \quad (1.25)$$

Où α est un facteur de pondération et k est le degré de non-linéarité. Dans la figure 1.11, nous choisissons $k = 0.5$. $k = 0$ représente le cas du B-JTC, discriminatif et moins robuste, alors que $k = 1$ représente le cas du NZ-JTC, robuste et moins discriminatif. Ainsi, il est possible de trouver un compromis entre robustesse et discrimination en faisant varier k selon le besoin. Pour cette raison, l'approche NL-JTC est adaptée à notre cas de suivi de nageur et nous l'utiliserons donc dans les méthodes présentées par la suite.

FIGURE 1.11 – Plan de corrélation de l'approche NL-JTC avec $k = 0.5$.

1.5.6 Critère de décision basé sur le *PCE*

Afin évaluer les techniques basées sur le JTC, le PCE (*Peak to Correlation Energy*) peut être calculé [15]. Il représente le rapport entre l'énergie contenue dans le pic de

corrélation et l'énergie contenue dans le plan de corrélation. Plus l'image cible et l'image référence sont similaires en terme d'apparence, plus le pic de corrélation est élevé et énergétique par rapport au plan de corrélation. Le PCE peut être obtenu par :

$$PCE = \frac{\sum_{i,j}^{n,m} E(i,j)}{\sum_{i,j}^{N,M} E(i,j)} \quad (1.26)$$

Où, $n*m$ est la taille de la zone autour du pic, alors que $N*M$ représente la taille du plan complet. Enfin, E correspond à l'énergie d'un point du plan. Cependant, le critère PCE devient significativement sensible dans le cas des plans de corrélation bruités. Pour pallier ce problème, d'autres critères peuvent être utilisés [7] tels que, par exemple, la comparaison des histogrammes de couleurs de la référence et la cible détectée en utilisant des mesures de distance comme : Bhattacharyya, L^2 (norme euclidienne) ou χ^2 (*khi - deux*) [58].

1.5.7 Discussion

Dans cette section, nous avons présenté plusieurs variantes de la corrélation JTC pour la détection et le suivi d'objet. Après avoir étudié ces variantes, nous avons retenu la méthode NL-JTC car elle propose le contrôle du degré de non-linéarité. Ceci nous permet de contrôler la finesse des pics selon le besoin. Dans notre cas, nous cherchons un facteur de non-linéarité qui permet d'obtenir le meilleur compromis entre la discrimination et la robustesse. Pour cela, nous ferons une étude afin de fixer ce degré de non-linéarité. Il est à noter que la variante NL-JTC que nous utilisons tout au long de notre étude consiste en un seuillage non-linéaire appliqué sur l'intensité du spectre joint non-zéro, comme présenté dans la formule 1.25. Cette méthode validée par les nombreuses publications dans la littérature [7, 12, 15, 57], donne de bons résultats pour le suivi. En revanche, sa limitation reste son faible pouvoir discriminant, de l'objet à suivre, qui se base principalement sur le contour. En effet, celui-ci peut subir des déformations (parfois majeures) engendrant une perte de l'objet à suivre. Afin de s'adapter à notre cas de suivi de nageur, nous étudierons d'autres techniques de description basées sur différents descripteurs que nous combinerons à la corrélation pour obtenir un système de suivi de nageurs robuste.

1.6 Suivi par des techniques basées sur les histogrammes

Les méthodes basées sur les histogrammes ont pour but de décrire numériquement les caractéristiques de couleur, de texture ou de contour d'un objet. Pour cela elles codent, dans chaque composante d'un histogramme, l'occurrence d'une information particulière dans l'image. Il est utile de noter qu'un unique histogramme peut être calculé sur toute l'image, on parle dans ce cas de description globale, mais qu'il est aussi possible de calculer plusieurs histogrammes, un pour chaque région de l'image, et on parle dans ce cas de description locale. Dans cette section, nous allons présenter les techniques les plus répandues dans la littérature comme histogrammes de couleurs, spatiogrammes, HOG (histogrammes de gradients orientés) et histogrammes de descripteur LBP (Motifs Binaire Locaux).

1.6.1 Histogramme de couleurs

Cette approche utilise la couleur afin de décrire l'objet à suivre. Dans le cas du suivi des nageurs, l'information de couleur est très importante car, souvent, il y a un contraste entre les différents objets présents dans le bassin. Par exemple, l'eau a une forte composante bleue, la peau une forte composante rouge et le bonnet peut avoir différentes couleurs. Dans ce contexte, l'idée principale de cette technique est de représenter chacune des deux images, référence et cible, par un histogramme où chaque composante code le nombre de pixels d'une couleur donnée [16–18]. Le nombre de composantes de l'histogramme, qui correspond aux intervalles de couleur pris en compte, est un paramètre de la méthode à définir à l'aide d'un ensemble d'apprentissage.

La comparaison entre les deux histogrammes, qui correspond à la ressemblance entre l'image référence et l'image cible, s'obtient grâce au calcul de la distance en utilisant les coefficients de Bhattacharyya, L^2 ou χ^2 (*chi*²) [58]. Dans le cas du suivi, cette méthode s'appuie sur l'algorithme de recherche exhaustive en définissant une zone d'intérêt dans laquelle l'histogramme de l'image de référence est comparé avec les histogrammes des images cibles. L'image sélectionnée correspond alors à celle ayant donné une distance minimum par rapport à la référence.

Malgré sa sensibilité par rapport au changement des conditions d'éclairage et aux occultations, cette méthode de suivi basée sur l'histogramme de couleurs est très robuste, notamment, par rapport à l'orientation de l'objet à suivre. Pour cette raison, nous l'utiliserons par la suite pour notre application de suivi des nageurs.

1.6.2 Spatiogramme

Même si les histogrammes ont montré leur intérêt pour représenter les couleurs d'une image, ils ne prennent pas en compte les informations spatiales qui peuvent fortement influencer la discrimination entre l'objet à suivre et le reste de la scène. Pour contourner cette limitation, Birchfield et Rangarajan ont récemment introduit le concept de Spatiogramme [59], qui ajoute des informations spatiales, moyenne spatiale et covariance des pixels, aux histogrammes. Cependant, cette approche très discriminante rend parfois le suivi trop sensible par rapport à l'orientation et à la déformation. De ce fait, cette méthode est peu adaptée au milieu aquatique qui présente de fortes déformations liées à l'eau.

1.6.3 HOG

Cette méthode est basée sur la construction d'histogrammes permettant de stocker l'orientation des gradients dans une image [21]. L'idée des auteurs est que l'apparence et la forme d'un objet peuvent être caractérisées par la distribution des gradients d'intensité, ou par les contours, sans la nécessité de connaître leurs emplacements.

Le calcul du gradient de chaque pixel s'effectue en appliquant un filtre dérivatif $1-D$ centré, dans les directions horizontales et verticales. Puis, pour chaque pixel l'angle et l'amplitude du gradient sont stockés dans un histogramme appelé histogramme de gradient orienté.

Cette méthode robuste utilise principalement l'information de gradient pour effectuer le suivi. Dans le cas du suivi, certaines déformations locales, dues aux éblouissements par exemple, vont affecter les contours, et donc les gradients, et rendre la méthode peu discriminante. Pour pallier ces limitations, nous combinerons cette approche avec d'autres méthodes pour tirer parti de ses avantages et renforcer la robustesse de notre algorithme de suivi des nageurs

1.6.4 LBP

Les motifs binaires locaux (LBP - *Local Binary Patterns*) [19, 20] sont des descripteurs qui décrivent chaque pixel d'une image par leur texture, calculée par rapport au niveau de gris relatif des pixels voisins. Le point fort de cette approche est sa robustesse par rapport à la variation des conditions d'éclairage obtenue grâce à la relation locale entre le pixel et ses voisins. Pour chaque pixel le descripteur LBP est obtenu grâce à l'algorithme de la figure 1.12. Premièrement, une relation entre les intensités d'un pixel central et de ses voisins est calculée en soustrayant l'intensité de ceux-ci par l'intensité

du pixel central. Ensuite, on code par le bit 1 les voisins qui ont une intensité plus forte que le pixel central et par le bit 0 les autres. Grâce à une pondération, chaque pixel est décrit par un nombre entre 2^0 et 2^7 (8 bits). Enfin, la somme de ces nombres donne la code LBP du pixel central et le descripteur final est une matrice de codes LBP qui peut être représentée par un histogramme.

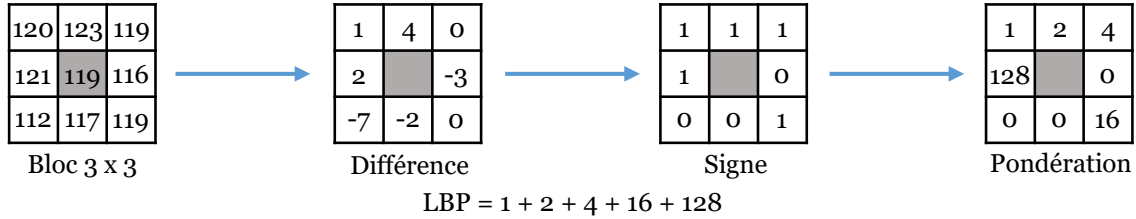


FIGURE 1.12 – *Processus de calcul du motif binaire local (LBP).*

Plusieurs variantes des LBP existent. Parmi les plus connues, on peut citer les LBP uniformes qui permettent de surmonter le problème de la sensibilité au bruit et les LBP invariants en rotation qui ajoutent une invariance à la rotation.

Les descripteurs LBP uniformes ne s'intéressent qu'aux codes dits "uniformes". Ceux-ci sont les codes qui contiennent au plus deux transitions de 0 vers 1 ou 1 vers 0. Les autres codes sont considérés comme non-uniforme (bruit). Par exemple : 11100011 contient deux transitions, donc c'est un code uniforme par contre 11010101 ne l'est pas car il contient six transitions. La représentation de l'image se fait en calculant l'histogramme des codes uniformes puis en gardant tous les codes non-uniformes dans la première composante de l'histogramme [20].

Pour avoir une représentation robuste par rapport à la rotation il suffit de considérer tous les codes qui contiennent le même nombre de 1 et de 0 comme étant les mêmes codes, ce que l'on appelle LBP *rotation-invariant*. Par exemple : 11100101 = 10010111 = 11111000 [20].

Il est possible de considérer comme identique les codes uniformes qui disposent d'un même nombre de 1 et de 0. Ainsi, les LBP obtenus permettent d'avoir une approche robuste par rapport au bruit et à la rotation mais sensible aux déformations et aux occultations [20]. De part ses bonnes performances et sa robustesse par rapport au bruit et à la rotation nous utiliserons par la suite l'approche LBP *uniforme rotation-invariant*, en association avec d'autres techniques, afin de renforcer notre suivi des nageurs.

1.6.5 Discussion

Dans cette section nous avons présenté des approches de suivi par histogrammes basées sur différents descripteurs : la couleur, le gradient et la texture. Dans le prochain chapitre, nous utiliserons les approches basées sur l'histogramme de couleurs, HOG et LBP afin d'effectuer des tests et de les adapter à notre approche de suivi des nageurs. Partant du principe que chaque méthode a des avantages et des limitations, nous proposerons de fusionner ces techniques afin de créer une nouvelle approche plus robuste et plus discriminante. Dans ce cadre, nous présentons dans la suite une étude générale traitant de la fusion de descripteurs.

1.7 Fusion de descripteurs

La fusion de données est très répandue dans les domaines du traitement d'image et du signal, notamment dans les systèmes de reconnaissance biométrique [60, 61] qui exigent une robustesse élevée. Pour ce type de systèmes, la fusion de données peut être faite aux niveaux des différents modules : acquisition, paramètres extraits, scores et décisions. Dans le reste de cette section, nous allons répondre aux questions suivantes : pourquoi ? et comment ? utilisons-nous la fusion de données dans le domaine du suivi de nageurs.

1.7.1 Pourquoi la fusion

Malgré l'avancement de la recherche en traitement d'image ainsi que la robustesse et la précision des approches utilisées pour le suivi, cela reste insuffisant. En effet, la majorité des approches ont été proposées et validées selon des critères environnementaux bien définis. Dans les environnements non-contrôlés, comme dans notre cas, plusieurs difficultés sont souvent rencontrées. Nous citons, à titre d'exemple, les difficultés suivantes :

- Bruit lié au capteur utilisé.
- Changement des conditions d'éclairage de la scène filmée.
- Occultation partielle ou totale de l'objet à suivre.
- Déformation de la forme de la cible.
- Changement d'échelle.

Afin de surmonter ces difficultés, la recherche récente vise à introduire la fusion de données pour bénéficier des avantages de plusieurs types de données qui, fusionnées, permettent de réduire les limitations évoquées ci-dessus.

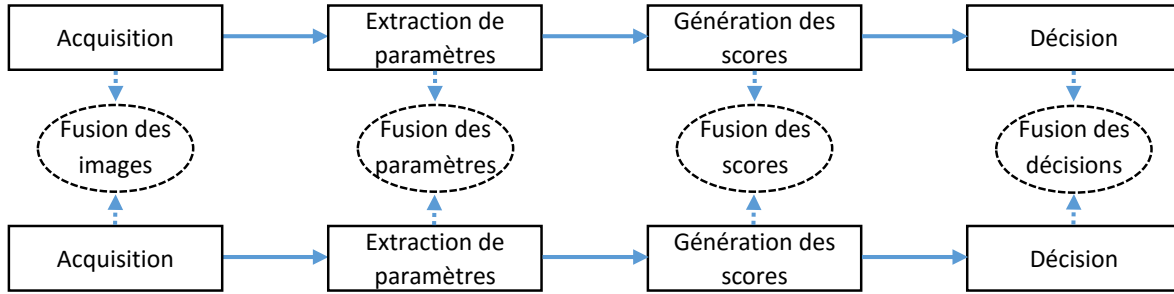


FIGURE 1.13 – Niveaux de fusion possibles pour deux systèmes de suivi. La fusion au niveau des scores est la plus répandue, cependant, les autres niveaux sont applicables selon les objectifs du système.

1.7.2 Niveaux de fusion

Comme nous l'avons mentionné dans le préambule de cette section, et comme pour les systèmes de reconnaissance biométrique, nous pouvons distinguer différents niveaux pour la fusion de données dans les méthodes de suivi. La figure 1.13 montre ces niveaux de fusion pour deux systèmes de suivi distincts. Une fois la fusion effectuée au niveau d'un des modules de suivi, le reste des modules est effectué une seule fois.

1.7.2.1 Au niveau de l'acquisition

Dans ce niveau, les données capturées doivent être homogènes pour pouvoir les fusionner. Cela doit être fait de manière que l'on puisse créer de nouvelles données à partir de celles qui sont issues des différents capteurs. Par exemple, il est possible de fusionner plusieurs images capturées par différents types de caméras (caméra thermique, caméra 4K, etc.) mais il est nécessaire d'ajuster leur résolution respective au moment de la fusion.

1.7.2.2 Au niveau des paramètres extraits

Ce niveau est moins limité par rapport à la fusion que celui du niveau des données acquises. Il consiste à combiner les descripteurs obtenus en appliquant les différentes méthodes d'extraction des paramètres (LBP, HOG, etc.). En revanche, il peut être difficile de fusionner ces paramètres à cause de la différence entre leurs intervalles de valeurs et leur interprétation. Souvent, la fusion, dans ce niveau, se fait par concaténation des paramètres extraits [62].

1.7.2.3 Au niveau des scores de détection

Le score est une valeur attribuée à chaque détection pour mesurer la ressemblance entre la référence et la cible. C'est le niveau le plus utilisé pour la fusion car il peut être appliqué à tous les types de systèmes, contrairement aux niveaux précédents. Après une étape de normalisation de scores (détaillée dans la section 1.7.3), les valeurs de scores de détection seront dans un espace de dimension limitée qui reflète la confiance envers la détection.

1.7.2.4 Au niveau des décisions

Ce niveau de fusion concerne les décisions finales. Dans ce cas, les systèmes à fusionner s'exécutent séparément jusqu'à la dernière étape de décision. Ensuite, la décision finale consiste à favoriser une des décisions ce qui revient à l'utilisation d'un sous ensemble de systèmes de suivi. Ceci représente le principal point faible de ce type de fusion.

Après cette étude sur les différents niveaux de fusion possibles, nous retenons la fusion au niveau des scores. En effet, celui-ci semble être le plus adapté à notre application de suivi de nageurs car il permet de générer de nouveaux scores, plus pertinents, en se basant sur des scores issus d'approches variées.

1.7.3 Normalisation

Les scores finaux (après la fusion) représentent des facteurs de confiance robustes issus d'une fusion réelle de plusieurs systèmes. Avant de fusionner les scores issus de chaque système de suivi, il est important de noter qu'ils peuvent avoir une nature différente. Certains systèmes produisent des scores de vraisemblance (similarité) tandis que d'autres produisent des scores de distance (dis-similarité). De plus, chaque système peut avoir des intervalles de variations des scores différents, par exemple, entre 0 et 1 pour un premier système et entre -10 et 1000 pour un second. Dans ce cadre, afin de pouvoir paramétrer le poids donné à chaque méthode, il est nécessaire de normaliser leur dynamique afin de les rendre cohérentes, par exemple $[0, 1]$.

La normalisation consiste à projeter un score, i.e. un plan dans notre cas, sur un espace de valeurs cible, souvent, en changeant sa moyenne (position) et son écart type (l'échelle), on parle souvent de données centrées réduites. Pour cela, plusieurs techniques existent [60,61]. Parmi les plus répandues, on cite la technique *Z-score* [61] qui normalise la valeur $p(i, j)$ d'un ensemble de données noté p par :

$$p_{nor}(i, j) = \frac{(p(i, j) - \mu)}{\sigma} \quad (1.27)$$

Avec μ et σ qui représentent respectivement la moyenne et l'écart type de l'ensemble de données à normaliser. Cependant, cette méthode ne permet pas d'obtenir des valeurs dans l'intervalle $[0, 1]$ et a l'inconvénient de changer la distribution des données.

Pour pallier ce problème, la technique *Min-Max* [61] peut être utilisée. Cette technique consiste à normaliser les données dans un intervalle $[0, 1]$ tout en conservant leur distribution originale. La normalisation de la valeur $p(i, j)$ est donné par :

$$p_{nor}(i, j) = \frac{(p(i, j) - \min(p))}{(\max(p) - \min(p))} \quad (1.28)$$

1.7.4 Types de fusion

Grâce à l'étape de normalisation précédente, les scores obtenus partagent une dynamique commune. Ainsi, ces scores peuvent être combinés par l'application d'un opérateur de fusion. Ce dernier permet d'obtenir des nouveaux scores S synthétisant les informations issues des scores des N différents systèmes fusionnés S_i pour $i = 1..N$. Les opérateurs les plus répandus sont [60, 61] :

— Fusion par minimum :

$$S = \min(S_i) \quad (1.29)$$

— Fusion par maximum :

$$S = \max(S_i) \quad (1.30)$$

— Fusion par moyenne :

$$S = 1/N \sum_{i=1}^N S_i \quad (1.31)$$

— Fusion par médiane :

$$S = \text{mediane}(S_i) \quad (1.32)$$

— Fusion par produit :

$$S = \prod_{i=1}^N S_i \quad (1.33)$$

— Fusion par sommation simple :

$$S = \sum_{i=1}^N S_i \quad (1.34)$$

— Fusion par sommation pondérée :

$$S = \sum_{i=1}^N \omega_i S_i \quad (1.35)$$

Le choix du type de fusion dépend des techniques de suivi utilisées et des objectifs visés par le système final. Le dernier opérateur est souvent le plus adapté car il permet de combiner chaque score en lui attribuant un poids. Celui-ci peut être vu comme un facteur de confiance qu'il est possible de faire varier en fonction de la robustesse de chacune des techniques de suivi utilisées.

1.8 Prédiction de mouvements

Afin de concevoir des systèmes de suivi performants et rapides, il est nécessaire d'éviter de chercher l'objet à suivre dans l'image entière, au risque d'augmenter les détections de faux positifs et d'alourdir le temps de calcul. Pour cela, il est utile de restreindre la région de recherche à la zone correspondant à la dernière détection de l'objet à suivre. Afin d'actualiser l'emplacement de cette zone de recherche, il est indispensable de connaître l'historique de la position de l'objet à suivre afin d'estimer ses mouvements et ainsi prédire son prochain déplacement. Dans ce contexte, nous allons présenter les modèles temporels les plus connus pour la prédiction de mouvement.

1.8.1 Filtre de Kalman

Le filtre de Kalman est un estimateur statistique récursif [63–66]. Dans ce type de méthodes, l'estimation de l'état courant dépend nécessairement des états précédents et des mesures actuelles. Pour des modèles linéaires gaussiens, le filtre de Kalman donne une solution optimale récursive, grâce à des expressions analytiques de moyenne et de covariance de la distribution gaussienne $P(x_t|x_{1..t})$. Les filtres de Kalman sont basés sur des systèmes dynamiques linéaires discrétisés dans le domaine temporel. L'état du système $x_t \in R^d$ à l'instant t est calculé en utilisant l'état du système à l'instant $t - 1$:

$$x_t = A_t x_{t-1} + w_t \quad \text{avec } w_t \sim N(0, \sigma_t) \quad (1.36)$$

Où A_t est le modèle de transition de l'état et w_t est un bruit gaussien de moyenne nulle et de covariance σ_t . L'observation y_t de l'état réel x_t est faite selon l'équation suivante :

$$y_t = B_t x_t + v_t \quad \text{avec } v_t \sim N(0, \sigma_t) \quad (1.37)$$

Avec B_t qui représente le modèle d'observation localisant l'espace d'état par rapport à l'espace observé et v_t le bruit d'observation qui est supposé gaussien de moyenne nulle et de covariance σ_t .

Comme tous les filtres stochastiques, le filtre de Kalman s'applique en deux étapes : prédiction et mise à jour [63]. La phase de prédiction se base sur l'état estimé à l'instant précédent et l'historique des estimations pour produire une estimation fiable de l'état

courant. Dans l'étape de mise à jour, les observations de l'instant courant sont utilisées pour corriger et ajuster l'état prédit dans le but d'obtenir une estimation plus précise.

Plusieurs variantes ont été proposées pour généraliser le filtre de Kalman classique en se basant sur l'hypothèse de la non-linéarité. Parmi les variantes non linéaires les plus connues, on peut citer EKF (*Extended Kalman Filter*) [63] et UKF (*Unscented Kalman Filter*) [67]. Malgré les bons résultats de cette approche, celle-ci est peu adaptée à notre application de suivi de nageurs car elle nécessite de prendre en compte les spécificités de mouvements des nageurs.

1.8.2 Filtre particulaire

La technique du filtrage particulaire fait appel à la méthode de *Monte Carlo* [68]. Celle-ci consiste à utiliser des tirages aléatoires pour réaliser le calcul d'une quantité déterministe. En répétant un grand nombre de fois une expérience, de façon indépendante, on obtient une approximation de plus en plus fiable de la vraie valeur de l'espérance du phénomène observé.

Le filtrage particulaire connaît actuellement un fort développement dans de nombreux domaines dont les sciences de l'ingénieur et les sciences et techniques de l'information et de la communication tel que la localisation, la robotique, le traitement du signal audio ou encore la vision par ordinateur [69–72]. Il s'agit d'une méthode stochastique dans laquelle des particules explorent l'espace d'état, ces particules interagissent sous l'effet d'un mécanisme de sélection qui concentre automatiquement les particules dans les régions d'intérêt de l'espace d'état. Chaque particule est munie d'un poids qui représente la probabilité que l'état qu'elle décrit soit celui du processus à estimer.

Le suivi d'objet peut être formalisé dans un cadre stochastique Bayésien représenté par un système Markovien caractérisé par un vecteur d'état x_t à l'instant t . Un ensemble de mesures z_t représente l'évolution du système. La distribution du vecteur d'état initial $p(x_0)$ et l'évolution temporelle du vecteur d'état $x_k(p(x_k|x_{k-1}))$ sont supposées connues. L'objectif du filtrage particulaire est d'estimer $p(x_k|z_{1..k})$. Cette méthode montre de bonnes performances mais reste complexe à mettre en place.

1.8.3 Mean-shift

Mean-shift est une approche d'analyse de l'espace de données qui a pour but de localiser son maximum [18, 73–75]. Elle consiste à faire converger un point vers le maximum local le plus proche en appliquant un algorithme itératif. Etant donné x_i , n points tel que $i = 1..n$ dans l'espace R^d de dimension d , le vecteur *Mean-shift* $M_\sigma(x)$ pour un

noyau gaussien K de covariance σ est donné par :

$$M_{\sigma}(x) = \frac{\sum_{i=1}^n x_i K\left(\frac{x-x_i}{\sigma}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{\sigma}\right)} - x \quad (1.38)$$

Le filtrage *Mean-shift* est obtenu en appliquant à chaque itération le déplacement donné par cette équation.

Dans le domaine du traitement d'images, cette technique est utilisée pour localiser une région précise dans une image avec la contrainte que cette zone soit proche, en terme de similarité d'histogrammes, d'une zone de référence. La distance entre les deux histogrammes est mesurée à l'aide de leur distance de Bhattacharyya [58]. La recherche de la cible est effectuée en appliquant plusieurs itérations de *Mean-shift*, en commençant par la position estimée, à partir des coordonnées de la cible détectée dans l'itération précédente. L'approche *Mean-shift* est adaptée aux techniques de suivi basées sur les histogrammes. Dans ce cadre, nous l'utiliserons et l'adapterons dans notre approche de suivi utilisant des histogrammes.

1.8.4 Discussion

Dans cette section, nous avons présenté des techniques d'analyse et de prédiction du mouvement permettant d'optimiser le suivi. Dans la suite nous utiliserons l'approche *Mean-shift* qui est adaptée aux méthodes basées sur les histogrammes. Cependant, nous verrons qu'il est intéressant de prendre en compte les spécificités du milieu aquatique et particulièrement les spécificités de la natation afin d'améliorer la prédiction de mouvement du nageur.

1.9 Systèmes de suivi

Le suivi des objets à partir de séquences vidéos a de nombreuses applications [11, 21, 76–78] dans le domaine de la vision par ordinateur dont : la vidéo surveillance, le suivi et la reconnaissance de personnes, la création d'environnements intelligents, le suivi des véhicules, etc. Le suivi des objets permet de déterminer leurs trajectoires dans chaque image de la vidéo et les données statistiques extraites peuvent être utilisées pour compter, surveiller et évaluer le mouvement de ces objets.

Dans cette section nous montrerons comment sont combinées les étapes de suivi citées ci-dessus et nous présenterons en détail différents systèmes de suivi. Parmi les systèmes les plus répandus dans la littérature, nous avons choisi le système de suivi de personnes, de suivi de visages et de suivi de véhicules. Pour conclure nous aborderons les applications de suivi pour le domaine de la natation.

1.9.1 Suivi de personnes

Le suivi des personnes fait l'objet de plusieurs études [11, 21, 79–81]. Dalal and Triggs [21] ont proposé un système basé sur une description par HOG (Histogramme de Gradient Orienté) et le classifieur par SVM (*Support Vector Machine*). Ce système nécessite une étape d'apprentissage où les vecteurs HOG sont extraits pour chaque classe, “personne” et “non personne”, afin d'entraîner le classifieur SVM. L'étape de suivi se fait en parcourant chaque image de la vidéo par une fenêtre glissante, afin d'extraire les vecteurs HOG, puis en les classifiant dans l'une des deux classes apprises, à savoir “personne” ou “non personne”.

Jiang et al [11] ont proposé d'améliorer le système proposé par Dalal et Triggs en fusionnant les histogrammes de gradient orienté avec les histogrammes de couleurs (figure 1.14). Dans le but de surmonter les problèmes d'occultations partielles, ils ont proposé d'analyser séparément la partie supérieure et la partie inférieure du corps humain. Dans leurs travaux, Wang et al [79] combinent les descripteurs HOG et LBP pour avoir une description locale basée sur les contours et la texture des parties du corps humain. Ainsi ils obtiennent un système de suivi robuste aux occultations qui peut être combiné aux algorithmes *Mean-shift* et SVM afin d'explorer les images de la vidéo pour y localiser les personnes.



FIGURE 1.14 – Système de suivi de personnes basé sur la fusion des histogrammes de gradient orienté et les histogrammes de couleurs [11]. Exemple de suivi de personnes dans trois images successives prise par une caméra de surveillance. En rouge, les personnes détectées et le nombre affiché au dessus de la tête de chaque personne correspond à son identifiant.

Le système de suivi proposé par Li et al [80] est composé de deux étapes, la détection et le suivi. La détection de la tête et des épaules est assurée par la méthode de classification Adaboost sur des descripteurs HOG. Chaque zone détectée est suivie par un filtre particulaire utilisant des descripteurs locaux modélisant l'apparence des cibles et offrant une robustesse dans les environnements encombrés.

Enfin, en projetant ces applications de suivi de personnes sur notre sujet de suivi de nageurs, nous pouvons remarquer certaines ressemblances, notamment, la forme

globale de la cible à suivre et les difficultés liées aux occultations et aux déformations. En revanche, nous constatons une grande différence en termes de profils de mouvement et d'environnement de travail. En conclusion, nous pouvons tirer profit de quelques idées comme : les techniques adaptées au suivi, la fusion des différentes techniques de suivi et la décomposition de la cible en plusieurs zones.

1.9.2 Suivi de visages

La détection et le suivi de visages sont souvent associés aux applications de surveillance, de contrôle d'accès et de reconnaissance biométrique faciale. Plusieurs études ont été publiées dans la littérature concernant ce domaine [16, 23, 76, 82–84]. La méthode la plus répandue dans le domaine de la détection de visages et celle proposée par Viola et Jones [76] qui se base sur les caractéristiques pseudo-Haar et le classifieur Adaboost. Cette méthode nécessite un apprentissage des classes “visage” et “non visage”. Le classifieur permet de sélectionner les caractéristiques utiles et de séparer les exemples positifs “visage” des exemples négatifs “non visage” par cascade de décision.

Katz et al [7] ont proposé un système de détection de chute des personnes âgées basé sur le suivi de visage et l'analyse du mouvement. Ces informations leur permettent de détecter la chute et ainsi limiter les accidents. Comme le montre la figure 1.15 leur système de suivi de visage se décompose en trois modules : l'initialisation par l'algorithme de Viola et Jones de détection du visage. Puis, un module de suivi basé sur la technique de corrélation NL-JTC. Et pour finir, un module de mesure de confiance qui consiste en la comparaison de l'histogramme de couleurs de la cible détectée avec celui de la référence. La mesure χ^2 est utilisée pour la comparaison de ces histogrammes de couleurs et la génération d'une valeur de confiance. Celle-ci est comparée avec un seuil pour valider le suivi ou revenir à la phase d'initialisation dans le cas d'un décrochage du suivi.

Junxiang et al [16] ont proposé une méthode de suivi de visage basé sur le filtre particulaire où le visage à suivre est défini par un histogramme de couleur. Sachant que la couleur de la peau humaine est différente de la couleur de la majorité des objets, Kuchi et al [85] propose de modéliser la couleur de la peau par une gaussienne dans le but de détecter le visage. Dans leurs travaux, Lee et al [82] extraient le visage de l'arrière-plan par la méthode de différentiation d'image, où la soustraction de deux images successives permet d'extraire l'objet en mouvement i.e. le visage.

D'autres chercheurs ont proposé des systèmes de suivi multimodaux. Par exemple, Chuan Xu et al [23] ont développé un système de suivi basé sur les descripteurs LBP, qui décrivent la texture, et l'information de couleur de la peau. Vadakkepat et al [83] ont proposé un système de détection et de suivi pour les robots. Ce système se base sur les réseaux de neurones, la détection de peau et de visage ainsi que sur l'algorithme

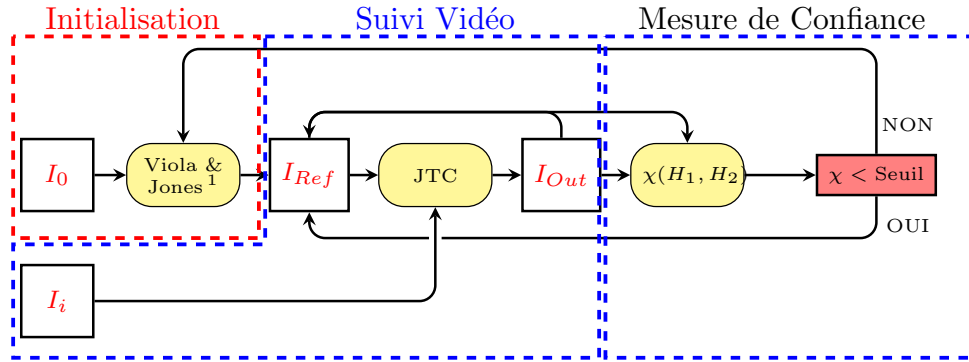


FIGURE 1.15 – Système de suivi et détection de chutes des personnes âgées basé sur l’algorithme de Viola and Jones pour la détection de visage, NL-JTC pour le suivi et l’histogramme de couleurs pour l’évaluation du suivi.

Mean-shift pour le suivi. Spors et Rabenstein [84] ont utilisé l’Analyse en Composantes Principales pour localiser les yeux dans des régions de peau, détectées par segmentation. A partir de cette localisation, le visage est détecté puis suivi en utilisant un filtre de Kalman.

Notons que les principes généraux de ces approches peuvent être utilisés dans le cadre du suivi des nageurs. Pour ce faire, nous pouvons commencer par détecter la référence initiale (tête du nageur) par des méthodes par apprentissage. Ensuite, nous pourrions suivre cette cible en utilisant des technique basées sur la corrélation NL-JTC, sur l’histogramme de couleurs, etc. Enfin, nous pourrions fusionner ces informations pour évaluer les performances des nageurs.

1.9.3 Suivi de véhicules

Les systèmes de suivi de véhicules offrent des données brutes essentielles pour une large gamme d’applications telles que, la gestion et le contrôle du trafic, le routage, la navigation et la sécurité routière. Etant donné le grand intérêt de ce domaine beaucoup de recherches ont été faites pour améliorer ces systèmes [77, 78, 86–89].

Dans ce type d’applications, les caméras utilisées sont souvent statiques, ce qui favorise les techniques de soustraction de fond connues pour leur simplicité et leur rapidité. Des variantes de ces techniques ont été proposées dans la littérature. Magee [77] a proposé de modéliser l’arrière-plan par un modèle de mélange de gaussiennes afin d’extraire les objets en mouvement, i.e. les véhicules. Un filtre de Kalman permet ensuite de prédire leur prochain emplacement. Dailey et al [86], quant à eux, utilisent la différenciation des images sur la vidéo, pour extraire le fond, puis appliquent le détecteur

de contour de Sobel [90] pour localiser les véhicules plus précisément.

Long et al [87] proposent un système qui combine la soustraction de fond avec la comparaison des histogrammes de couleurs basée sur la mesure de Bhattacharyya. Leur objectif est d'enlever l'interférence entre l'arrière-plan et la densité de probabilité de couleur des véhicules, comme le montre la figure 1.16. Avidan [88] a proposé un système, appelé SVT (*Support Vector Tracking*), assurant le suivi des véhicules par classifieur SVM.

Enfin, nous pouvons noter que ces approches de suivi de véhicules utilisent généralement des systèmes de prise de vue statique. En particulier, nous constatons l'utilisation conjointe de la soustraction de fond et des histogrammes de couleurs permettant de gagner en robustesse. Cette idée est très intéressante pour notre cas de suivi de nageurs car elle peut permettre d'appliquer des prétraitements sur l'image dans le but de faciliter et d'optimiser le suivi.

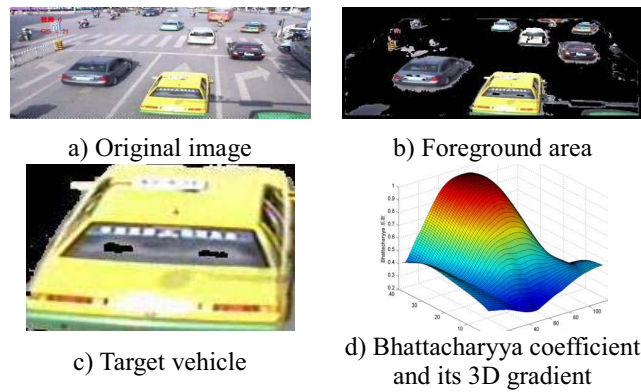


FIGURE 1.16 – *Système de suivi des véhicules. La première étape est la soustraction de fond (b) de l'image originale (a). Ensuite, une comparaison des histogrammes de couleurs basée sur le coefficient de Bhattacharyya (d) est faite pour détecter la cible (c).*

1.9.4 Discussion

La présentation de ses systèmes de suivi permet d'avoir une vision globale de l'enchaînement des différents modules constituant un système de suivi complet. Ainsi, nous nous inspirerons des différentes optimisations proposées afin de développer un système de suivi de nageur performant. Celui-ci pourra contenir un module de prétraitement permettant de restreindre la zone de recherche, un module de suivi qui pourra être optimisé par une fusion de descripteurs et un module de prédiction de mouvement.

1.10 Applications relatives à la natation

Etant donnée la popularité de la natation ainsi que la forte compétitivité entre les nations, un grand intérêt est donné à l'étude de leurs performances. Dans ce cadre, plusieurs études ont été menées afin d'évaluer les performances au cours des entraînements ou des compétitions. Dans cette section, nous allons présenter, un ensemble d'applications de suivi relatives à ce domaine.

1.10.1 Analyse des performances des nageurs à partir de vidéos

L'objectif de ces expérimentations est de faire une étude cinématique et dynamique du mouvement des nageurs dans le but d'aider les entraîneurs [24,91,92]. Ainsi, ceux-ci peuvent avoir des statistiques précises, sur les nageurs qu'ils accompagnent, permettant d'analyser leurs techniques de nages, de détecter leurs points faibles, et ainsi d'améliorer leur niveau. Les études existantes évaluent, le plus souvent, la vitesse, les techniques de départ (plongeon), la reprise de nage, le taux de frappes des mains, le retournement utilisé dans le virage (*culbute*), le cycle de respiration, etc.

Pogalin et al [29] ont développé un système d'enregistrement de vidéos, utilisable pendant les séances d'entraînement, afin d'aider les entraîneurs à évaluer quantitativement et qualitativement les nageurs. Le système proposé utilise plusieurs caméras équidistantes permettant de couvrir toute la piscine. Les vidéos acquises par ces caméras doivent être rectifiées en corrigeant la distorsion introduite par les lentilles. Les images des vidéos sont recalées afin d'avoir une vidéo panoramique qui couvre toute la piscine puis une transformation de perspectives est faite pour régulariser la géométrie. Ensuite, pour suivre les nageurs ils appliquent deux méthodes : la détection du nageur par une méthode de soustraction de fond, utilisant la moyenne des images de la vidéo pour sa modélisation, et une méthode de segmentation basée sur la définition d'un intervalle de couleur de la peau composée principalement de rouge. Ce système a donné des résultats intéressants sur des séquences vidéo de très courtes durées. Cependant, deux limitations majeures apparaissent. Tout d'abord la nécessité d'une intervention manuelle de l'utilisateur puis la lourdeur du dispositif de prise de vue à mettre en place qui est non adaptable aux compétitions.

1.10.2 Projet Poséidon pour la détection de noyade

Le projet Poséidon est un système, intelligent et autonome, de détection de noyade développé par Eyeview Systems, entreprise Néerlandaise spécialisée dans les systèmes de sécurité et de santé pour les piscines. Le projet Poséidon est installé dans plus de



FIGURE 1.17 – *Système d'enregistrement des vidéos d'entraînement. Trois caméras sont placées au bord du bassin sur des trépieds afin de permettre l'obtention d'une vue de dessus qui assure une meilleure analyse de performances.*

120 piscines en Europe, Amérique du nord, Japon et Australie. Il est conçu pour aider les sauveteurs à détecter la noyade des nageurs débutants, les localiser en temps réel pour ainsi les secourir le plus rapidement possible. Ce système est équipé d'un réseau de caméras étanches installées au fond de la piscine permettant de surveiller et d'analyser en permanence les activités des nageurs (principalement au fond de la piscine). La technique de surveillance est basée sur la reconnaissance de texture, le volume et le mouvement du nageur. Dans le cas où le système détecte une absence de mouvement d'un nageur sous la surface de l'eau pendant quelques secondes, il le considère en état critique et une alarme est déclenchée afin que les sauveteurs interviennent [93].

Cependant, des différences majeures existent entre le projet Poséidon et le suivi des nageurs en compétitions. En effet, dans cette approche, le suivi du nageur est effectué grâce à plusieurs caméras sous la surface de l'eau et dans un environnement calme contrôlé et sans éclaboussures. Il est à noter que ce type de système de prise de vue est interdit dans les compétitions. Finalement ce système se base sur le suivi de la personne de manière globale afin de détecter l'absence de mouvement correspondant à une situation de noyade. Pour ces différentes raisons, l'approche du projet Poséidon ne permet pas le suivi des nageurs pendant les compétitions avec une seule caméra de manière instantanée et robuste.



FIGURE 1.18 – *Projet Poséidon pour la détection de noyade. Caméras étanches placées dans le fond du bassin permettant la détection des personnes qui dépassent un certain niveau de sécurité (noyade).*

1.11 Conclusion

Dans ce chapitre, nous avons présenté un aperçu général de l'état de l'art des systèmes de suivi. Pour cela, nous avons détaillé les différents modules en suivant le schéma synoptique global d'un système de suivi classique. Nous avons commencé par une présentation des techniques de prétraitements pouvant améliorer le suivi spécialement dans le milieu aquatique. Puis, nous avons détaillé les méthodes de détection connues en les regroupant dans deux ensembles : méthodes basées sur la corrélation et méthodes numériques. Etant donnée l'expansion des systèmes basés sur la fusion multi-descripteurs, nous avons consacré une partie pour présenter l'objectif sous-jacent à l'utilisation de la fusion ainsi que son principe et ses différents types. Ensuite, nous avons exposé des systèmes de suivi utilisant les techniques détaillées dans les parties précédentes. Enfin, nous avons conclu notre chapitre par une présentation des applications relatives au domaine de la natation qui est le cadre global de notre thèse. Dans le but de proposer un système de suivi robuste et efficace pour les nageurs, nous proposons dans le chapitre suivant un ensemble de prétraitements adaptés au milieu aquatique facilitant le suivi.

Chapitre 2

Prétraitements adaptés aux vidéos du milieu aquatique : Analyse, choix et validation

Sommaire

2.1	Introduction	46
2.2	Spécification du milieu	46
2.3	Segmentation du bassin	48
2.3.1	Etudes basées sur les espaces couleur	48
2.3.2	Etude basée sur la détection des lignes	52
2.3.3	Discussion	53
2.4	Calibrage	53
2.4.1	Objectifs du calibrage	54
2.4.2	DLT pour le calibrage du bassin	55
2.4.3	Correction de perspective	56
2.5	Localisation de la région d'intérêt	56
2.5.1	Extraction du couloir	57
2.5.2	Différence d'images	59
2.5.3	Décomposition en blocs	60
2.5.4	Seuillage et classification	61
2.5.5	Elimination des faux blocs	61
2.5.6	Détection de mouvement du nageur	62
2.6	Initialisation du suivi par l'approche <i>Scaled Composite JTC</i>	63
2.6.1	Génération de la base de données	63
2.6.2	Application de la technique NL-JTC	65
2.6.3	Référence composite re-dimensionnée	66
2.6.4	Pré-localisation de la région d'intérêt	67

2.6.5	Décision basée sur PCE	68
2.7	Conclusion	69

2.1 Introduction

L'objectif de ce chapitre est de mettre en place et de valider un ensemble de prétraitements nécessaires à l'optimisation du suivi des nageurs. En particulier, nous cherchons à localiser le couloir de nage, restreindre la zone d'intérêt et détecter automatiquement le nageur. Ainsi, il nous sera possible d'initialiser le processus de suivi avec une image de référence de la tête du nageur que nous aurons détectée automatiquement. D'autre part, la localisation de la région d'intérêt nous permettra de pré-localiser le nageur dès son premier mouvement dans le bassin et de restreindre la région d'intérêt autour de la cible à suivre tout au long de la course.

2.2 Spécification du milieu

La Fédération Internationale de la Natation (FINA) a établi des normes pour les piscines de compétition. Pour les courses longues la longueur du bassin est de 50 m tandis qu'il ne fait que 25 m pour les courses courtes [94]. Les piscines de compétition sont généralement couvertes et chauffées afin d'assurer leur utilisation durant toute l'année, et pour se conformer plus facilement aux règlements de la FINA concernant la température, l'éclairage et l'équipement de l'arbitrage automatique. Une piscine olympique doit absolument répondre à toutes les normes exigées par la FINA pour les Jeux Olympiques et pour les grands événements comme les championnats du monde et les championnats nationaux. Le tableau 2.1 résume les normes imposées par la FINA pour les grands bassins olympiques.

Propriété	Valeur conventionnelle
Longueur	50 m
Largeur	25 m
Profondeur	2 au minimum (3 m recommandé)
Nombre de couloirs	10 (de 2,5 m de largeur chacun)
Température	25-28 °C
Éclairage	1500 lux

TABLE 2.1 – les normes de la FINA pour l'élégibilité des piscines olympiques de longues courses.

Afin qu'un bassin de 50 m soit éligible pour organiser des grands événements, il est indispensable qu'il fasse 50 m de longueur par 25 m de largeur, divisé en huit couloirs de 2,5 m de largeur chacun. Deux autres couloirs supplémentaires de 2,5 m de chaque côté du bassin (couloirs 0 et 9) ont été ajoutés aux 8 couloirs traditionnels lors du congrès de la FINA de 2009 [94]. D'autre part, la profondeur du bassin n'est pas fixe mais elle est limitée à 2 m au minimum. D'autres critères sont aussi réglementés par la FINA, par exemple la couleur de la ligne de couloir ou le positionnement des drapeaux indicateurs de retournement pour la course de dos (5 m depuis chaque bord). La température de l'eau doit être maintenue à 25-28 °C et le niveau d'éclairage à plus de 1500 lux. Des *Touchpads* sont montés sur les deux bords de bassins afin de mesurer automatiquement les temps de passages et d'arrivées des athlètes aux extrémités. La figure 2.1 résume les réglementations dimensionnelles imposées par la FINA.

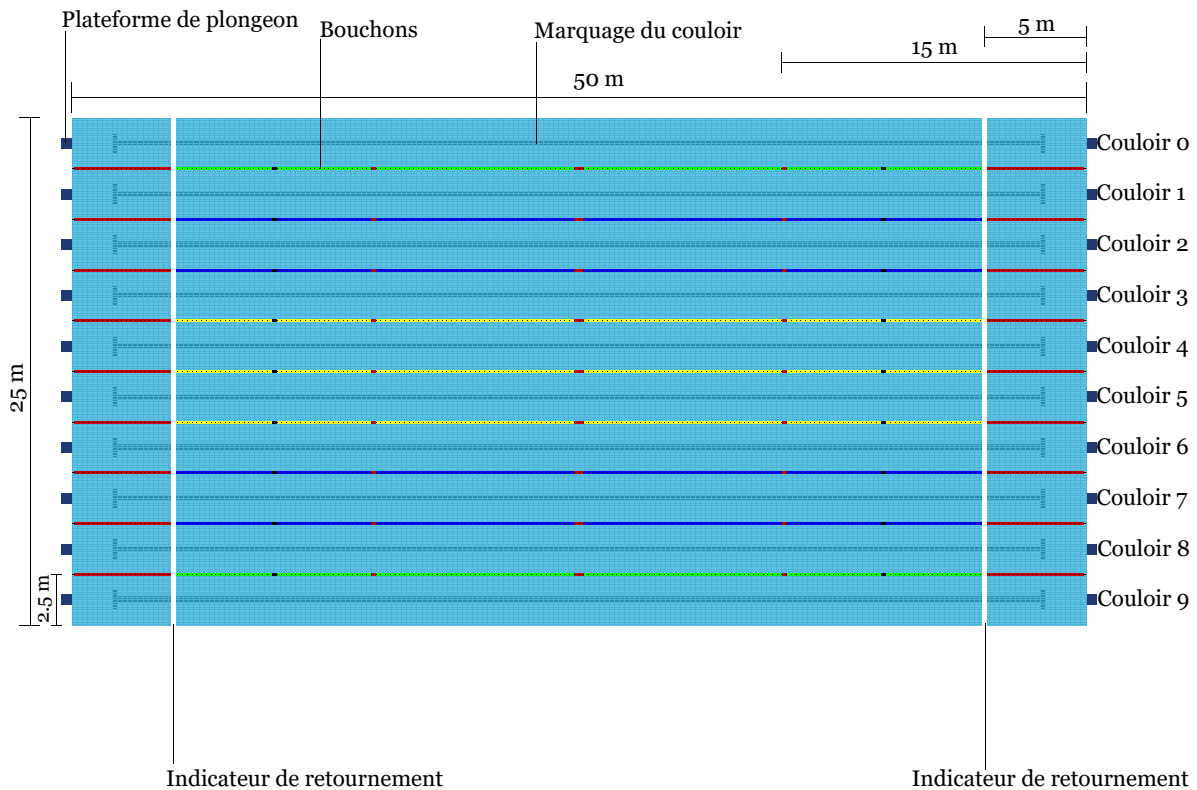


FIGURE 2.1 – Mesures conventionnelles des bassins olympiques reconnus par la FINA pour les longues courses.

Ces réglementations assurent le bon déroulement des compétitions, mais nous servent aussi de repères référentiels. Ces repères permettent de calibrer le bassin en calculant la relation géométrique entre les coordonnées métriques connues des points repères dans le bassin et les coordonnées pixels correspondant aux mêmes points repères. Une fois la vidéo calibrée, il est possible de passer simplement des coordonnées métriques réelles

aux coordonnées pixels dans l'image et vice versa. Par conséquent, nous serons capables de calculer différentes mesures :

- La Position du nageur détecté dans l'image en termes de coordonnées métriques.
- La distance parcourue durant un temps déterminé.
- La vitesse instantanée et moyenne du nageur.
- Les temps de passage par les différents indicateurs réglementés dans le bassin.
- L'estimation de la taille des nageurs dans les différents couloirs de l'image.

Notons que les réglementations concernant l'éclairage, les couleurs de lignes et la couleur de fond peuvent être utiles pour la segmentation du bassin. En particulier, nous utiliserons ces différentes données pour extraire des informations sur le bassin telles que sa taille, les couloirs de nages et les lignes d'eau.

2.3 Segmentation du bassin

Dans cette partie, nous analysons les vidéos de compétitions dans le but d'en extraire des informations relatives à l'environnement aquatique (i.e. le bassin) susceptibles de nous aider à suivre le nageur et calculer sa vitesse. En d'autres termes, nous cherchons à segmenter le bassin et à reconnaître les différentes régions de la scène, à savoir : les bords du bassin, les couloirs, l'eau, la peau, etc. Pour cela, nous avons mené plusieurs recherches basées sur des prétraitements des images de bassins détaillés dans le chapitre. Dans cette section, nous présentons des études basées sur la nature des espaces couleur, sur la détection de peau et sur la détection des lignes.

2.3.1 Etudes basées sur les espaces couleur

L'espace couleur est une organisation spécifique qui permet de créer, décrire et visualiser les couleurs. Dans le domaine de la vision par ordinateur, la couleur est souvent représentée par des triplets. A titre d'exemple, nous citons les espaces RGB et HSV qui représentent les espaces couleur les plus connus en traitement d'images. Par la suite, nous étudierons la possibilité de segmenter des images de bassins selon ces deux espaces.

2.3.1.1 Espace RGB

RGB est un espace additif tri-chromatique composé de trois composantes : rouge, verte et bleue. Il est utilisé par plusieurs dispositifs : ordinateurs, télévisions, caméras, etc. Les systèmes de prise de vue que nous utilisons dans cette thèse sont basés sur l'espace RGB. En utilisant cette représentation, chaque image de la séquence vidéo se

compose de trois matrices où chacune contient des valeurs codées sur 8 bits représentant l'intensité de la couleur concernée (rouge, verte, ou bleue).

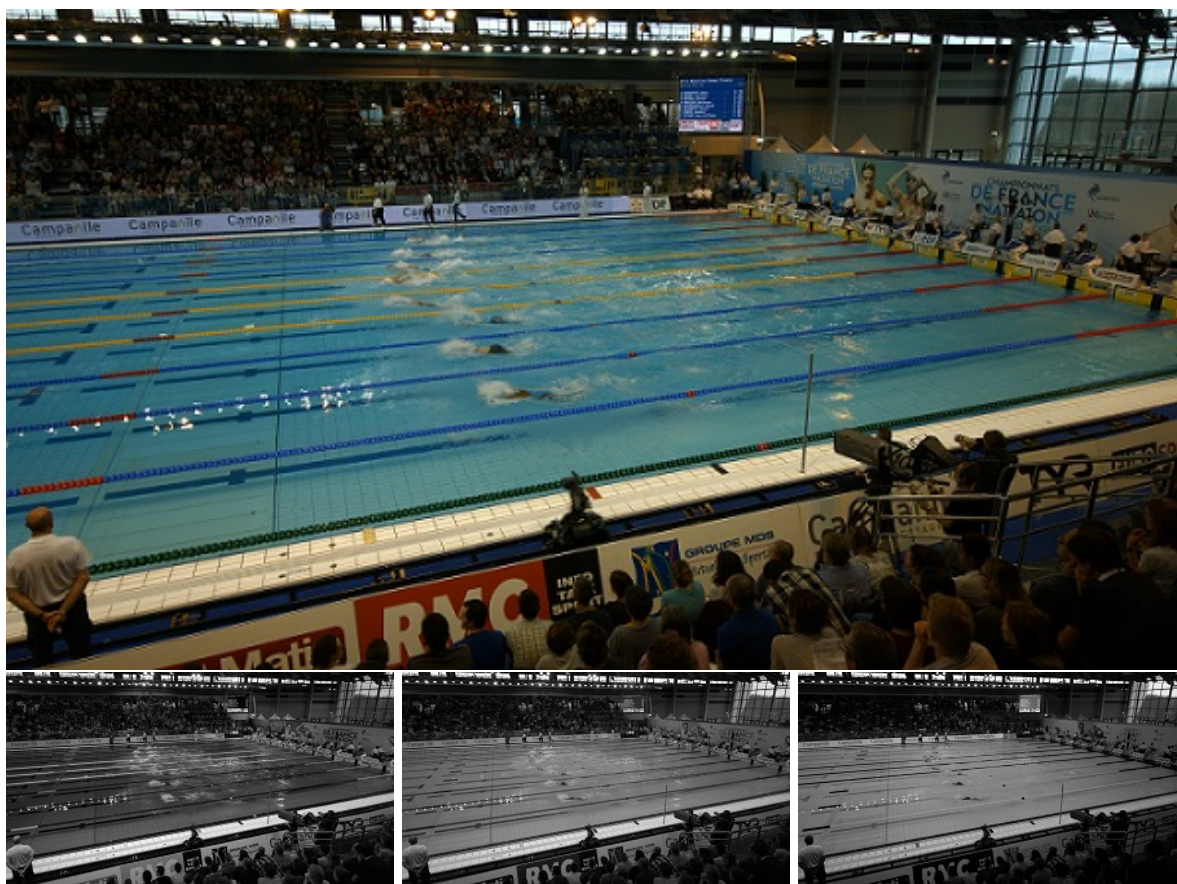


FIGURE 2.2 – Image d'un bassin de compétition affichée dans l'espace RGB. Les images en bas correspondent, de gauche à droite, aux composantes rouge, verte et bleue (représentées en niveaux de gris). Championnats de France de Natation - Limoges, Avril 2015

La figure 2.2 présente un exemple d'une image de la moitié du bassin de 50 mètres représentée dans l'espace couleur RGB. Les trois images en bas, présentent, de gauche à droite, les composantes rouge, verte et bleue (représentées en niveaux de gris). Dans la première composante, nous constatons que les hautes intensités (i.e. le blanc) correspondent aux bouchons rouges et à tous les objets contenant une couleur rouge ainsi qu'à la peau des nageurs qui contient aussi une forte composante rouge. Même constat concernant les composantes verte et bleue, où les hautes intensités correspondent aux objets de même couleur (respectivement vert et bleu). En particulier, l'image de la composante bleue est très intéressante car elle permet de faire ressortir le contraste entre l'eau, qui contient une forte intensité de bleu, et le reste du bassin. Ce constat sera pris en compte dans les chapitres suivants pour nos approches de suivi basées sur

CHAPITRE 2. PRÉTRAITEMENTS ADAPTÉS AUX VIDÉOS DU MILIEU AQUATIQUE : ANALYSE, CHOIX ET VALIDATION

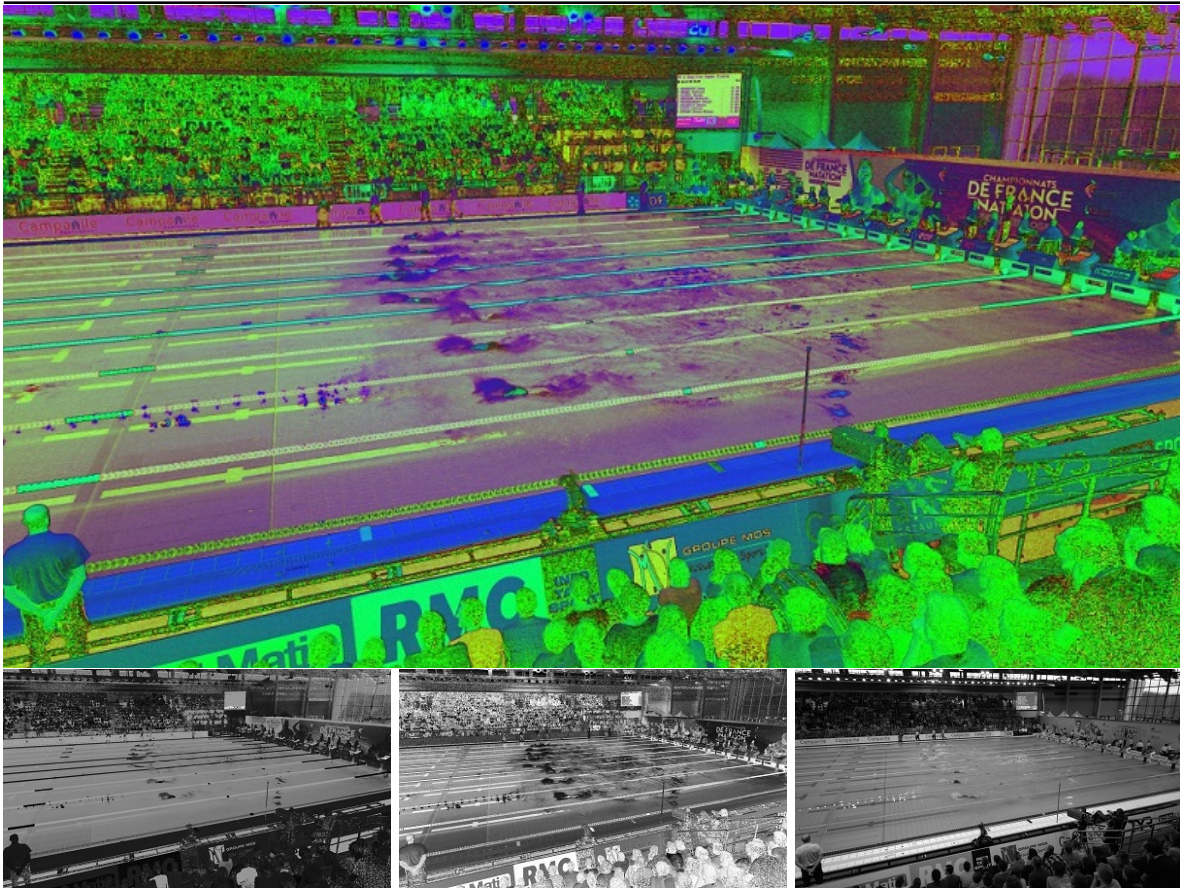


FIGURE 2.3 – Image d'un bassin de compétition affichée dans l'espace HSV. Les images du bas correspondent, de gauche à droite, à la composante Teinte, Saturation et Valeur. Championnats de France de Natation - Limoges, Avril 2015

l'histogramme de couleur afin d'augmenter le pouvoir discriminant de ces approches.

2.3.1.2 Espace HSV

L'espace couleur HSV consiste en une transformation non-linéaire de l'espace RGB dans le but de le rendre plus intuitif et plus pertinent pour des applications spécifiques, notamment dans le domaine de la vision par ordinateur. Cet espace est composé d'un triplet : *Hue* qui représente la teinte, *Saturation* qui correspond à l'intensité de la couleur et *Value* qui représente la description de la brillance.

L'image principale de la figure 2.3 correspond à une transformation de l'image visible sur la figure 2.2 de l'espace RGB vers l'espace HSV. La décomposition des composantes de l'image en HSV génère les trois images affichées en dessous. De gauche à droite, l'image de la composante *H* présente une description robuste des couleurs en termes

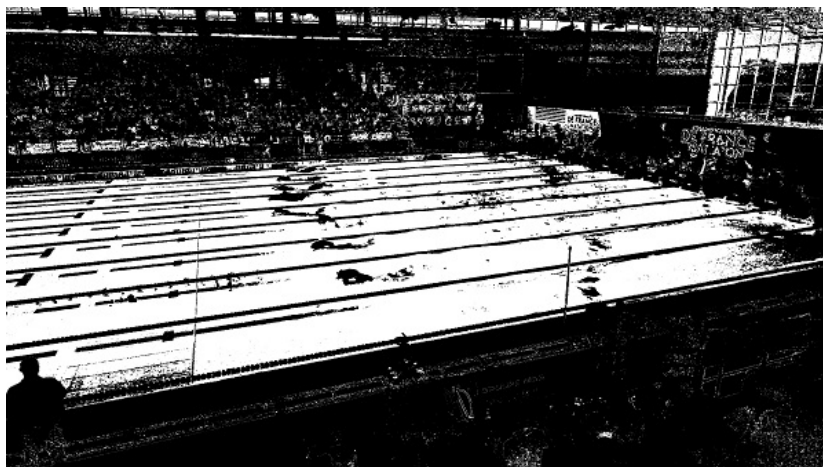


FIGURE 2.4 – *Segmentation de l'eau par seuillage sur l'image de la composante Teinte de la figure 2.3. Championnats de France de Natation - Limoges, Avril 2015*

de teintes. Ceci permet, entre autre, d'éliminer les effets de reflets et d'éclaboussure et donne ainsi des couleurs homogènes aux différents objets. L'image de la composante S apporte l'information d'intensité à la première image H . La dernière composante V décrit la brillance et permet de mettre en évidence les reflets et les éclaboussures. Dans notre cas, nous nous intéressons particulièrement à la composante H pour son homogénéité. En effet, elle permet de faire ressortir l'eau et les différents objets présents dans le bassin. Finalement, ceci nous permet, comme le montre la figure 2.4, d'extraire les couloirs en appliquant un seuillage sur les valeurs de cette composante.

2.3.1.3 Détection de peau

La détection de peau représente un prétraitement important qui peut faciliter la tâche de pré-localisation du nageur. La seule caractéristique qui permet de différencier la peau du reste de la scène est sa couleur. Par conséquent, nous avons déterminé un intervalle de valeurs correspondant à la peau sur les espaces couleur RGB, HSV et sur leurs différentes composantes. Ceci a été fait après des tests réalisés sur des séquences vidéo de compétitions. En effet, nous nous intéressons particulièrement à la segmentation appliquée sur la composante R du modèle RGB car elle permet une bonne extraction de la peau qui contient une forte intensité de rouge. Cependant, d'autres objets peuvent apparaître parmi les détections potentielles de la peau car ils possèdent des valeurs de rouge qui sont inclus dans l'intervalle sélectionné. La figure 2.5 présente un exemple de détection de la peau en n'utilisant que la composante R . Dans cette figure, nous constatons que de grandes parties de la peau ne sont pas détectées à cause des éclaboussures. D'autre part, nous notons que des bouchons rouges sont détectés par erreur. Cependant, ces derniers peuvent être ignorés grâce à leur appartenance aux

lignes des couloirs. Pour cela, nous présentons par la suite une méthode de localisation précise de ces lignes d'eau.

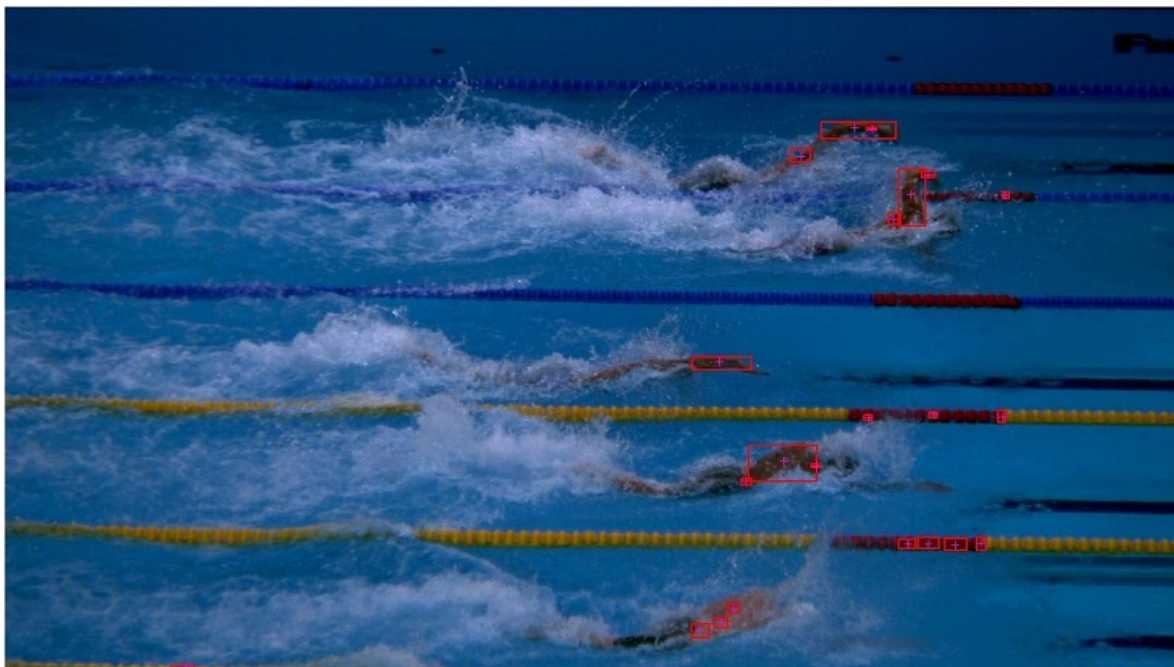


FIGURE 2.5 – *Détection de peau utilisant un intervalle de valeurs sur l'image issue de la composante rouge. Ici, les rectangles rouge représentent les zones de peau localisées. Championnats du monde de natation - Barcelone, Juillet 2013*

2.3.2 Etude basée sur la détection des lignes

Dans le but de restreindre la zone d'intérêt et de pré-localiser le nageur, nous allons tirer profit des caractéristiques géométriques du bassin présentées dans la section 2.2 afin de détecter les lignes délimitant les couloirs de nage. Pour cela, nous appliquons la technique de Hough [53, 54] que nous avons introduite dans la section 1.4.3.3 du chapitre 1. Pour cela, nous commençons, d'abord par une binarisation de l'image en utilisant le filtre Canny pour ne garder que l'information de contour. Le résultat de l'application de ce filtre est affiché dans la figure 2.6.a. Sur cette image binaire de contour, nous appliquons le détecteur de lignes Hough afin de détecter toutes les lignes présentes dans l'image, comme le montre la figure 2.6.b. Ensuite, il est nécessaire de les classifier pour éliminer celles qui ne correspondent pas aux lignes d'eau délimitant les couloirs de nages. Pour cela nous distinguons deux types de lignes détectées selon leur allure : les lignes qui sont parallèles aux lignes de couloirs et celles qui ne le sont pas. Ces dernières sont faciles à détecter et à éliminer. En revanche, il est difficile de déterminer les lignes qui sont parallèles aux lignes de couloirs et qui correspondent à

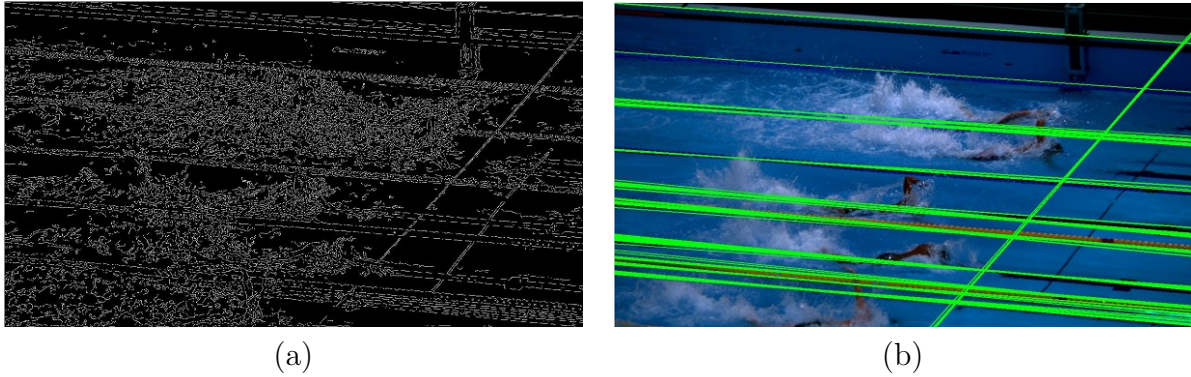


FIGURE 2.6 – Détection de lignes basée sur la technique de Hough. (a) Représente le résultat de la binarisation et du filtre de Canny. (b) Représente le résultat de la détection de ligne par la méthode de Hough avec en vert les détections obtenues. On peut noter de fausses détections causées par les rappels de lignes d’eau situés au fond de la piscine. Championnats du monde de natation - Barcelone, Juillet 2013

autre chose, par exemple, aux indicateurs de couloirs situés au fond du bassin (lignes, généralement noires, situées sous les bouchons).

2.3.3 Discussion

L’objectif principal de cette section est de segmenter les images dans le but d’extraire les différents objets existant dans le bassin. Ceci doit permettre de préparer la séquence vidéo pour des traitements ultérieurs qui consistent à détecter, suivre et évaluer précisément les performances du nageur. Les prétraitements présentés dans cette partie ouvrent des pistes très intéressantes que nous utiliserons tout au long de notre thèse. En revanche, ils ne sont pas concluants en termes de précision et d’adaptation aux différentes situations rencontrées sur le terrain. Pour cette raison, nous proposons une nouvelle solution qui consiste au calibrage du bassin dans le but de le segmenter et d’effectuer des mesures précises dans l’image. Ce choix est argumenté par la sensibilité et l’importance des mesures que nous devons effectuer pour estimer précisément les performances des nageurs.

2.4 Calibrage

En analysant le cadre de travail de la thèse, i.e. la natation, nous avons ciblé différentes difficultés liées aux spécificités du milieu aquatique. En particulier, la localisation des couloirs de nage ainsi que l’extraction des lignes d’eau est une tâche

difficile qui reste néanmoins cruciale au bon fonctionnement d'une technique de suivi. En effet, ces informations sont nécessaires pour prédire la future localisation du nageur ou encore pour estimer les distances parcourues, nécessaires au calcul de la vitesse. Afin de contourner les limitations des algorithmes présentés précédemment, nous proposons de calibrer le bassin afin d'établir le lien entre les coordonnées métriques et pixels issues respectivement du bassin et de l'image. Pour cela, nous considérons le plan d'eau comme un espace 2D dans lequel les positions métrique et pixel, de certains points clés, sont connues. Par la suite, nous présentons les objectifs du calibrage et l'application de la technique DLT pour calibrer le bassin en s'inspirant des travaux réalisés par la FFN dans le cadre de la thèse de Marc Elipot [24].

2.4.1 Objectifs du calibrage

Le calibrage peut être utilisé pour corriger les déformations issues des erreurs de perspectives (figure 2.7.b) et des aberrations correspondant aux lentilles de la caméra (figure 2.7.c). A cause de ces distorsions, les pixels de l'image sont mal placés. Cependant, cela n'implique pas toujours que l'information est perdue et qu'elle ne peut pas être reconstruite. Ces problèmes peuvent être résolus en mesurant les paramètres extrinsèques de la caméra. Afin de calculer ces paramètres, nous utilisons la technique de calibrage DLT (Direct Linear Transformation) [24, 25], voir section 1.3 du chapitre précédent. Cette technique nous permet de connaître la relation entre les coordonnées métriques et les coordonnées pixels qui permet à son tour de corriger les perspectives et d'associer une distance (en mètres) à chaque pixel du bassin. Dans notre cas de suivi des nageurs, nous appliquons la technique de calibrage DLT sur le bassin afin de pouvoir extraire le couloir de nage, corriger la perspective, prédire la position et la trajectoire du nageur et calculer les différentes mesures de distance et de vitesse qui permettent d'analyser ses performances.

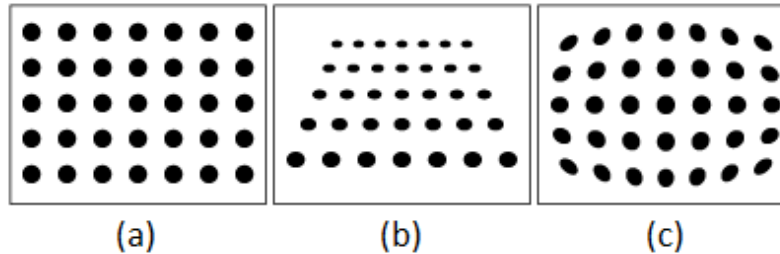


FIGURE 2.7 – Grille de calibrage. (a) Image originale. (b) Image avec une perspective déformée. (c) Distorsion non-linéaire liée aux lentilles.

2.4.2 DLT pour le calibrage du bassin

Les vidéos de compétitions de natation peuvent être calibrées en se basant sur les différents marqueurs du bassin. Etant donné que ce dernier peut être considéré comme un plan en deux dimensions il est possible d'utiliser la technique DLT simplifiée à la place de la DLT classique [24, 25]. Dans celle-ci, l'équation 1.1 de la technique DLT est simplifiée en annulant la troisième coordonnée z et donne l'équation suivante :

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1x_1 & -v_1y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -u_nx_n & -u_ny_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -v_nx_n & -v_ny_n \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_7 \\ L_8 \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_n \\ v_n \end{bmatrix} \quad (2.1)$$

Pour résoudre cette équation, nous avons besoin de calculer les paramètres de calibrage $L_{1..8}$ qui représentent les inconnus de ce système d'équations. Pour cela, il suffit de connaître les coordonnées pixels (x_i, y_i) et les coordonnées métriques (u_i, v_i) de n points de la scène enregistrée avec $n \geq 4$.

Dans notre cas, les piscines olympiques qui organisent les grands évènements se basent généralement sur la norme internationale en termes de dimensions, largeurs des couloirs et couleurs de séparateurs de couloirs. Afin de calibrer le bassin de la figure 2.8, nous sélectionnons manuellement quatre points dont nous avons déjà mesuré avec précision les coordonnées métriques, comme le montre la figure 2.8. Il est à noter que l'origine $(0, 0)$ dans nos calculs correspond au coin haut/droit du bassin. Les valeurs des coordonnées métriques et pixels sélectionnées remplacent respectivement les variables (u_i, v_i) et (x_i, y_i) dans l'équation 2.1. Cela nous donne un système de 8 équations et 8 inconnues $L_{1..8}$. La résolution de ce système permet de calculer les valeurs des paramètres de calibration $L_{1..8}$.

Une fois ces paramètres de calibrage $L_{1..8}$ calculés, le passage des coordonnées pixels aux coordonnées métriques se fait grâce à l'équation suivante :

$$\begin{cases} u = \frac{L_1x + L_2y + L_3}{L_7x + L_8y + 1} \\ v = \frac{L_4x + L_5y + L_6}{L_7x + L_8y + 1} \end{cases} \quad (2.2)$$

Comme le montre la figure 2.8, le calibrage permet de localiser précisément les couloirs de nage ainsi que la position des lignes d'eau.

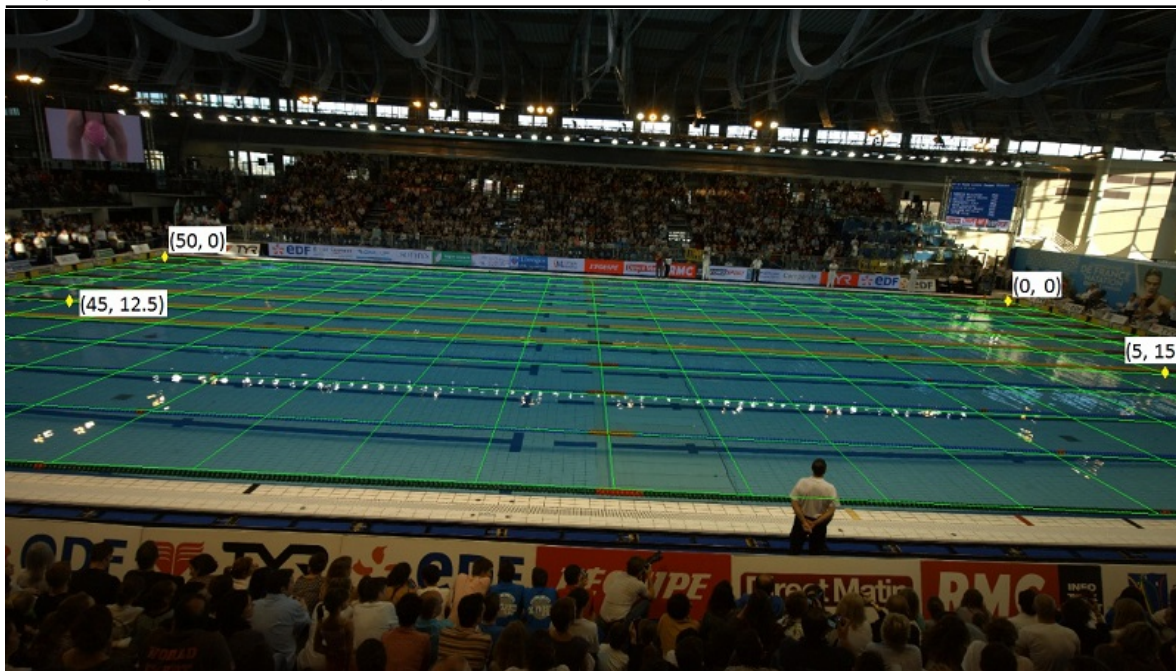


FIGURE 2.8 – Calibrage du bassin par la technique 2D DLT à partir de quatre points. Affichage d'une grille $50 \times 25 \text{ m}^2$ où chaque cellule représente une zone de $2.5 \times 2.5 \text{ m}^2$. Championnats de France de Natation - Limoges, Avril 2015

2.4.3 Correction de perspective

Afin de faire des mesures de distance précises dans l'image, il est important de faire une correction de perspective. Pour ce faire, nous nous appuyons sur les résultats de la technique de calibrage DLT. Ainsi il est possible d'assurer le passage des coordonnées pixels de l'image, qui contient dans notre cas un bassin avec des perspective déformées, aux coordonnées métriques réelles. Finalement, nous pouvons situer précisément le nageur dans le bassin afin de calculer les différentes mesures liées à son avancement. La figure 2.9 montre une image originale prise durant les championnats de France de Natation - Limoges 2015, par une caméra 4K et la reconstruction de cette image après la correction de perspective.

2.5 Localisation de la région d'intérêt

L'objectif de cette partie est de restreindre la région d'intérêt [1] autour du nageur durant toute la séquence vidéo. Cette étape est importante pour préparer la vidéo en vue d'un suivi robuste du nageur permettant d'effectuer des mesures précises nécessaire à l'évaluation de ses performances. Deux résultats sont attendus de ce processus de

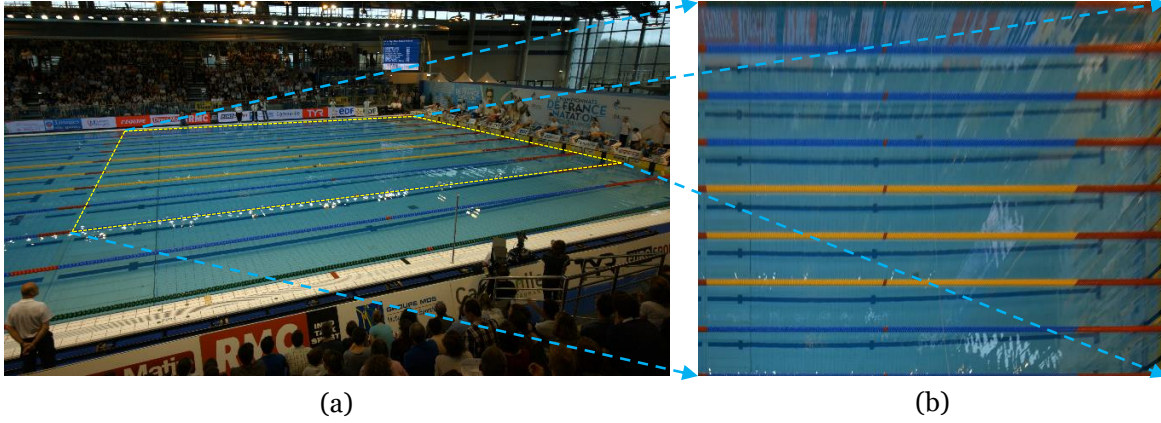


FIGURE 2.9 – Correction de perspective d'une image d'un bassin de compétition. (a) Image originale. (b) Image reconstruite après la correction de perspective. Championnats de France de Natation - Limoges, Avril 2015

localisation de la région d'intérêt, à savoir, l'instant de début de nage qui représente le premier mouvement du nageur dans le couloir de nage et la restriction de la région d'intérêt en pré-localisant le nageur durant la course.

Pour cela, nous proposons un processus automatique basé sur la détection de mouvement [11, 31, 95] du nageur dans le couloir concerné. Pour ce faire, nous utilisons des techniques de prétraitement d'image et nous prenons en compte des aprioris liés à la natation et aux spécificités du bassin. Comme données d'entrée, ce processus nécessite la connaissance des paramètres de calibrage de la caméra, du numéro de couloir ainsi que des dimensions du bassin. Dans le reste de cette section, nous allons détailler les différentes étapes du processus de localisation de la région d'intérêt suivant l'enchaînement présenté dans la figure 2.10.

2.5.1 Extraction du couloir

La connaissance des dimensions du bassin et des couloirs de nage nous permet d'extraire précisément le couloir contenant le nageur concerné. Ce dernier est défini lors de la phase d'initialisation du suivi en sélectionnant le nageur à suivre. L'extraction du couloir concerné se fait selon la formule suivante :

$$\begin{cases} x_1 = 0 \\ x_2 = Longueur \\ y_1 = NumCouloir * Largeur \\ y_2 = (NumCouloir + 1) * Largeur \end{cases} \quad (2.3)$$

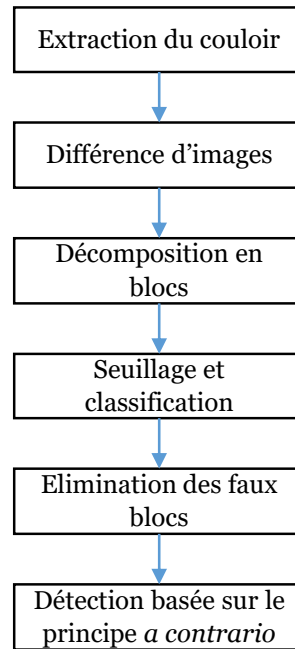


FIGURE 2.10 – *Processus général de localisation de la région d'intérêt.*

avec $NumCouloir$ la position initiale du nageur par rapport à l'axe des y sachant que l'origine $(0, 0)$ correspond au coin haut/droit du bassin, $Longueur = 50$ m la longueur du bassin et $Largeur = 2.5$ m représente la largeur du couloir. x_1, x_2, y_1 et y_2 représentent les coordonnées des deux lignes délimitant le couloir contenant le nageur concerné. Il est à noter que ces mesures sont calculées en mètres et qu'il est donc nécessaire d'obtenir les coordonnées pixels correspondantes. Pour cela, nous tirons parti du calibrage et de la fonction $MetrePixel$ permettant d'assurer ce passage, voir équation 2.2. Ainsi, nous obtenons les huit coordonnées (x_{ij}, y_{ij}) des quatre points délimitant le couloir avec :

$$(x_{ij}, y_{ij}) = MetrePixel(x_i, y_j) \quad (2.4)$$

Finalement, nous extrayons un rectangle contenant les quatre points délimitant le couloir et nous lui appliquons un masque. Le résultat est affiché dans la figure 2.11.



FIGURE 2.11 – *Extraction du couloir et application d'un masque en tenant compte des quatre points : x_1, x_2, x_3 et x_4 . Les coordonnées métriques de ces points sont connues et leurs coordonnées pixels sont calculées en utilisant les résultats du calibrage. Championnats de France de Natation - Limoges, Avril 2015*

2.5.2 Différence d'images

La différence d'images [28,31] est une technique de détection de mouvement utilisée particulièrement sur les vidéos capturées par des caméras qui filment de manière statique (sans déplacement de la caméra), ce qui est notre cas. Une première possibilité est de considérer l'objet en mouvement en tant que premier plan qui sera segmenté et extrait en effectuant une soustraction entre l'image i de la séquence vidéo et l'image de fond. Où cette dernière est une image de la scène sans l'objet à suivre. Dans notre cas, la première image où le couloir est vide, avant le plongeon du nageur, peut être considérée comme une image de fond. Cette technique montre de très bons résultats pour le suivi en cas d'arrière-plan statique. Cependant, elle reste très sensible aux variations d'éclairage et au mouvement des différentes composantes de l'arrière-plan de la scène.

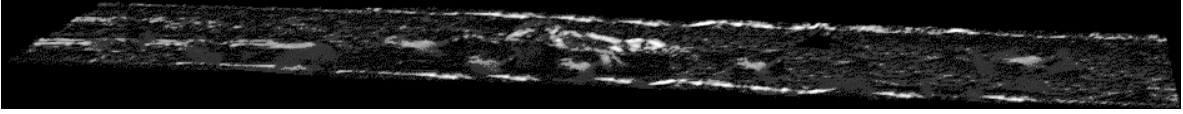


FIGURE 2.12 – *Différence d'images entre l'image i et la première image de la séquence vidéo contenant un couloir vide. Championnats de France de Natation - Limoges, Avril 2015*

Cependant, dans notre cas, le fond n'est pas totalement statique, particulièrement lorsque le nageur commence à nager. Ce dernier génère des éclaboussures et des vagues sur toute la longueur du couloir. Ceci crée un bruit important lors de la soustraction de l'image i contenant un nageur et la première image du couloir vide, comme le montre la figure 2.12. Néanmoins, nous avons remarqué des variations minimales entre les images successives excepté dans la zone contenant le nageur où nous remarquons des variations significatives. Par conséquent, nous optons pour la différence d'images successives afin de détecter le nageur en se basant sur son mouvement. Dans le but de diminuer le bruit dû au mouvement de l'eau, nous proposons d'appliquer un filtre Médian sur les deux images successives avant leur soustraction. L'image de différence est calculée grâce à l'équation suivante :

$$Dif f_i = |Mediane(im_i) - Mediane(im_{i-1})| \quad (2.5)$$

Dans l'image de différence $Dif f_i$, les hautes intensités correspondent principalement à la zone de mouvement du nageur. Ceci est illustré dans la figure 2.13 où nous remarquons que l'image de différence est moins bruitée que celle de la figure 2.12 issue de la différence entre l'image i et la première image du couloir vide. Pour cette raison, nous retenons la différence d'images successives pour le reste de notre étude.

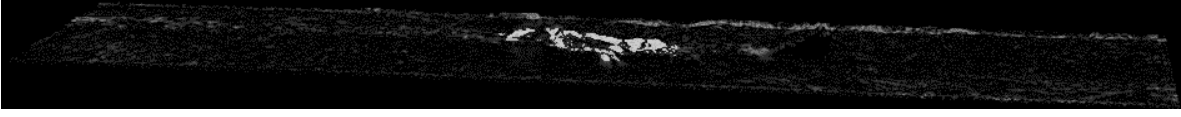


FIGURE 2.13 – *Différence d'images successives $i - 1$ et i . Les hautes intensités correspondent à la zone de mouvement du nageur et de l'eau en mouvement autour de lui. Championnats de France de Natation - Limoges, Avril 2015*

2.5.3 Décomposition en blocs

L'image de différence contient des pixels de différentes intensités. En effet, la valeur de l'intensité correspond au degré de variation de la couleur du pixel. Ce qui nous permet de détecter les objets en mouvement dans la scène. Cependant, dans notre cas, ce n'est pas uniquement le nageur qui est en mouvement mais aussi l'eau et les reflets de lumière l'entourant. Dans ce cadre, nous proposons de décomposer l'image de différence en blocs afin d'étudier localement les intensités et de prendre une décision concernant la détection de mouvement dans chaque bloc du couloir. Nous décomposons donc l'image de différence en blocs de taille $b * b$ tel que :

$$b = \frac{|y_{11} - y_{22}|}{Nbre_{bloc}} \quad (2.6)$$

sachant que la distance $|y_{11} - y_{22}|$ représente la largeur du couloir et $Nbre_{bloc}$ représente le nombre de blocs souhaité, dans notre cas à 20 blocs. Ceci permet de faire une étude locale consistante favorisant l'extraction des zones contenant un mouvement important. Afin de décider si un bloc contient une zone de mouvement, ou non, il est nécessaire d'établir un seuil adapté.

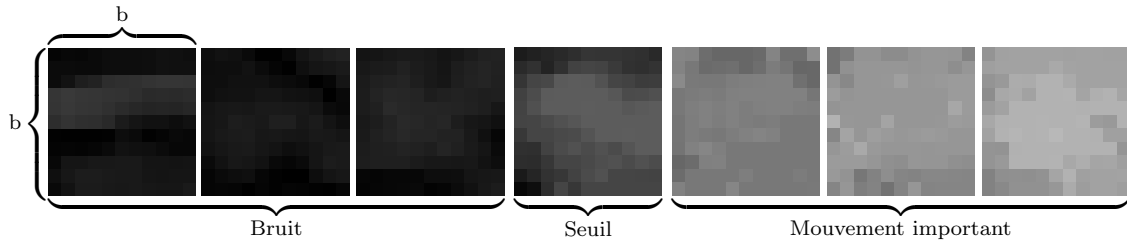


FIGURE 2.14 – *Exemple de classification des blocs selon un modèle de bruit. Le bloc du centre consiste au seuil. Les blocs de gauche sont classifiés comme étant du bruit de fond. Les blocs de droite sont classifiés comme représentant un mouvement important qui sera à analyser pour déterminer sa nature (nageur ou non). Championnats de France de Natation - Limoges, Avril 2015*

2.5.4 Seuillage et classification

Cette étape consiste à appliquer un seuillage sur les blocs afin de les classer (nageur/fond) [2]. Pour chaque bloc de l'image de différence, nous calculons la moyenne locale de ses intensités, que nous comparons à un seuil bien défini ($Seuil_1$). Ce seuil est calculé avant le plongeon du nageur dans le but de mesurer le mouvement de l'eau et la variation liée aux reflets dans un état initial (piscine lisse sans mouvement majeur). Afin de calculer le seuil, nous procédons de la même façon sur deux images successives du couloir vide. Nous calculons la différence entre ces deux images filtrées par un filtre Médian. Ensuite, nous décomposons l'image de différence en blocs et nous calculons la moyenne de chaque bloc. Enfin, le seuil ($Seuil_1$) correspond à la valeur maximale du bruit lié au mouvement aléatoire de l'eau. En d'autres termes, $Seuil_1$ correspond à la valeur maximale des moyennes des blocs mesurée dans l'état de repos sur un couloir vide. Dans ce contexte, la figure 2.14 présente un exemple de classification par rapport au modèle de bruit, où le seuil du bruit correspond au bloc du centre. Les blocs situés à gauche sont considérés comme étant du bruit et correspondent à ceux ayant une moyenne d'intensité inférieure à celle du seuil. En revanche, les blocs à droite correspondent à ceux ayant une moyenne d'intensité supérieure à $Seuil_1$ et ils représentent les zones contenant un mouvement important qui peut correspondre à un nageur. Ces derniers sont localisés et labellisés pour des traitements ultérieurs afin de localiser le nageur.

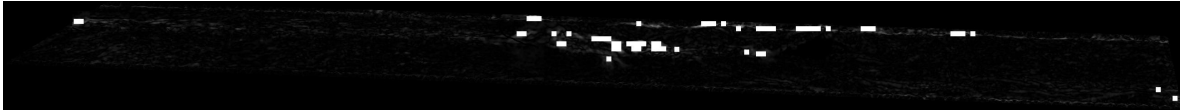


FIGURE 2.15 – *Décomposition en blocs, classification et seuillage. Les blocs blancs correspondent aux zones ayant un mouvement important qui peut correspondre à un nageur. Championnats de France de Natation - Limoges, Avril 2015*

2.5.5 Elimination des faux blocs

Grâce aux étapes précédentes, nous sommes en mesure de détecter les éléments de la scène qui sont en mouvement entre deux images successives. Cependant, nous avons remarqué, après plusieurs tests, que ces zones détectées correspondent au nageur, au mouvement des lignes délimitant le couloir et aux reflets de la lumière. Afin d'éliminer ces deux derniers éléments, nous nous basons sur deux critères principaux : la position et la surface. La position de chaque bloc détecté aide à déterminer sa nature et décider s'il peut correspondre à un nageur. Par exemple, dans la figure 2.15, nous pouvons distinguer les blocs qui correspondent au mouvement des lignes délimitant le couloir d'eau. Afin d'affiner la détection, nous éliminons deux lignes de blocs autour de chaque ligne délimitant le couloir. D'autre part, pour traiter le cas des reflets, nous utilisons un

critère de surface car les blocs représentant des reflets sont généralement détectés sous forme de blocs isolés. L'ensemble des blocs qui reste correspondent au nageur détecté dans l'image i , comme présenté dans la figure 2.16.



FIGURE 2.16 – Affinement des détections en éliminant les faux blocs correspondant aux lignes d'eau et aux reflets. Championnats de France de Natation - Limoges, Avril 2015

2.5.6 Détection de mouvement du nageur

2.5.6.1 Principe *a contrario*

L'approche *a contrario* est une approche statistique basée sur les tests d'hypothèses pour détecter des événements géométriques significatifs dans les images. L'idée de base de cette approche est inspirée du principe de la perception visuelle dit *non-accidentale*, qui est appelé aussi le principe de *Helmholtz* [96, 97]. Dans leur livre, Desolneux et al. [97] résument ce principe comme suit : “à chaque fois qu’une déviation de l’aspect aléatoire apparaît, une structure est perçue”. Ici, la structure est définie par son opposé à savoir le bruit. Dans le cas d’absence de structure les événements sont indépendants et se comportent aléatoirement tandis que la structure se diffère par un comportement plus organisé. La méthode *a contrario* a été appliquée pour différents problèmes de détection. Nous citons par exemple, la détection de contour dans [98, 99], la reconnaissance de formes dans [100] et la détection de points d’intérêt rigides pour le *matching* entre les images dans [101]. Nous proposons, par la suite, d’appliquer le principe *a contrario* pour la détection du mouvement du nageur.

2.5.6.2 Détection de mouvement par la méthode *a contrario* : adaptation

Afin de détecter automatiquement le nageur dès son plongeon dans l’eau, nous avons appliqué le principe *a contrario* qui est proposé dans la littérature [96, 97]. En effet, le modèle de bruit est considéré comme étant une distribution uniforme indépendante sur les éléments considérés. Dans notre cas, nous avons considéré le mouvement aléatoire de l’eau dans le couloir vide comme un modèle de bruit. Pour cette raison, nous avons pu établir un seuil dynamique et pertinent ($Seuil_1$) qui dépend du mouvement de l’eau du couloir concerné. Durant toute la période où l’image de différence $Diff_i$ contient un bruit aléatoire, nous nous considérons dans un état de repos. Une fois qu’un mouvement structuré est apparu entre les deux bords du couloir, la moyenne d’intensité des

blocs concernés va augmenter et dépasser le $Seuil_1$. Ceci nous permet de détecter et localiser automatiquement le nageur et déterminer sa direction. La figure 2.17 présente les différentes étapes du processus de localisation de la région d'intérêt utilisant le principe *a contrario* pour la détection de mouvement, commençant par le modèle de bruit calculé sur le couloir vide jusqu'à la restriction de la région d'intérêt autour du nageur.

2.6 Initialisation du suivi par l'approche *Scaled Composite JTC*

L'approche de localisation de la région d'intérêt introduite dans la section précédente permet de détecter l'apparition d'un nageur dans le couloir ainsi que de restreindre la région d'intérêt autour de lui. Cependant, notre objectif principal consiste à développer un système automatique précis de suivi et d'évaluation des nageurs. Pour cela, nous utilisons le principe *a contrario* afin de détecter le nageur, déterminer l'instant exact de son apparition dans le couloir et le suivre globalement tout au long de la vidéo. Cependant, il reste à traiter l'aspect de la précision du suivi. Pour ce faire, il est nécessaire de tenir compte des différentes difficultés qui résident essentiellement dans la déformation et dans l'occultation du corps du nageur. En effet, la tête du nageur est la partie du corps du nageur qui permet d'obtenir le meilleur compromis entre la visibilité et la rigidité (moins de déformation). Par conséquent, nous proposons dans les prochains chapitres un système optimisé de suivi de nageurs basé principalement sur la tête comme partie du corps à suivre. Ce système nécessite en entrée, la région d'intérêt de l'objet à suivre et une référence de celui-ci. La première donnée d'entrée peut être retrouvée en appliquant l'approche de localisation de la région d'intérêt basée sur le principe *a contrario*. Tandis que la deuxième fera l'objet de cette section où nous proposons une technique de détection automatique de la tête du nageur en se basant sur la technique de corrélation NL-JTC [15, 57] appliquée sur une référence composite mise à l'échelle construite à partir d'une base de données pré-générée. Enfin, la cible détectée sera utilisée comme référence pour le système de suivi optimisé que nous présentons dans les chapitres suivants. Ces travaux ont fait l'objet d'une publication d'un article traitant ce problème [1].

2.6.1 Génération de la base de données

Afin de détecter la tête du nageur, il est nécessaire d'avoir une description a priori de cette partie du corps à suivre. Pour ce faire, nous avons généré une base de données d'apprentissage en se basant sur des séquences vidéos que nous avons enregistrées durant des compétitions nationales (Championnats de France de natation - Limoges 2015)

CHAPITRE 2. PRÉTRAITEMENTS ADAPTÉS AUX VIDÉOS DU MILIEU AQUATIQUE : ANALYSE, CHOIX ET VALIDATION

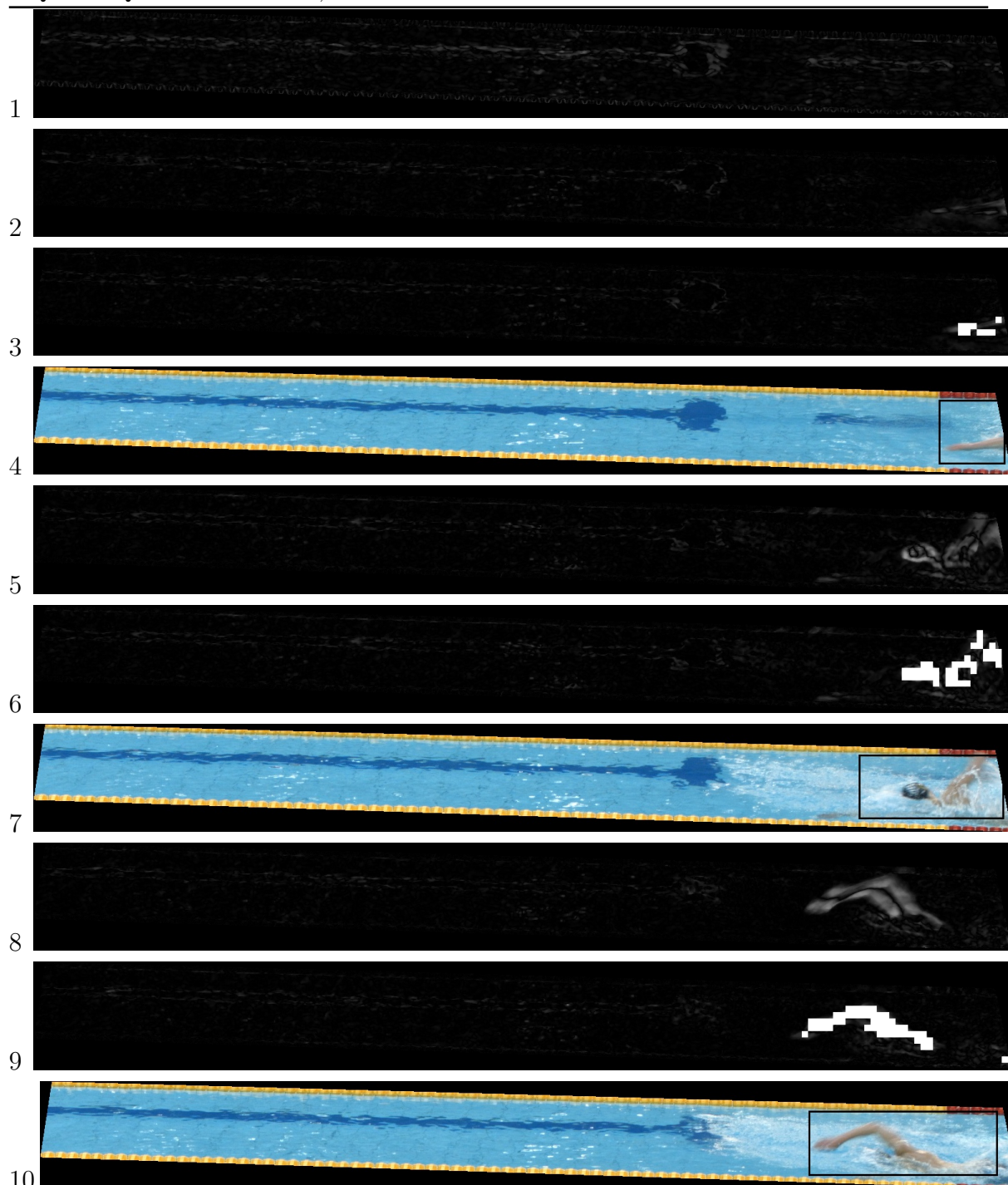


FIGURE 2.17 – Localisation de la région d'intérêt à base du principe *a contrario*. Image 1 : modèle de bruit obtenu à partir de la différence entre deux images du couloir vide. Images 2, 5 et 8 : images de différences entre deux images successives. Images 3, 6 et 9 : décomposition en blocs et seuillage par rapport à un seuil relatif au modèle de bruit. Images 4, 7 et 10 : localisation du nageur par la méthode *a contrario*. Championnats de France de Natation - Limoges, Avril 2015

et internationales (Championnats du monde de natation - Barcelone 2013, Kazan 2015). Ces vidéos ont été enregistrées en utilisant des caméras *4K* de très haute définition permettant d'avoir plus de détail sur les images pour extraire la tête du nageur de manière efficace. Notre base de données contient des têtes de nageurs que nous avons extraites à partir des différentes situations réelles apparaissant durant la nage. En particulier, les

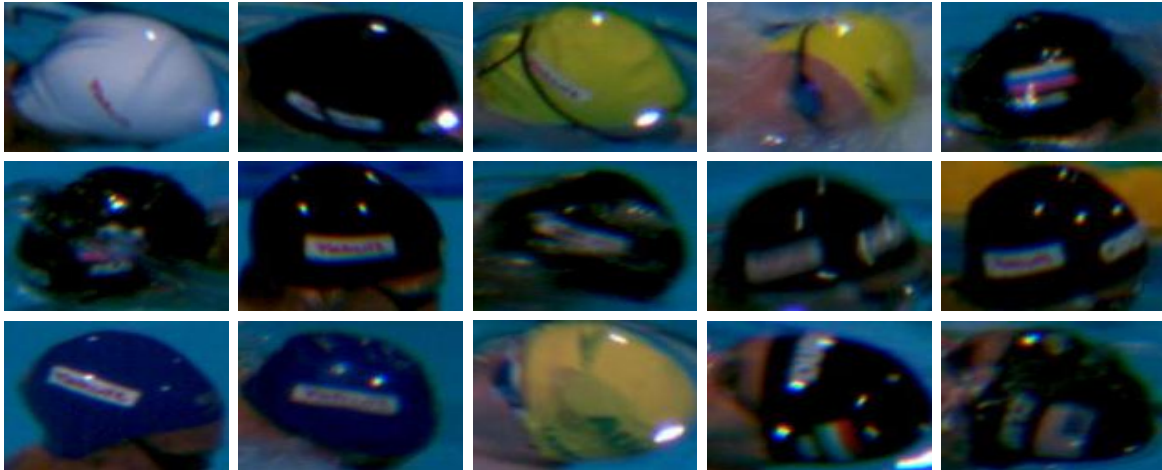


FIGURE 2.18 – Exemples de la base de données d'apprentissage. *World Championship, Barcelona 2013 and Kazan 2015.*

critères que nous avons pris en compte pour la génération de la base de données afin de couvrir la majorité des cas de figures sont :

- Âge : sénior ou junior
- Sexe : messieurs ou dames
- Type de nage : nage libre, dos, papillon ou brasse
- Direction de nage : aller ou retour

La figure 2.18 présente des exemple d'images de têtes/bonnets de différents nageurs dans des situations variées.

2.6.2 Application de la technique NL-JTC

Sachant que la couleur des bonnets portés par les nageurs peut varier, nous optons pour leur forme afin d'assurer une description pertinente et standard de la tête du nageur. Pour cette raison, nous choisissons la technique de corrélation NL-JTC qui est connue dans la littérature pour la détection de contour [7, 57, 102]. Comme nous l'avons présenté dans le chapitre précédent, le plan d'entrée de la technique NL-JTC comprend une image de référence et une image cible. La première image est une référence standard que nous générerons selon notre cas à partir de la base de données.

L'image cible, quant à elle, représente la région d'intérêt autour de la tête du nageur.

CHAPITRE 2. PRÉTRAITEMENTS ADAPTÉS AUX VIDÉOS DU MILIEU AQUATIQUE : ANALYSE, CHOIX ET VALIDATION

Celle-ci peut être extraite en utilisant l'approche de localisation de la région d'intérêt présentée dans la section précédente. Une fois que le plan d'entrée est généré, nous appliquons la technique NL-JTC, comme le montre la figure 2.19 et nous récupérons un plan de corrélation. L'analyse de ce dernier nous permet de prendre la décision concernant l'existence d'une cible ayant une forme ressemblant à la tête du nageur, ou non. Cette décision est prise sur un ensemble d'images de la vidéo et selon la valeur du PCE afin de choisir les meilleures cibles. Dans ce qui suit, nous allons détailler le processus de génération automatique de l'image de référence, l'image cible ainsi que la décision finale.

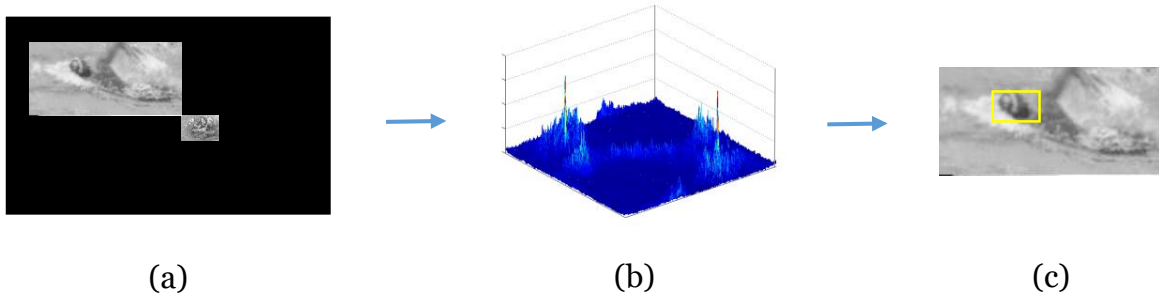


FIGURE 2.19 – Application de la technique NL-JTC pour la détection de la tête du nageur. (a) Plan d'entrée contenant une référence composite et une cible contenant la région d'intérêt issue du principe a contrario. (b) Plan de corrélation. (c) Détection et localisation de la tête dans la région d'intérêt. Championnats de France de Natation - Limoges, Avril 2015

2.6.3 Référence composite re-dimensionnée

A partir de la base de données pré-générée, nous choisissons n images de référence de têtes de nageurs relatives à notre cas. En pratique, nous fixons $n = 3$ afin d'avoir trois formes différentes de têtes de nageurs correspondant à la même situation, à savoir, le type de nage, la direction, l'âge, etc. Ensuite, les images de référence sélectionnées sont converties en niveaux de gris. Ceci est indispensable pour l'application de la technique NL-JTC et n'a aucune influence sur les résultats de la détection car l'information de couleur du bonnet est écartée.

Ensuite, nous appliquons le filtre composite [1, 103, 104] sur les n images dans le but de générer une seule image représentant une description de forme riche et contenant différentes têtes de nageurs. L'idée de base du filtre composite est de calculer une somme pondérée des images de référence, comme le montre l'équation suivante :

$$REF(x, y) = \sum \alpha_i * ref_i(x, y) \quad (2.7)$$



FIGURE 2.20 – Principe du filtre composite. À gauche : les images de référence. À droite : l'image composite qui consiste à une sommation pondérée des deux images de référence.

Avec α_i représente le coefficient de pondération qui peut être utilisé afin de favoriser la référence ref_i .

Enfin, cette référence composite doit être re-dimensionnée selon la taille de la tête du nageur concerné. Cette dernière est inconnue, afin de l'estimer, nous calculons le rapport en mètres entre la largeur standard de la tête et la largeur du couloir. Puis, en se basant sur ce rapport, la largeur du couloir en pixels et la fonction de calibrage, nous calculons la largeur estimée, en pixels, de la tête comme le montre l'équation suivante :

$$largeurTete_{pixel} = largeurCouloir_{pixel} * \frac{largeurTete_{metre}}{largeurCouloir_{metre}} \quad (2.8)$$

En utilisant le résultat de cette équation, nous re-dimensionnons notre référence composite afin de l'adapter à notre cas et selon la largeur du couloir concerné. Le processus de génération de la référence composite re-dimensionnée est résumé dans la figure 2.21

2.6.4 Pré-localisation de la région d'intérêt

La pré-localisation de la région d'intérêt est une étape importante pour développer une approche automatique précise de suivi des nageurs. Pour ce faire, nous nous basons sur le processus de localisation de la région d'intérêt basée sur le principe *a contrario* présenté dans la section précédente. Ceci nous permet de déterminer le moment de l'apparition du nageur dans le couloir ainsi que de restreindre la région d'intérêt. En effet, il est important de tenir compte des deux informations suivantes : la direction de nage et les blocs correspondant au mouvement du nageur. Afin de pré-localiser notre région d'intérêt, nous sélectionnons un rectangle de 2 m de longueur et d'une largeur de 1.5 m qui se termine au dernier bloc correspondant au mouvement du nageur en prenant en compte la direction de nage. Ces mesures sont transformées en pixels puis

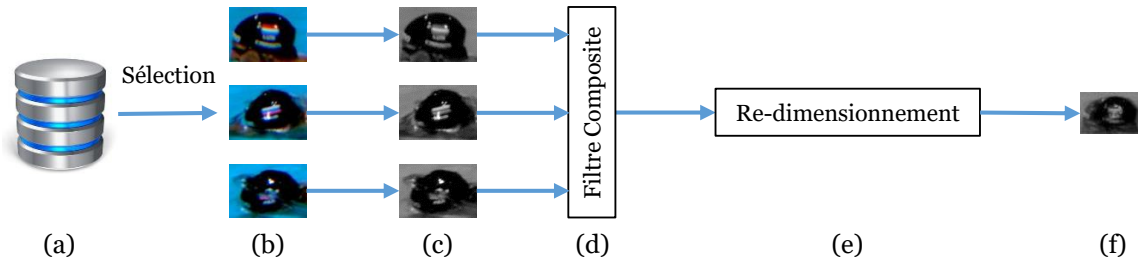


FIGURE 2.21 – Génération de la référence composite re-dimensionnée. (a) Base de données contenant les références de têtes. (b) Sélection des images relatives à la situation présente. (c) Passage de l'espace couleur RGB aux niveaux de gris. (d) Application du filtre composite pour la fusion des trois images. (e) Re-dimensionnement de l'image composite selon la taille du couloir concerné. (f) Résultat : référence composite re-dimensionnée Championnats de France de Natation - Limoges, Avril 2015

la région d'intérêt est extraite comme le montre la figure 2.22.

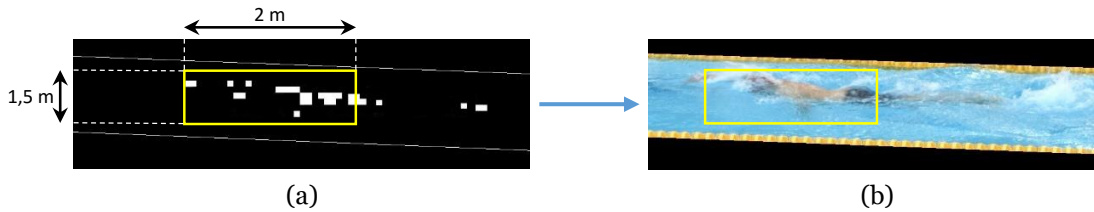


FIGURE 2.22 – Pré-localisation de la région d'intérêt. (a) Image de différence décomposée en blocs et seuillée. Le rectangle représente la région d'intérêt centrée dans la largeur du couloir en éliminant 0.5 m des deux côtés et qui s'étend sur une longueur de 2 m en commençant par le premier bloc correspondant au mouvement du nageur. (b) Projection du rectangle de localisation sur l'image originale. Championnats de France de Natation - Limoges, Avril 2015

2.6.5 Décision basée sur PCE

Notre objectif est de détecter avec précision la tête du nageur afin d'initialiser le système de suivi des nageurs que nous présenterons dans les prochains chapitres. Pour cela, nous appliquons l'approche proposée, nommée *Scaled Composite JTC* [1], afin de détecter le nageur dans les premières images de la séquence vidéo. Cette période est censée contenir les événements suivants : couloir vide, plongeon et reprise de nage. Durant le premier événement, nous appliquons seulement le processus de localisation de la région d'intérêt basée sur le principe *a contrario*. Une fois que le mouvement du

nageur est détecté dans le couloir, nous appliquons l'approche *Scaled Composite JTC* afin de détecter et localiser la tête du nageur dans chaque image de cette période. La durée de celle-ci est fixée entre 2 et 3 secondes (60 et 75 images). Ensuite, pour chaque détection potentielle, nous calculons la valeur du PCE qui nous servira de facteur de confiance. Enfin, les détections potentielles sont classées selon leurs valeurs PCE associées et la décision finale correspond à celles ayant les plus hauts PCE, comme le montre la figure 2.23. Dans notre cas, nous validons trois cibles détectées pour l'initialisation du système de suivi de nageur optimisé que nous proposons dans les deux prochains chapitres.

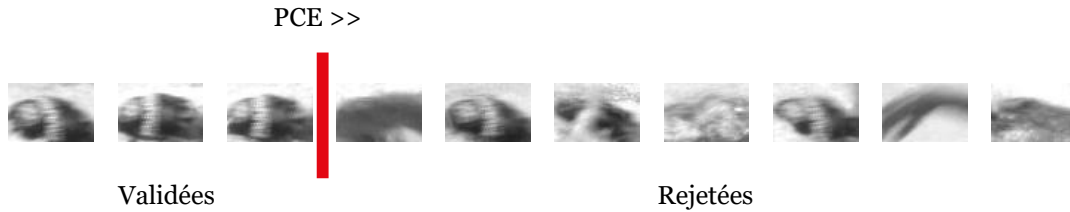


FIGURE 2.23 – Validation des détections potentielles selon un critère de décision basé sur le PCE. Les cibles détectées sont classées selon l'ordre décroissant de leurs valeurs de PCE. Les trois cibles potentielles validées sont celles ayant les plus hauts PCE. Championnats de France de Natation - Limoges, Avril 2015

2.7 Conclusion

L'objectif principal de ce chapitre était d'effectuer des prétraitements sur les séquences vidéo afin de les préparer pour des traitements ultérieurs plus précis. Nous avons commencé ce chapitre par une étude préliminaire de l'environnement de travail, où nous avons décrit les différentes caractéristiques des bassins dédiés aux compétitions de haut niveau. Puis, nous avons étudié des images de compétitions dans le but de segmenter le bassin et de détecter les différents objets présents dans la scène : eau, lignes de couloirs et nageur. En raison d'un manque de précision, nous avons introduit le principe de calibrage qui permet de modéliser la relation entre les coordonnées métriques et pixels. Ceci a permis de localiser le bassin dans l'image, d'extraire les lignes délimitant les couloirs de nage et d'effectuer des mesures précises dans l'image.

Une fois la séquence vidéo calibrée et le bassin segmenté, nous avons proposé une approche de localisation de la région d'intérêt basée sur le principe *a contrario*. Pour ce faire, nous avons extrait le couloir concerné, puis, nous avons calculé l'image de différence entre deux images successives afin de détecter le mouvement du nageur, en le comparant avec un modèle de bruit. Cette approche sert principalement à restreindre

CHAPITRE 2. PRÉTRAITEMENTS ADAPTÉS AUX VIDÉOS DU MILIEU AQUATIQUE : ANALYSE, CHOIX ET VALIDATION

la zone de recherche autour du nageur pour suivre précisément sa tête. Le suivi de la tête du nageur que nous présenterons par la suite nécessite une détection automatique de la référence initiale.

Pour cela, nous avons proposé l'approche de la référence composite re-dimensionnée (*Scaled Composite Reference*) basée sur le NL-JTC. L'idée de base consiste à créer une base de données contenant des têtes de nageurs classées selon les différentes situations. Puis, trois images sont choisies, en fonction de la situation présente, afin de générer une référence standard en appliquant le filtre composite. Cette référence composite est re-dimensionnée selon la taille du couloir concerné. Ensuite, en appliquant la technique NL-JTC, elle est utilisée comme image de référence pour détecter la tête du nageur concerné dans une région d'intérêt issue du principe *a contrario*. Enfin, nous choisissons, selon les valeurs des PCE, trois images de référence sur l'ensemble des images détectées durant une période déterminée. Ce processus est résumé dans la figure 2.24.

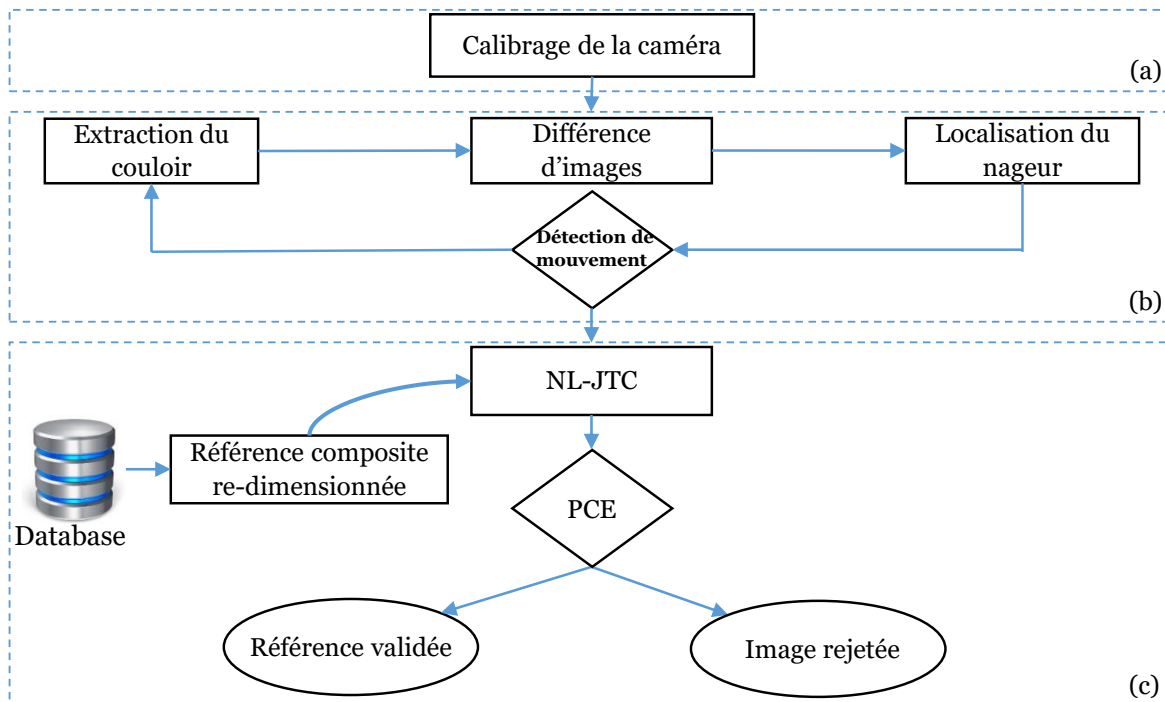


FIGURE 2.24 – Processus de détection automatique de la tête de nageur pour l'initialisation d'un système de suivi précis. (a) Calibrage de la caméra. (b) Processus de localisation de la région d'intérêt. (c) NL-JTC composite re-dimensionnée pour la détection d'une référence pertinente pour le système de suivi optimisé.

Chapitre 3

Système de suivi de nageurs : Analyse, choix et validation

Sommaire

3.1	Introduction	71
3.2	Processus général de suivi de nageurs	72
3.3	Suivi par corrélation	72
3.3.1	NL-JTC pour le suivi des nageurs	73
3.3.2	Choix du coefficient de non-linéarité	73
3.3.3	Limitations de NL-JTC pour le suivi	74
3.3.4	Etude et optimisation du plan de corrélation	77
3.4	Suivi par des techniques basées sur les histogrammes . . .	83
3.4.1	Processus général	83
3.4.2	Histogramme de couleur pour le suivi des nageurs	84
3.4.3	LBP pour le suivi des nageurs	87
3.4.4	HOG pour le suivi des nageurs	89
3.4.5	Avantages et limitations des techniques basées sur les histogrammes	89
3.5	Conclusion	91

3.1 Introduction

Etant donnée la popularité de la natation ainsi que la forte compétitivité entre les nations, un grand intérêt est donné à l'étude des performances des nageurs. Dans ce cadre, l'objectif de cette thèse est de développer un système permettant d'évaluer

et d'améliorer leurs stratégies de nage à travers l'analyse de leur vitesse instantanée. Ce système doit être automatique et sans capteur porté par le nageur. Pour cela, nous proposons, dans ce chapitre, un système de suivi de nageurs à partir des séquences vidéo dans le but d'estimer leurs performances. En effet, deux approches seront présentées, un suivi par corrélation basé sur la technique JTC et un suivi par histogrammes basé sur l'histogramme de couleur, LBP et HOG.

3.2 Processus général de suivi de nageurs

Afin de remplir les objectifs de la thèse, nous proposons un système de suivi de nageurs qui se compose des modules suivants :

Prétraitements (Chapitre 2) : Ce module a été détaillé dans le chapitre précédent. L'idée de base consiste à calibrer la caméra pour s'adapter à la géométrie du bassin et obtenir une correspondance entre les coordonnées métriques et les coordonnées pixels dans l'image. Ceci va nous permettre de faire des mesures dans l'image, extraire le couloir concerné, restreindre la région d'intérêt, prédire la trajectoire du nageur et détecter la référence initiale de suivi (tête du nageur à suivre).

Suivi de nageurs (Chapitres 3-4) : A partir de la référence initiale, une région d'intérêt est prédite, puis, le nageur est suivi en utilisant une des deux approches que nous détaillons par la suite : suivi par corrélation et suivi par histogrammes. Après chaque détection du nageur, sa position en mètres est calculée afin d'estimer la distance parcourue. Enfin, la cible détectée peut être utilisée pour l'actualisation de la référence et assurer la cohérence temporelle.

Estimation de performances (Chapitre 5) : Le/la nageur(se) sera évalué(e) selon les courbes de vitesse instantanée qui sont générées au fur et à mesure du processus de suivi.

L'objectif de ce chapitre consiste à présenter en détail notre système proposé pour le suivi de nageurs. Pour cela, nous allons faire une étude approfondie des deux approches de suivi de nageurs dans le but de cibler leurs limitations et optimiser le module de suivi. Ce dernier représente une étape cruciale pour l'étude des performances des nageurs qui sera présentée par la suite.

3.3 Suivi par corrélation

Les techniques basées sur la corrélation ont montré leur efficacité pour le suivi dans les environnements contrôlés et semi-contrôlés [7, 13, 15]. Afin de répondre aux besoins de cette thèse nous avons cherché à adapter ces approches pour les rendre performantes dans le cas du suivi des nageurs. Pour cela, nous avons analysé les avantages et les

limitations des variantes de la technique de corrélation JTC présentées dans le chapitre 1 section 1.5 et nous avons retenu la technique NL-JTC. Ce choix est argumenté par la capacité de cette technique à contrôler la finesse des pics de corrélation permettant le contrôle de la robustesse et de la discrimination.

3.3.1 NL-JTC pour le suivi des nageurs

Afin de suivre le nageur durant sa course, nous avons proposé et développé un système de suivi basé sur la technique de corrélation NL-JTC [15, 57]. Pour cela, nous considérons sa tête/bonnet comme référence. Cette dernière sera sélectionnée dans l'image initiale im_i à l'instant i de la séquence vidéo grâce à la méthode *Scaled Composite JTC* [1] présentée dans le chapitre 2. Ainsi, la cible correspond à l'image de l'instant $i + 1$ de la séquence. Cependant, dans le but de simplifier les calculs et de limiter les erreurs de suivi, nous restreignons la zone d'intérêt autour de la dernière position détectée. Le plan d'entrée (f) de la technique NL-JTC, présenté dans la figure 3.4, contient une image de référence $Ref(x - d, y - d)$ et une image cible $Cible(x, y)$ séparées d'une distance d .

D'abord, nous appliquons la Transformée de Fourier sur le plan d'entrée pour obtenir un spectre joint. Afin de générer le plan de corrélation, nous appliquons la Transformée de Fourier inverse sur l'intensité de ce spectre joint. Puis, nous appliquons le Filtre Zéro pour éliminer le pic d'autocorrélation. Ceci consiste en la multiplication élémentaire du plan de corrélation par une matrice unitaire de même taille contenant des zéros au centre. Ensuite, la fonction de non-linéarité est appliquée avec une valeur de k qui sera déterminée en trouvant le bon compromis entre la discrimination et la robustesse. Ceci sera détaillé dans la section 3.3.2 suivante. La position du centre de la cible correspond à la valeur maximale du plan de corrélation.

Pour assurer la cohérence temporelle du suivi, la référence et la région d'intérêt devront être actualisées pour la prochaine itération. Pour ce faire, la référence peut être actualisée par la cible détectée dans l'image im_{i+1} . Ainsi, la région d'intérêt peut être extraite de l'image $i + 2$ dans la même position que la cible détectée et avec une taille trois fois plus grande que la taille de la référence. Enfin, l'itération $i + 1$ est relancée avec un plan d'entrée qui contient la nouvelle référence et la nouvelle région d'intérêt.

3.3.2 Choix du coefficient de non-linéarité

La particularité de la technique NL-JTC consiste en son coefficient de non-linéarité qui contrôle la largeur du pic. Cette dernière a une influence sur la capacité de discrimination et de robustesse qui représentent deux notions importantes dans les techniques

CHAPITRE 3. SYSTÈME DE SUIVI DE NAGEURS : ANALYSE, CHOIX ET VALIDATION

de détection et de suivi. En effet, la première permet de faire la différence entre les différents objets dans une scène donnée et la cible recherchée, tandis que la seconde permet d’avoir une détection de la cible dans tous les cas de figures (rotation, bruit, occultations...). Le plus souvent, ces deux notions s’opposent et il est difficile de combiner la discrimination et la robustesse avec une seule méthode. En revanche, pour la corrélation basée sur l’approche NL-JTC, un compromis entre ces deux notions peut être obtenu grâce au coefficient de non-linéarité k . Afin de déterminer l’intervalle de bon compromis pour notre cas de suivi de nageur, nous avons calculé la valeur du PCE (voir chapitre 1, section 1.5.6) pour le même plan d’entrée avec différentes valeur du coefficient de non-linéarité.

Valeur de k	0.1	0.3	0.5	0.7	0.9
PCE	0.023	0.029	0.012	0.004	0.001

TABLE 3.1 – *Evolution de la valeur du PCE en fonction de la valeur du coefficient de non-linéarité k . Tests effectués sur l’image de référence de la figure 3.1a et l’image cible de la figure 3.1b.*

La table 3.1 présente l’évolution de la valeur du PCE en fonction de la valeur du coefficient de non-linéarité k . Le plan d’entrée utilisé contient la référence et la cible présentées dans la figure 3.1a et 3.1b. Les valeurs élevées du PCE correspondent à $k = 0.1, 0.3$ et 0.5 . Comme le montre les figures 3.1c, 3.1d et 3.1e, ces cas font apparaître des pics de corrélation très fins avec moins de bruit, ce qui implique une forte discrimination. Dans le cas de $k \geq 0.3$, nous remarquons que la valeur et la largeur des pics sont accrues, signifiant une augmentation de la robustesse des détections. Dans le cas de notre application de suivi de nageur, nous retenons un coefficient de non-linéarité $k \in [0.3, 0.5]$ qui permet d’obtenir le meilleur compromis entre la discrimination et la robustesse.

3.3.3 Limitations de NL-JTC pour le suivi

Malgré ses avantages, l’utilisation de la technique NL-JTC pour le suivi de nageur reste insuffisante. En effet, cette méthode fait apparaître des décrochages (perte de la cible à suivre) au cours du processus de suivi. Cela est dû aux limitations de cette technique dans le cas de l’environnement étudié. En effet, le milieu aquatique, ainsi que le suivi de nageurs, induit les faiblesses suivantes :

- Actualisation de la référence avec une cible qui correspond à une mauvaise détection.
- Changement de la forme et rotation de la tête du nageur durant la nage.
- Sensibilité importante aux déformations du contour.
- Corrélation entre les bords de la référence et de la cible.

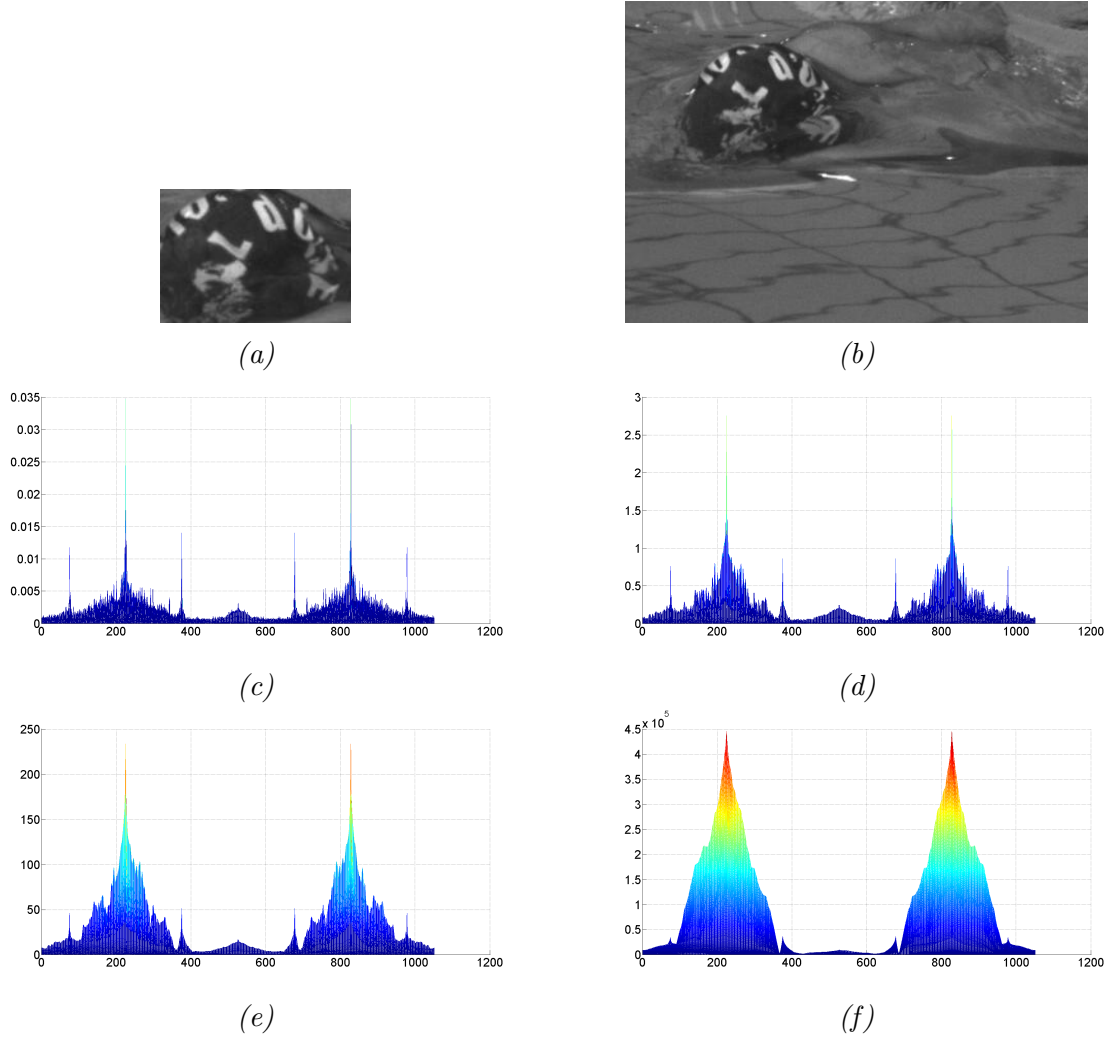


FIGURE 3.1 – Variation des plans de corrélation en fonction de la valeur du coefficient de non-linéarité k . (a et b) Référence et cible constituant le plan d'entrée de la technique NL-JTC. (c, d, e et f) Plans de corrélation avec $k = 0.1, 0.3, 0.5$ et 0.8 respectivement. A chaque fois que la valeur de k augmente, les pics s'élargissent en améliorant la robustesse et diminuant la discrimination.

- Effet de flou lié au mouvement du nageur et à la caméra utilisée.
- Changement d'échelle.

Afin de surmonter ces limitations, nous proposons dans la section suivante une étude approfondie du plan de corrélation permettant d'appuyer les optimisations que nous proposons.



FIGURE 3.2 – Exemple d’images pouvant être utilisées pour localiser la zone potentielle de corrélation. (a) image cible pour le coin haut/gauche. (b) image cible pour le coin haut/droit. (c) image cible pour le coin bas/gauche, (d) image cible pour le coin bas/droit. (e) image de référence.

3.3.3.1 Extraction de la zone potentielle de corrélation

Le plan de corrélation classique, tel que présenté dans la section précédente, peut être séparé en deux zones distinctes : l’une porteuse d’informations, répartie entre les coins haut/gauche et bas/droit, l’autre non porteuse d’informations, répartie entre les coins bas/gauche et haut/droit, tel que clairement identifiable sur la figure 3.4b. Grâce à cette observation, nous proposons de restreindre le plan de corrélation à un sous plan contenant uniquement la zone de corrélation représentant la similarité entre la référence et la cible. Ainsi, “la zone potentielle de corrélation” pourra être fusionnée avec notre carte de scores couleurs et limitera la possibilité de trouver de faux pics dans le reste du plan. Notre but est donc de localiser précisément cette zone potentielle de corrélation. Pour cela, nous nous basons sur les trois relations entre la position des deux pics de corrélation et leur position dans le plan d’entrée, à savoir :

1. La distance entre les pics sur le plan de la corrélation est égale à deux fois la distance entre le centre de l’image référence et le centre de l’objet à détecter dans la cible sur le plan d’entrée.
2. La pente de la droite passant par les pics est égale à celle passant par le centre de l’image référence et le centre de l’objet à détecter dans la cible.
3. Le plan de corrélation est symétrique par rapport à son centre.

Localiser et extraire cette zone potentielle de corrélation revient à chercher la position précise de ses quatre coins. Pour illustrer cette démarche, nous proposons d’utiliser le NL-JTC sur quatre plans d’entrée, chacun composé d’une image cible de Léna, coupée de telle sorte que son visage soit placé exactement sur l’un des quatre coins, voir figure 3.2a-d, et du visage de Léna comme référence, voir figure 3.2e. Dans le premier cas, figure 3.2, les distances de_x et de_y , voir figure 3.3, entre les centres de la référence et de la même image dans la cible sont données par :

$$\begin{cases} de_x = l_r/2 + l_c - l_r/2 = l_c \\ de_y = h_r/2 + h_c - h_r/2 = h_c \end{cases} \quad (3.1)$$

Où l_r , h_r , l_c et h_c représentent respectivement la hauteur et largeur des images référence et cible. D'après les relations 1 et 2, établissant le lien entre les plans d'entrée et de sortie, nous savons que les distances ds_x et ds_y entre les pics dans le plan de sortie, voir figure 3.3b sont données par :

$$\begin{cases} ds_x = 2de_x = 2l_c \\ ds_y = 2de_y = 2h_c \end{cases} \quad (3.2)$$

Sachant que le plan de corrélation est symétrique, il est possible de calculer les coordonnées du pic à partir du centre du plan, voir figure 3.3b. A partir de l'équation 3.2, les coordonnées $pic1_x$ et $pic1_y$ du pic de corrélation, correspondant au coin haut/gauche, sont données par :

$$\begin{cases} pic1_x = l_p/2 - (ds_x)/2 = l_p/2 - l_c \\ pic1_y = h_p/2 - (ds_y)/2 = h_p/2 - h_c \end{cases} \quad (3.3)$$

Où l_p et h_p représentent respectivement la hauteur et largeur du plan d'entrée. Les coordonnées des autres coins (2 – haut/droite, 3 – bas/gauche et 4 – bas/droite) peuvent être calculées de la même manière :

$$\begin{cases} (pic2_x, pic2_y) = (l_p/2 - l_r, h_p/2 - h_c) \\ (pic3_x, pic3_y) = (l_p/2 - l_c, h_p/2 - h_r) \\ (pic4_x, pic4_y) = (l_p/2 - l_r, h_p/2 - h_r) \end{cases} \quad (3.4)$$

Ainsi, les dimensions ZPC_l et ZPC_h de la zone potentielle de corrélation peuvent être calculées par :

$$\begin{cases} ZPC_x = l_c - l_r \\ ZPC_y = h_c - h_r \end{cases} \quad (3.5)$$

Grâce à cette extraction, la détection de l'objet dans la cible revient à localiser simplement le pic maximal de corrélation dans la zone potentielle de corrélation. En effet, dans cette nouvelle approche, la position de l'objet dans l'image cible est donnée par les mêmes coordonnées.

3.3.4 Etude et optimisation du plan de corrélation

La création du plan d'entrée pour les architectures JTC est une étape importante, car la position du pic de corrélation va dépendre de l'emplacement des images cible et

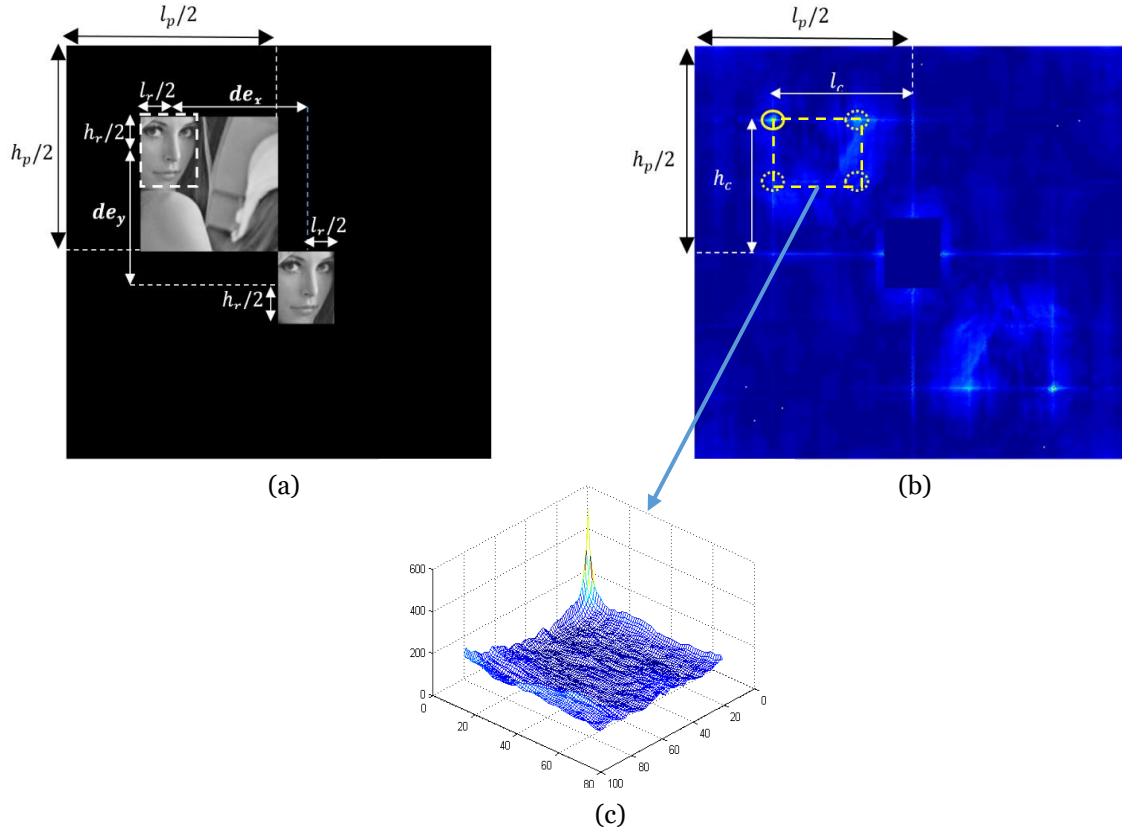


FIGURE 3.3 – Localisation et extraction de la zone potentielle de corrélation. (a) plan d'entrée pour la localisation du coin haut/gauche. (b) plan de corrélation. Le cercle continu représente le pic de corrélation issu du plan d'entrée alors que les autres cercles représentent les pics de corrélation pour les autres plan d'entrées contenant une des cibles présentées dans la figure 3.2 (coin haut/droit, bas/gauche et bas/droit). Un filtre est appliqué au centre du plan de corrélation pour enlever le pic d'autocorrélation. (c) « zone potentielle » de corrélation extraite du plan complet.

référence dans ce plan. Le plan d'entrée standard proposé dans la littérature [15, 55] est construit de la manière suivante :

$$\begin{cases} (c_x, c_y) = (p, p) \\ (r_x, r_y) = (p + c_x, p + c_y) \end{cases} \quad (3.6)$$

Où les couplets (c_x, c_y) et (r_x, r_y) représentent respectivement les coordonnées des coins haut/gauche de la cible et de la référence dans le plan d'entrée. Avec p qui représente le décalage entre l'image cible et le bord du plan et permet d'éviter que l'information utile, i.e. le pic de corrélation, sorte du plan de corrélation. Cette valeur est fixée em-

piriquement par rapport à la taille de la référence, voir figure 3.4. Finalement, la taille du plan d'entrée est donnée par :

$$\begin{cases} l_p = 2 * l_c + 2 * p \\ h_p = 2 * h_c + 2 * p \end{cases} \quad (3.7)$$

Où (l_p, h_p) , (l_c, h_c) représentent respectivement la largeur et la hauteur du plan d'entrée et de l'image cible.

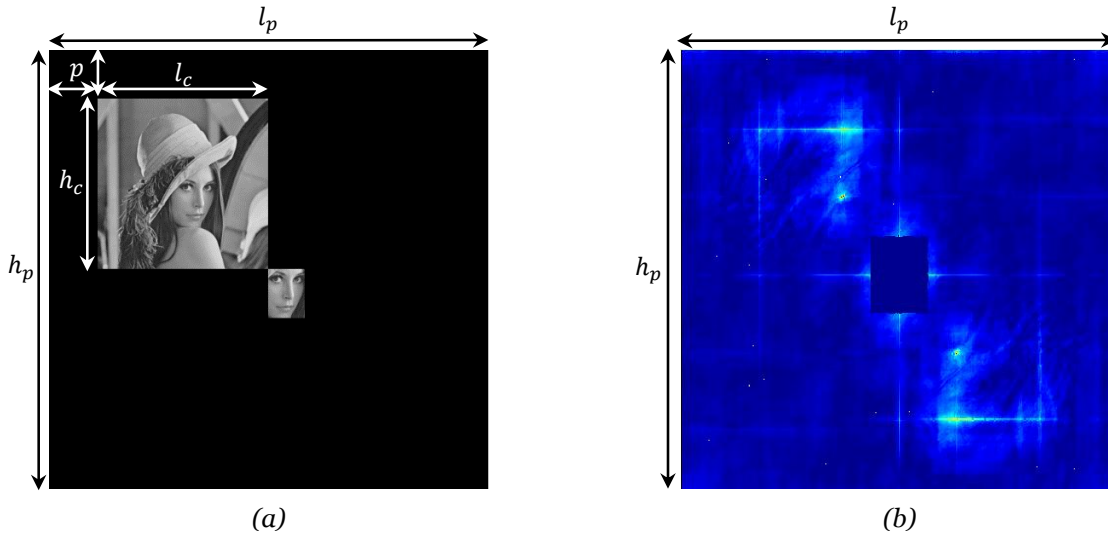


FIGURE 3.4 – Application de la technique NL-JTC : (a) plan d'entrée où les images sont placées en diagonale. (b) plan de corrélation. Ici, un filtre est appliqué au centre du plan de corrélation pour enlever les résidus du pic d'autocorrélation.

3.3.4.1 Nouvelle organisation du plan d'entrée

Le plan de corrélation, obtenu par les méthodes de la littérature, comporte deux zones potentielles de corrélation (en haut à gauche et en bas à droite) ainsi que deux zones ne comportant que très peu d'informations (en bas à gauche et en haut à droite), comme le montre la figure 3.4. Afin d'améliorer le temps de calcul, nous proposons de réduire la taille du plan d'entrée à son strict minimum en réduisant, autant que possible, les zones non porteuses d'informations. Pour cela, nous proposons le plan d'entrée réduit, illustré par la figure 3.5, construit de la manière suivante :

$$\begin{cases} (c_x, c_y) = (p, p) \\ (r_x, r_y) = (p + l_c, p + h_c/2 - h_r/2) \end{cases} \quad (3.8)$$

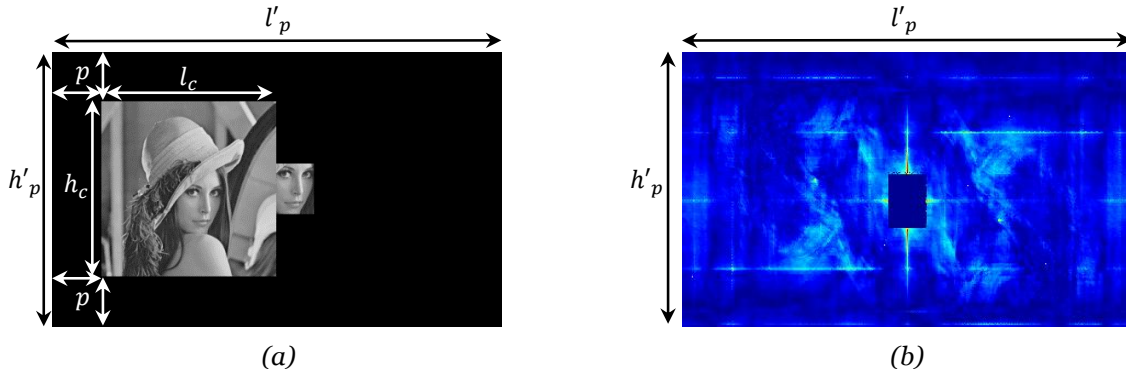


FIGURE 3.5 – Nouvelle organisation du plan d'entrée, où les images sont placées horizontalement, afin de réduire le temps de calcul de la méthode NL-JTC. (a) plan d'entrée réduit. (b) plan de corrélation réduit.

Avec cette nouvelle organisation, le nouveau plan réduit a une taille différente de ceux de la littérature présenté dans la figure 3.4. En effet, le plan d'entrée proposé permet l'élimination des zones non porteuses d'informations et la réduction de la taille des plans qui devient :

$$\begin{cases} l'_p = 2 * l_c + 2 * p \\ h'_p = h_c + 2 * p \end{cases} \quad (3.9)$$

Le gain de taille des plans obtenu par notre méthode par rapport à ceux de la littérature est donné par :

$$gain = \frac{h_c}{2 * h_c + 2 * p} \simeq 50\% \quad (3.10)$$

Ainsi, notre méthode de suppression des zones non porteuses d'information permet de réduire le plan d'entrée de la technique NL-JTC et donc de réduire le temps de calculs. Une fois le plan de corrélation généré, nous pouvons chercher le pic de corrélation qui se trouve dans la zone potentielle décrit dans la section précédente.

3.3.4.2 Mesure de confiance pour la détection

Dans le système que nous proposons, la validation de la cible détectée est une étape cruciale dans le processus de suivi du nageur. Pour cela, il est nécessaire de mesurer la confiance que l'on a pour chacune des cibles détectées. Grâce à cela, nous établissons trois seuils de validation : $s_1 \geq s_2 \geq s_3$.

Le seuil s_1 est le plus exigeant et sert à valider la possibilité de l'actualisation de la référence. Cette dernière ne peut être remplacée par la cible détectée que dans le cas

d'une forte ressemblance.

Le seuil s_2 permet d'améliorer la robustesse de la prédiction en créant une position référentielle correspondant à la position de la cible détectée. Elle sera actualisée à chaque fois que le critère de ressemblance est supérieur à s_2 .

Finalement, s_3 est le seuil minimal qui permet de considérer une cible comme étant une bonne détection. Ceci implique que, si dans toutes les images de la séquence, la cible a une valeur de ressemblance supérieure à s_3 alors sa position sera prise en compte pour les différentes mesures de la distance parcourue et de la vitesse.

Le PCE est le critère utilisé pour mesurer la ressemblance entre la référence et la cible pour les techniques basées sur la corrélation. Ce critère de ressemblance représente le rapport entre l'énergie du pic de corrélation et l'énergie du plan. Cependant, dans le cas du suivi des nageurs, les plans de corrélation sont très bruités et la valeur du PCE ne représente pas une mesure robuste de ressemblance. Afin de palier cela, nous nous inspirons de l'approche proposée par Katz et al. [7] où ils mesurent la ressemblance entre les histogrammes de couleurs, de la référence et de la cible, grâce à la mesure du χ^2 . Ainsi, afin de mieux quantifier la ressemblance, nous proposons d'utiliser la distance de Bhattacharyya. Celle-ci a l'avantage de donner une mesure significative $\in [0, 1]$, au lieu de $[0, +\infty[$ dans le cas du χ^2 .

3.3.4.3 Filtre composite temporel

Nous avons constaté que lorsque l'on utilise une seule référence pour le suivi, il devient difficile de suivre le nageur durant une période de plusieurs secondes. En effet, si nous utilisons la première image de référence, la corrélation diminue avec le temps et si nous actualisons toujours par la dernière cible détectée, il devient impossible de récupérer le suivi après un décrochage. Pour résoudre ces problèmes, nous proposons de garder la première référence, pour assurer la robustesse, ainsi que la cible actualisée. Ceci peut être effectué en profitant du principe du filtre composite proposé pour les architectures JTC [103, 104].

3.3.4.4 Suppression des artefacts de corrélation des bords

Lorsque la ressemblance des images référence et cible est faible (bruit, occultation, rotation...) la position du pic de corrélation est souvent erronée. En effet, dans ce cas, le pic d'énergie maximum se trouve fréquemment sur les bords de la zone potentielle de corrélation. Après une étude approfondie de ce phénomène, nous avons noté qu'une zone de forte corrélation apparaissait tout autour du sous plan et particulièrement dans ses coins. Cet artefact est dû à la corrélation des bords des images cible et référence

entre eux, comme le montre la figure 3.6. Pour éviter ces erreurs, nous proposons de restreindre la zone potentielle de corrélation de l'équation 3.5 de quelques pixels (5 dans notre implémentation) afin de supprimer la zone représentant la corrélation entre les bords des deux images. Cependant, ceci n'est possible que si l'on pose l'hypothèse que l'objet à détecter ne se situe pas sur l'un des bords de l'image cible. Pour prendre en compte ce cas de figure, il est possible de générer un plan d'entrée où l'image cible a une taille plus importante (5 pixels dans notre cas) afin de s'assurer que l'objet à détecter ne se trouve pas sur les bords de l'image.

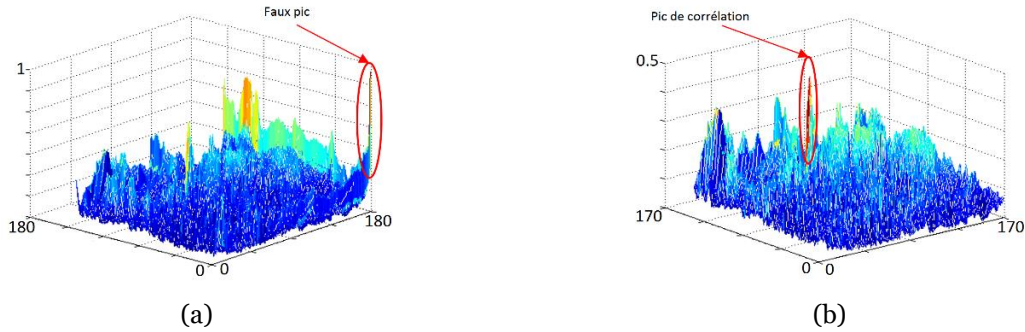


FIGURE 3.6 – *Suppression de la zone correspondant à la corrélation des bords. (a) – zone potentielle de corrélation complète où le grand pic correspond à la corrélation des bords (fausse détection). (b) – zone potentielle de corrélation restreinte après l'élimination des artéfacts de la corrélation des bords et où le grand pic correspond à la vraie détection.*

3.3.4.5 Limitation de l'effet de flou par le filtre de *Sobel*

Les techniques de corrélation, notamment le NL-JTC, sont très sensibles à la déformation du contour de l'objet à détecter. A cause du flou de mouvement fortement présent dans les vidéos que nous traitons, la détection et le suivi des objets est difficile voire impossible. Pour réduire ce problème, nous proposons d'aider le processus de corrélation par un rehaussement de contours effectué à l'aide du filtre de *Sobel*, par exemple, appliqué aux images cible et référence. Comme le montre la figure 3.7a et la figure 3.7b, où un flou de mouvement a été appliqué à l'image cible du plan d'entrée, le plan de corrélation original est très fortement bruité et ne permet pas la localisation/détection de la référence dans la cible. En revanche, après application de la méthode de filtrage proposée, le plan de corrélation est significativement amélioré et permet une localisation/détection correcte de la référence, voir figure 3.7c et figure 3.7d. Finalement, l'application du filtre de *Sobel* rehausse les contours floutés par le mouvement de l'objet à détecter, et permet de ce fait une localisation plus précise de l'objet grâce à un plan de corrélation moins bruité.

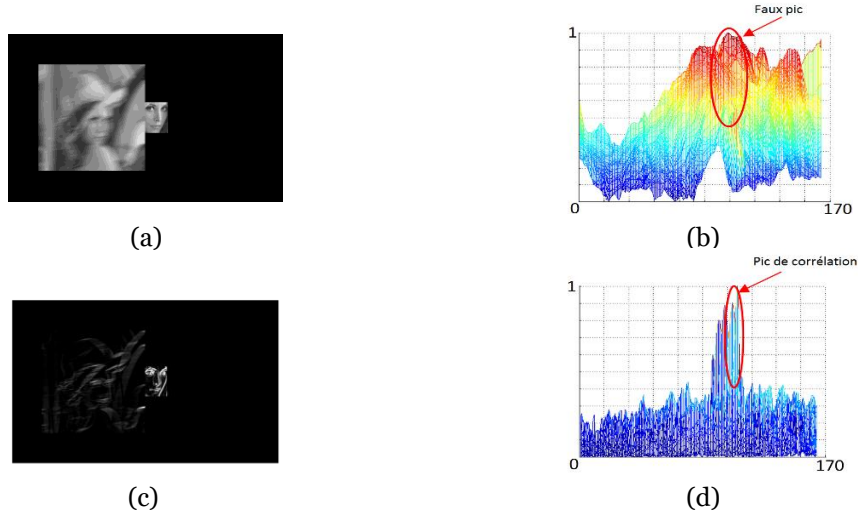


FIGURE 3.7 – Limitation de l'effet de flou par l'application du filtre de Sobel. (a) – plan d'entrée contenant une image cible floutée par un flou de mouvement. (b) – plan de corrélation fortement bruité issu de l'image (a) où la localisation est erronée. (c) – plan d'entrée filtré avec le filtre de Sobel. (d) – plan de corrélation débruité issu de l'image (c) où la localisation est correcte.

3.4 Suivi par des techniques basées sur les histogrammes

Les techniques basées sur les histogrammes consistent à représenter chacune des deux images, référence et cible, par un histogramme qui mesure la distribution d'une information donnée, appelée descripteur. Différents types de descripteurs peuvent être utilisés pour décrire les images, notamment la couleur, la texture en utilisant les LBP (motifs binaires locaux) [19, 20] et le gradient en utilisant les HOG (histogrammes de gradient orienté) [21]. En utilisant ces approches, le degré de ressemblance entre les images référence et cible est obtenu par comparaison de leur histogramme respectif grâce à une mesure adaptée. On peut noter entre autres la distance de Bhattacharyya ou encore le χ^2 [58].

3.4.1 Processus général

Dans le cadre du suivi des nageurs, nous sélectionnons la tête du nageur comme référence car elle représente la partie du corps la plus visible durant la nage. La zone de recherche, quant à elle est extraite automatiquement autour de la dernière position détectée, ou autour de la position de la référence sélectionnée dans le cas de la première

itération. La figure 3.8 présente un exemple d’une référence sélectionnée dans une image

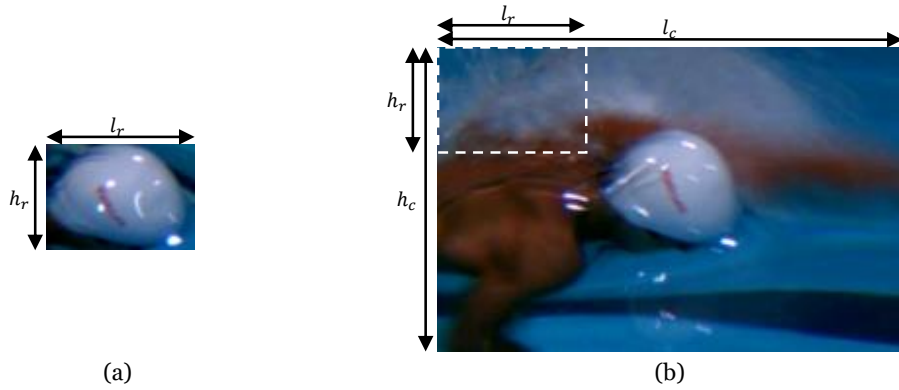


FIGURE 3.8 – Détection basée sur la comparaison d’histogrammes. (a) Image de référence. (b) Zone d’intérêt avec une fenêtre glissante, représentée par un rectangle en pointillé, parcourant la zone à la recherche de la cible. Pour chaque fenêtre une comparaison d’histogramme est effectuée afin de calculer un score de ressemblance.

à l’itération i et la zone d’intérêt à l’itération $i + 1$. Une recherche exhaustive peut être appliquée en comparant la référence avec une fenêtre de même taille (encadré en pointillé) parcourant toute la zone de recherche en utilisant le principe *Mean-shift* [73, 74] introduit dans le chapitre 1 section 1.8.3. Ici la comparaison est basée non pas sur les pixels eux-mêmes mais sur les histogrammes de couleur, LBP ou HOG calculés pour chacune des zones. Les scores de ressemblance calculés en parcourant la zone de recherche sont donnés par la formule suivante :

$$Scores = \frac{(l_c - l_r) * (h_c - h_r)}{pas} \quad (3.11)$$

Avec pas qui représente le pas de la fenêtre glissante permettant de fixer le niveau de précision (décalage entre chaque fenêtre glissante). Dans notre implémentation, nous l’avons fixé à 1 pixel afin d’avoir une précision maximale. Finalement, la cible correspond à la zone qui obtient le score maximum vis-à-vis de la référence. Dans notre cas, nous avons opté pour la distance de Bhattacharyya pour mesurer la ressemblance entre les histogrammes car elle génère des scores $\in [0, 1]$. Dans la suite de cette section, nous étudions l’apport des histogrammes (couleur, LBP et HOG) pour traiter le suivi des nageurs.

3.4.2 Histogramme de couleur pour le suivi des nageurs

Dans cette approche par histogramme de couleur [16–18], la principale information utilisée pour décrire un objet est la couleur. Pour le suivi des nageurs, nous générons un

histogramme de couleur pour la référence ainsi que pour la fenêtre glissante qui parcourt la zone de recherche. Pour ce faire, chaque composante (*bin*) de l'histogramme code le nombre de pixels d'un intervalle de couleur donné. Le nombre de composantes, qui correspond aux intervalles de couleurs pris en compte, est un paramètre de la méthode à définir empiriquement. Sachant que nous traitons des images codées sur 8 bits, à savoir 256 valeurs pour chaque pixel (car les images sont en niveaux de gris), nous choisissons de générer des histogrammes de 16 *bins* de 16 valeurs chacune. Chaque *bin* représente le nombre de pixels qui ont une valeur incluse dans son intervalle. Enfin, la recherche de la cible s'effectue, comme les autres techniques basées sur l'histogramme, par une recherche exhaustive dans la zone d'intérêt. La cible, dans ce cas, correspond à la partie qui ressemble le plus à la référence en terme de couleur. Afin de faire une détection pertinente, nous étudions dans la suite les différents types de couleur qui peuvent être utilisés pour différencier la bonne cible du reste de la scène.

3.4.2.1 Types de couleurs

Les images extraites des séquences vidéo acquises sont de type couleur RVB. Chaque image est constituée de trois matrices Rouge, Verte et Bleue. En utilisant ces informations nous pouvons reconstruire l'image avec différents types de couleur et choisir le plus adapté pour notre cas de suivi dans le milieu aquatique. L'image d'intensité (niveaux de gris) est constituée d'une seule matrice obtenue par la moyenne des trois composantes RVB. Le traitement de l'image d'intensité est simple, car il suffit de générer un histogramme pour chaque image. En revanche, Les zones ayant la même intensité et

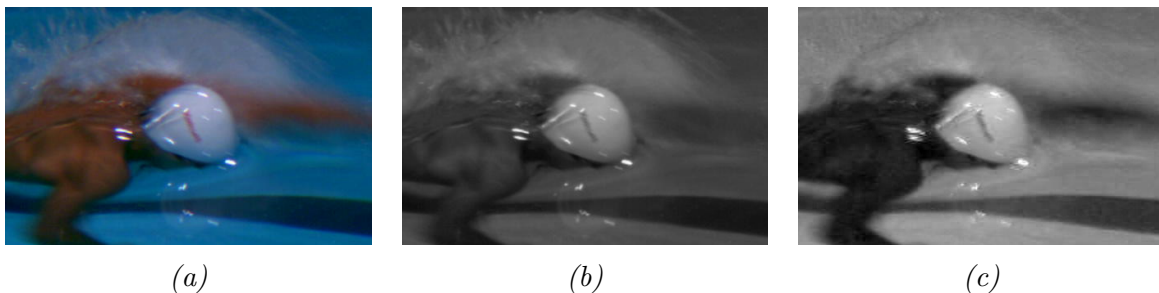


FIGURE 3.9 – Différents types de couleurs utilisé pour la génération de l'histogramme de couleur. (a) Image originale. (b) image d'intensité. (c) Image de la composante bleue.

des couleurs différentes seront regroupés dans un même *bin*. Pour pallier ce problème et avoir une description de couleur plus riche, il est possible de générer, pour chaque image, un histogramme de couleur qui consiste en une concaténation des histogrammes issus de chacune des composantes RVB. Cependant, dans ce cas, l'histogramme de couleur résultant est composé de 48 *bins* ce qui alourdit significativement le temps de calcul. Afin de tirer partie des deux approches, niveaux de gris et couleur, nous nous sommes orientés vers l'utilisation d'une seule matrice qui correspond à la composante bleue. Ce

CHAPITRE 3. SYSTÈME DE SUIVI DE NAGEURS : ANALYSE, CHOIX ET VALIDATION

choix est motivé par les caractéristiques du milieu aquatique, à savoir, l'eau entourant le nageur a une forte composante bleue permettant de la différencier facilement par rapport au nageur.

3.4.2.2 Comparaison des histogrammes de couleurs

Les histogrammes de couleur générés vont être comparés afin de prendre une décision quant au choix de l'image cible qui correspond à la référence dans la zone de recherche. Pour cela, nous calculons le coefficient de Bhattacharyya BC qui mesure le chevauchement et la ressemblance entre deux histogrammes selon la formule :

$$BC = \sum_{i=1}^{NB} \sqrt{(H_1 * H_2)} \quad (3.12)$$

Où NB représente le nombre de *bin* de chacun des deux histogrammes H_1 et H_2 .

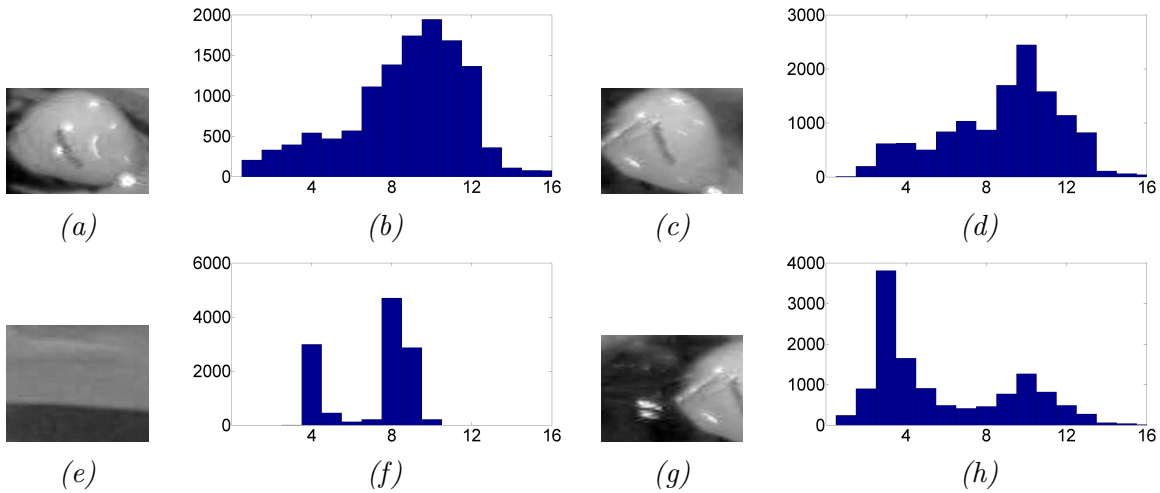


FIGURE 3.10 – Exemple de détection de nageurs basée sur les histogrammes de couleurs calculés à partir de la composante bleue des images. (a et b) Image de référence extraite de l'image (i) de la séquence vidéo et son histogramme de couleur. (c, d) Image cible extraite de l'image (i+1) de la séquence vidéo et son histogramme de couleur. (e, f, g et h) Images de test sélectionnées aléatoirement dans la zone de recherche extraite de l'image (i+1) de la séquence vidéo et leurs histogrammes.

La figure 3.10 présente des histogrammes de couleur générés à partir de quelques images, dont une référence, une cible et deux échantillons aléatoires dans la zone de recherche. Visuellement, nous remarquons une ressemblance plus marquée entre la distribution de l'histogramme de la figure 3.10b et l'histogramme de la figure 3.10d qui correspondent à la référence et la cible respectivement. Cette ressemblance est confirmée

par la valeur du coefficient de Bhattacharyya qui vaut 0.87. Tandis que la valeur de ce coefficient est de 0.4 pour la comparaison entre l'histogramme de la référence et celui de la figure 3.10f et 0.65 pour la comparaison entre l'histogramme de la référence et celui de la figure 3.10h. Finalement, nous pouvons voir que dans ce cas, la bonne cible a été détectée et que le suivi des nageurs par histogramme de couleur est efficace pour ces cas de figure. En revanche, La faiblesse de cette méthode reste la confusion entre les objets ayant la même couleur, même dans le cas où ils ont des formes différentes. Dans ce qui suit, nous étudions les techniques basées sur l'histogramme qui s'intéressent à décrire les informations de texture et de gradient.

3.4.3 LBP pour le suivi des nageurs

Le descripteur de motifs locaux [19, 20] consiste à générer un histogramme permettant de décrire l'information de texture contenue dans une image. Cette technique pour le suivi de nageurs s'applique de la même façon que la technique basée sur l'histogramme de couleur, mais sur des informations différentes. En effet, la génération des histogrammes ne s'effectue pas sur l'image elle-même mais sur une image de texture appelée "image des motifs locaux" ou "image LBP". Cette dernière est générée en calculant, pour chaque pixel de l'image, une valeur qui représente sa relation avec son voisinage (voir chapitre 1 - section 1.6.4).

La figure 3.11 présente les histogrammes LBP issus des mêmes images que celles utilisées pour la comparaison des histogrammes de couleur (voir figure 3.10). Où, la figure 3.11.a représente l'image de référence extraite de l'image (i) de la séquence vidéo. La figure 3.11.d représente l'image cible extraite de l'image (i+1) de la séquence vidéo. Les figures 3.11.g et j sont des images de test sélectionnées aléatoirement dans la zone de recherche extraite de l'image (i+1) de la séquence vidéo. Les figures 3.11.b, e, h et k représentent les images reconstruites après l'application de la technique LBP. Les histogrammes correspondants sont présentés dans les figures 3.11.d, f, i et l.

En faisant une comparaison visuelle entre les images LBP nous remarquons une forte ressemblance entre l'image 3.11.e et l'image de référence 3.11.b. Ce constat est confirmé par la ressemblance des histogrammes correspondants et la valeur du coefficient Bhattacharyya qui est égal 0.75. En revanche, l'image 3.11.h, quant à elle, est bien différente par rapport à la référence avec un coefficient Bhattacharyya de 0.3. Cela est expliqué par le fait que cette image LBP est moins texturée et est issue d'une partie vide du couloir de nage. Finalement, l'histogramme 3.11.l montre une ressemblance significative par rapport à l'histogramme de la référence 3.11.c. Malgré la différence entre les images LBP correspondantes, la valeur du coefficient Bhattacharyya est égale 0.72. Cette valeur est très élevée sachant qu'il s'agit d'une cible majoritairement occultée. Ce problème apparaît à cause de la faiblesse du descripteur LBP dans le cas d'images bruitées causé, dans ce cas, par le mouvement de l'eau.

CHAPITRE 3. SYSTÈME DE SUIVI DE NAGEURS : ANALYSE, CHOIX ET VALIDATION

Pour conclure, l'application de la technique LBP pour le suivi d'objet en général semble avantageux et l'information de la texture est très importante. Cependant, dans notre cas nous avons constaté qu'elle n'est pas suffisamment robuste car les images des nageurs sont très bruitées du fait des perturbations introduites par les spécificités du milieu aquatique.

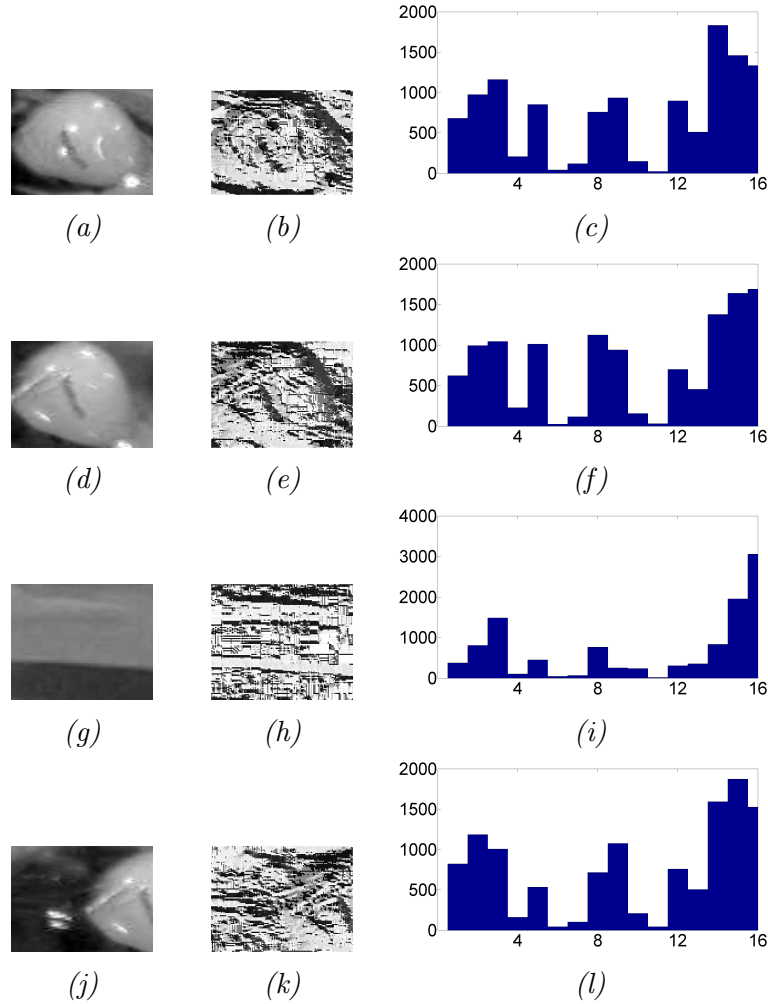


FIGURE 3.11 – Exemple de détection de nageur basée sur les descripteurs LBP. (a, b et c) Image de référence, image LBP correspondante et son histogramme LBP. (d, e et f) Image cible, image LBP correspondante et son histogramme LBP. (g, h, i, j, k et l) Images de test sélectionnées aléatoirement dans la zone de recherche, images LBP correspondantes et leurs histogrammes LBP.

3.4.4 HOG pour le suivi des nageurs

Afin de surmonter la faiblesse du descripteur LBP, deux informations très importantes peuvent être utilisées en supplément de la texture : l'amplitude et l'orientation des contours. Pour cette raison, nous proposons une approche de suivi de nageur par Histogramme de Gradient Orienté [21]. Cette technique consiste à représenter l'image par un histogramme regroupant les gradients selon leurs directions (voir chapitre 1 - section 1.6.3). Pour cela, une image de gradient est générée en appliquant un filtre dérivatif $1-D$ centré, puis l'histogramme de gradient orienté est calculé. Tel que, chaque *bin* représente la somme des amplitudes des gradients ayant une direction incluse dans un intervalle défini. Sachant que la direction du gradient peut varier de 0 à 360° , nous allons fixer le nombre de *bins* à 16 comme nous l'avons fait pour les techniques précédentes. Cela signifie que chaque *bin* concerne un intervalle de 22.5° , ce qui est suffisant pour renforcer la discrimination de la technique.

Afin d'utiliser les descripteurs HOG pour le suivi des nageurs, premièrement, une image de gradient est générée pour l'image de référence ainsi que pour la fenêtre glissante dans la zone de recherche. Ensuite, pour chaque image de gradient, les orientations et les amplitudes des gradients sont calculées et classifiées dans un histogramme appelé HOG. Finalement, la décision est prise en utilisant le coefficient de Bhattacharyya pour comparer les histogrammes.

La figure 3.12 montre un exemple de détection de nageurs basée sur le descripteur HOG. Les résultats obtenus ressemblent à ceux obtenus par le descripteur LBP avec une amélioration au niveau de la discrimination. En effet, nous remarquons une forte ressemblance (coefficient de Bhattacharyya = 0.93) entre les histogrammes 3.12c et 3.12f qui correspondent respectivement à l'image de référence et à la bonne cible. D'autre part, une faible ressemblance (coefficient de Bhattacharyya = 0.15) est obtenue pour l'histogramme 3.12i, et une forte ressemblance (coefficient de Bhattacharyya = 0.84) pour l'histogramme 3.12l, tous deux comparés à la référence. Pour conclure, les informations d'orientation et d'amplitudes des gradients apportées par le descripteur HOG améliorent légèrement la discrimination. En revanche, cette technique reste sensible au bruit et à la déformation de l'objet à suivre et donc difficile à mettre en place dans le cadre du suivi des nageurs.

3.4.5 Avantages et limitations des techniques basées sur les histogrammes

Les avantages et les limitations des techniques basées sur les histogrammes pour le suivi de nageurs dépendent des caractéristiques du milieu aquatique ainsi que de la nature du descripteur utilisé. Les techniques basées sur l'histogramme de couleur sont

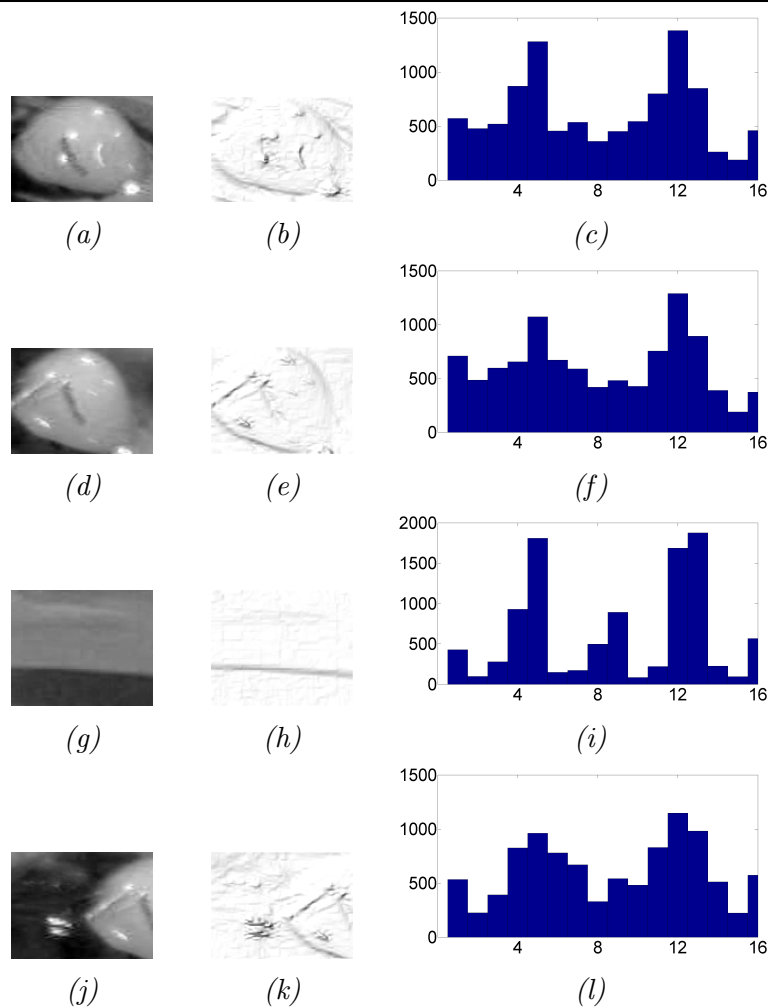


FIGURE 3.12 – Exemple de détection de nageur basée sur les descripteurs HOG. (a, b et c) Image de référence, image de gradient correspondante et son histogramme HOG. (d, e et f) Image cible, image de gradient correspondante et son histogramme HOG. (g, h, i, j, k et l) Images de test sélectionnées aléatoirement dans la zone de recherche, images de gradient correspondantes et leurs histogrammes HOG

très fiables dans le cas où la couleur du bonnet du nageur est différente de celle des objets alentour (eau, bouchons, peau, ...). Dans le cas inverse, cette technique donne de faibles résultats.

Cependant, les techniques basées sur les descripteurs LBP et HOG se montrent très efficaces dans le cas où l'objet à suivre conserve sa forme. Ceci n'est pas forcément le cas pour un nageur dans un bassin, car la tête du nageur peut subir des occultations et du bruit liés aux éclaboussures entraînant une baisse des performances de ces deux techniques utilisant principalement l'information de texture.

Enfin, un inconvénient commun à ces techniques vient de la recherche exhaustive de la cible dans la zone d'intérêt qui alourdit les calculs. Afin de pallier ce problème, nous proposons d'utiliser une adaptation, pour les histogrammes, des images intégrales proposée par Viola et Jones [76]. Cela permet de ne pas recalculer l'intégralité de l'histogramme lorsque deux zones ont un recouvrement important.

3.5 Conclusion

Dans ce chapitre nous avons proposé un système complet de suivi de nageurs. Où, nous avons détaillé deux types d'approches de suivi : les approches basées sur la corrélation NL-JTC et les approches basées sur les histogrammes. Pour la première, une étude approfondie a été faite sur le plan de corrélation. Ceci nous a menés à apporter des contributions importantes, notamment, la réduction du plan d'entrée, l'extraction de la zone potentielle de corrélation et le débruitage du plan de corrélation. Ces contributions ont permis d'optimiser la technique NL-JTC et de l'adapter à notre cas de suivi de nageurs. Suite à quelques faiblesses de cette technique, basée seulement sur l'information de contour, nous avons proposé des approches de suivi de nageurs à base d'histogrammes de couleur, LBP et HOG.

Enfin, nous avons conclu que chacune des approches proposées pour le suivi de nageurs montre des avantages importants et des limitations envers des points spécifiques. Particulièrement, les approches basées sur la corrélation NL-JTC ont montré de bons résultats en terme de précision pour la localisation, mais elles restent sensibles par rapport à la rotation et à la déformation des contours. En revanche, les approches basées sur l'histogramme de couleurs ont montré une certaine robustesse pour le suivi des nageurs mis à part dans les cas d'une confusion éventuelle avec des objets de la même couleur. Cette méthode souffre aussi d'imprécisions pour la localisation.

Par conséquent, nous allons, par la suite, proposer des systèmes optimisés de suivi des nageurs basés sur la fusion de ces données en étudiant profondément la complémentarité qui existe entre ces techniques.

Chapitre 4

Système de suivi de nageurs : Optimisation

Sommaire

4.1	Introduction	94
4.2	Prédiction de la trajectoire	95
4.2.1	Localisation du couloir de nage	95
4.2.2	Direction de nage	96
4.2.3	Vitesse moyenne du nageur	96
4.2.4	Position référentielle validée	97
4.2.5	Discussion	97
4.3	Approche de suivi multipiste	98
4.3.1	Principe du suivi multipiste	98
4.3.2	Approche de suivi multipiste uni-descripteur	99
4.3.3	Approche de suivi multipiste multidescripteur	100
4.3.4	Discussion	101
4.4	Approche de suivi par fusion dynamique	102
4.4.1	Inspiré du principe de débruitage	102
4.4.2	Principe de la fusion dynamique	103
4.4.3	Génération des plans de scores	103
4.4.4	Normalisation des scores	104
4.4.5	Fusion des plans	105
4.4.6	Analyse du plan de fusion : précision et robustesse	105
4.4.7	Facteur de confiance	106
4.4.8	Expérimentation sur des images synthétiques	106
4.4.9	Discussion	109
4.5	Approche de suivi multizone	109

4.5.1	Inspiré par l'idée du multipiste	110
4.5.2	Processus du suivi multizone	111
4.5.3	Détections potentielles	112
4.5.4	Critère de décision complexe	112
4.5.5	Discussion	117
4.6	Conclusion	118

4.1 Introduction

Grâce au système décrit dans le chapitre précédent, il est possible de suivre automatiquement un nageur. Cependant, il est nécessaire, à ce stade, de l'optimiser afin d'assurer la précision de l'évaluation des nageurs. Les choix des techniques précédentes ont été faits en étudiant leurs performances par rapport à notre sujet : le suivi des nageurs. Dans ce cadre, nous avons choisi, implémenté, optimisé et testé les techniques suivantes : NL-JTC (*Non-linear Joint Transform Correlator*), Histogramme de couleurs, LBP (*Local Binary Patterns*), HOG (*Histogram of Oriented Gradient*). Cependant, nous avons constaté que chacune de ces techniques avait des avantages et des limitations liées principalement au fait que chaque approche se base sur un seul descripteur : contour, couleur, texture. Par conséquent, nous proposons de nouvelles approches de suivi basées sur la fusion de données.

Dans ce chapitre nous allons détailler nos différentes contributions permettant d'optimiser notre système de suivi des nageurs.

1. Nous commencerons par l'optimisation du module de prédiction de la trajectoire qui a pour but de maintenir la zone d'intérêt autour du nageur. Cela nous permet d'éviter la perte de la région d'intérêt mais aussi de parvenir à la génération d'une cible valide qui contient toujours la tête du nageur à suivre.
2. Puis, nous proposons une approche multipiste qui consiste à la fusion des décisions issues de plusieurs techniques de suivi.
3. Ensuite, nous présenterons l'approche de fusion dynamique des plans de scores permettant de décrire un nouveau descripteur optimisé, basé sur une fusion des informations de couleur et de contour de la référence et de la cible.
4. Enfin, nous détaillerons l'approche multizone qui a pour objectif de surmonter le problème de l'occultation partielle ou totale de la tête du nageur. Le principe de base de cette approche consiste à suivre en parallèle deux ou plusieurs sous-cibles (dans notre cas : la tête et le maillot de bain) liées entre elles par des relations géométriques (distance, angle...) bien définies permettant de retrouver la cible occultée.

4.2 Prédiction de la trajectoire

Le système de suivi de nageurs que nous avons détaillé dans le chapitre précédant est bien adapté aux conditions d'entraînements où le nageur avance lentement. En effet, le nageur garde souvent la tête hors de l'eau et les éclaboussures sont réduites nous permettant de suivre le nageur et d'analyser ses performances avec plus de précision. Cependant, dans le cas général, comme le montre la figure 4.1, nous serons face à des courses rapides avec de nombreuses éclaboussures et d'importantes occultations qui engendreront des décrochages et des imprécisions des résultats.

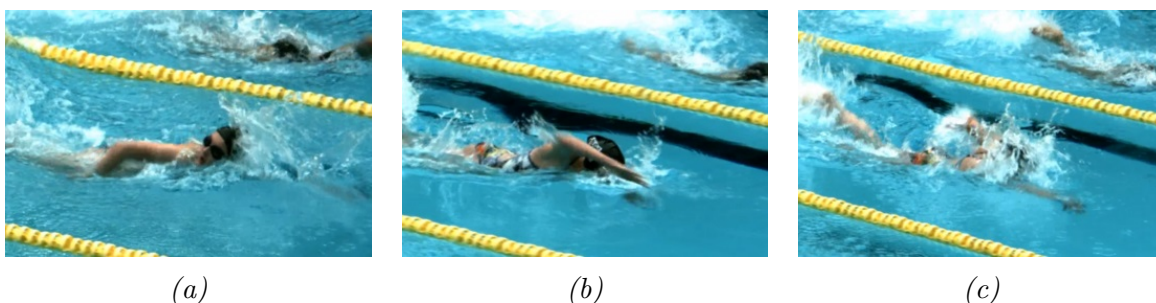


FIGURE 4.1 – Exemples des difficultés du suivi de nageurs. (a) rotation de la tête (b) couleur du fond similaire à celle de la cible (bonnet) (c) éclaboussures et occultations.

Afin de pallier ces problèmes, nous proposons un module optimisé pour la prédiction de la position du nageur. Ce module ne sert pas uniquement à pré-localiser et restreindre la zone de recherche de la cible mais permet aussi d'estimer la position du nageur en cas d'occultation. La prédiction se fait en prenant en compte les différentes propriétés de l'environnement aquatique obtenues grâce à la calibration qui nous permet de connaître la correspondance entre les coordonnées pixels dans l'image et les coordonnées métrique réelles. Afin d'assurer une prédiction pertinente, nous allons étudier les différents critères qui aident à prédire la position du nageur tout en simulant la perception humaine qui construit une vision globale du bassin et particulièrement du couloir concerné, ensuite, nous pré-localiserons le nageur selon sa vitesse moyenne estimée comme présenté dans la figure 4.2.

4.2.1 Localisation du couloir de nage

Dans le chapitre 2 section 2.2, nous avons présenté une description générale des bassins olympiques qui adhère aux normes de l'organisation internationale FINA. En particulier, nous avons présenté les dimensions des bassins, des couloirs et des différents marqueurs présents dans celui-ci.

La connaissance de ces informations nous permet de localiser automatiquement le couloir contenant le nageur à suivre. Ce dernier est défini lors de la phase d'initialisation du suivi en sélectionnant le nageur à suivre. L'extraction du couloir concerné se fait selon la formule suivante :

$$\begin{cases} y_1 = \left\lfloor \frac{y}{l} \right\rfloor * l \\ y_2 = (\left\lfloor \frac{y}{l} \right\rfloor + 1) * l \end{cases} \quad (4.1)$$

avec, y la position initiale du nageur par rapport à l'axe des ordonnées sachant que l'origine (0,0) correspond au coin haut/droit du bassin. $l = 2.5$ m représente la largeur du couloir. y_1 et y_2 représentent respectivement les coordonnées y des deux lignes délimitant le couloir contenant le nageur concerné.

4.2.2 Direction de nage

La direction de nage est une information indispensable pour la prédiction de la prochaine position du nageur dans le bassin. Elle consiste en une valeur binaire désignant le sens de nage. La position du début de la nage peut définir cette direction selon sa distance par rapport au bord. Ceci n'est valable que dans le cas où le système de suivi est appliqué au début de la course. En effet, l'application du système peut être faite sur des séquences vidéo de nage choisies dans des moments précis de la course. Par conséquent, nous avons besoin de généraliser notre fonction de la recherche de direction de nage. Pour cela, nous utilisons la formule suivante :

$$s = \text{sign}(x_0 - x_1) \quad (4.2)$$

avec, x_0 et x_1 qui représentent respectivement la position initiale du nageur et sa position détectée après i images par rapport à l'axe des abscisses. i est à choisir suffisamment grand pour éviter les imprécisions de détections qui peuvent falsifier la direction. En effet, durant ces n premières images la région d'intérêt est figée jusqu'à ce que nous déterminions la direction de nage pour qu'elle puisse ensuite se décaler selon la prédiction. Afin de répondre à ces contraintes, nous choisissons $i = 5$. s représente la direction de nage : de gauche à droite si la valeur de s est négative et de droite à gauche sinon.

4.2.3 Vitesse moyenne du nageur

Le record de France et d'Europe en 50 m nage libre messieurs est détenu par Frédérick Bousquet lors des championnats de France d'avril 2009. Il a établi une performance de 20.94 s ce qui est équivalent à une vitesse moyenne de 2.39 m/s. Cette vitesse

peut varier selon le type de nage et le niveau du nageur. Pour cette raison, durant les dix premières images, nous estimons la vitesse moyenne du nageur approximativement par rapport au type de nage dans une intervalle de $[1, 2m/s]$. Ensuite, elle sera ajustée au fur et à mesure du suivi. La vitesse moyenne consiste en la distance parcourue entre les positions x_0 et x_i par le nageur dans une durée δt , comme le montre la formule suivante :

$$V_{moy} = \frac{|x_0 - x_i|}{\delta t} \quad (4.3)$$

Ainsi, après chaque itération, nous actualisons la vitesse moyenne selon la position du nageur détecté et nous estimons la distance qui sera parcourue dans la prochaine itération.

4.2.4 Position référentielle validée

La première position référentielle validée (X_r, Y_r) est celle qui est sélectionnée à l'initialisation du suivi, voir chapitre 2. Cette position sera actualisée durant le suivi dans le cas où le critère de décision de la détection dépasse un certain seuil de confiance. En combinant toutes ces informations, la zone d'intérêt peut être prédite à la position (X_p, Y_p) par la formule suivante :

$$\begin{cases} X_p = X_r + k * V_{moy} * \frac{1}{T} - \frac{W}{2} \\ Y_p = Y_r - \frac{H}{2} \end{cases} \quad (4.4)$$

Sachant que w et h sont la longueur et largeur de la référence sélectionnée. X_r et Y_r sont les coordonnées de la dernière position référentielle validée. k représente le nombre d'images depuis la dernière position référentielle validée. $H = 3 * h$ et $W = 5 * w$ représentent respectivement la longueur et la largeur de la zone d'intérêt.

4.2.5 Discussion

La technique de prédiction proposée dans cette section est très utile pour la restriction de la zone de recherche autour de la cible à suivre et ainsi éviter le décrochage permanent du suivi. De plus, elle permet de prédire la position de la cible dans le cas où le facteur de confiance de la détection est faible, Par exemple, dans le cas d'une déformation ou d'une occultation de la cible à suivre. Cependant, cette prédiction reste globale et afin de faire face aux différentes difficultés du suivi de nageurs, nous proposons par la suite des techniques dédiées à l'optimisation de l'approche de suivi.

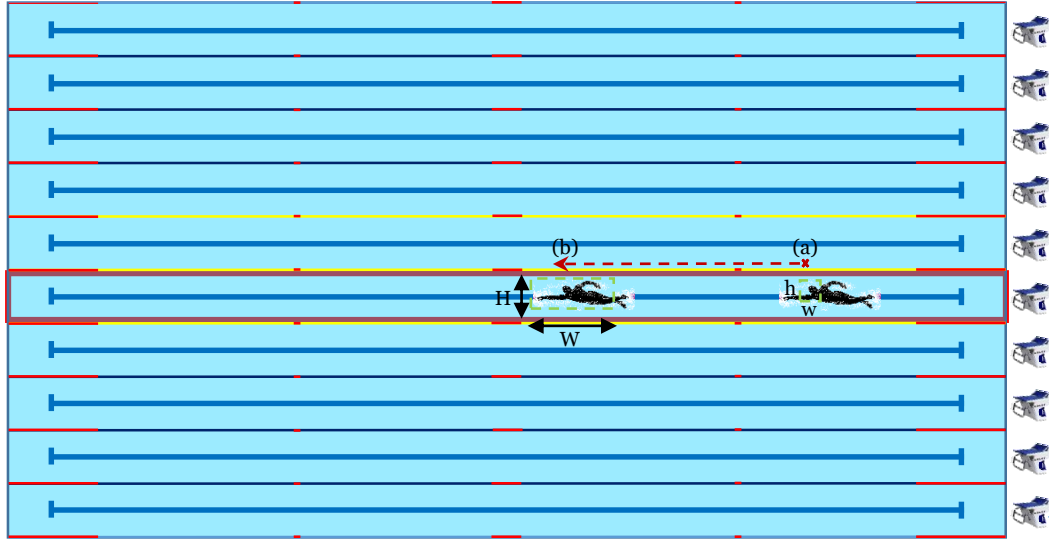


FIGURE 4.2 – *Prédiction de la position du nageur en tenant compte des informations suivantes : son couloir, sa direction de nage, sa dernière position référentielle validée et sa vitesse moyenne. (a) Dernière position référentielle validée. (b) Position prédite après une durée donnée.*

4.3 Approche de suivi multipiste

Après une analyse approfondie des vidéos de compétitions de natation mises à notre disposition par la FFN, nous avons noté plusieurs difficultés liées aux spécificités du milieu aquatique. A titre d'exemple, on peut citer le mouvement de l'eau, les éclaboussures, la variation de l'apparence du nageur et les occultations. Afin de surmonter ces problèmes, nous avons proposé dans le chapitre 3 des systèmes de suivi de nageur à base de techniques connues, notamment NL-JTC, histogramme de couleur, LBP et HOG. Chacune de ces techniques a montré des avantages et des inconvénients. Par conséquent, nous proposons un système multipiste qui tire parti des points forts des quatre techniques citées précédemment pour construire un système de suivi de nageur robuste.

4.3.1 Principe du suivi multipiste

Dans le but d'améliorer les résultats du suivi, nous proposons une approche multipiste (multi-descripteurs) qui consiste à combiner plusieurs techniques de reconnaissance pour renforcer la précision et la robustesse du suivi. Le choix des techniques à utiliser pour le suivi multipiste vient de la nécessité d'une description globale de la référence

à suivre. Nous avons choisi l'histogramme de couleur pour avoir une description robuste de celle-ci, le corrélateur optique NL-JTC pour décrire la forme, LBP pour la description de la texture et HOG pour renforcer la description des contours. Dans ce nouveau système, l'image $i_{0,0}$ est utilisée comme référence pour lancer plusieurs pistes de suivi parallèlement. Chaque piste s'exécute indépendamment durant n itérations (empiriquement $n = 5$), puis la référence est actualisée en choisissant la meilleure cible pour relancer la prochaine itération du processus de suivi multipiste.

4.3.2 Approche de suivi multipiste uni-descripteur

Dans le chapitre précédant, nous avons adapté et optimisé plusieurs techniques pour notre cas de suivi de nageurs. En particulier, la technique NL-JTC a montré de bonnes performances mais nous avons souligné des décrochages possibles, notamment quand la cible est partiellement occultée. Afin de mieux comprendre les raisons de ces décrochages, nous avons étudié profondément le plan de corrélation dans ces cas de figure. En analysant le plan de corrélation issu de la technique NL-JTC avec un coefficient de non-linéarité $k = 0.3$, nous avons constaté que le plan est parfois bruité, spécialement dans le cas de déformations du contour de la cible. Ceci diminue la hauteur du "vrai pic" de corrélation et donne la possibilité aux autres pics d'être plus importants que lui. Ces "faux pics" correspondent généralement à des zones ayant un contour similaire à celui de la vraie cible. Par exemple, dans la figure 4.3, la vraie détection correspond au troisième plus haut pic alors que les autres correspondent respectivement au bras et au bas du dos du nageur.

Cette observation nous a motivé à proposer une approche multipiste uni-descripteur basée sur la technique de corrélation NL-JTC. L'idée de base consiste à retrouver les m zones correspondant aux m plus hauts pics. Dans notre cas, il est suffisant de prendre $m = 4$ pour avoir une forte probabilité de contenir la vraie détection. Afin de prendre une décision sur les m zones, nous proposons de comparer les histogrammes de couleurs des n images avec celui de l'image référence. Comme nous l'avons précisé dans le chapitre précédant, et afin de profiter des caractéristiques du milieu aquatique, nous calculons des histogrammes de la composante bleue des images.

Ensuite, nous comparons ces histogrammes en utilisant la mesure de ressemblance de Bhattacharyya dans le but de choisir, parmi les m zones, celle qui ressemble le plus à la référence. Cette approche permet d'optimiser le suivi basé sur la technique de corrélation NL-JTC car nous ne nous limitons pas à la détection du plus grand pic mais nous vérifions aussi les autres possibilités dans les cas où des différences, liées au bruit, existent entre l'image de référence et la cible.

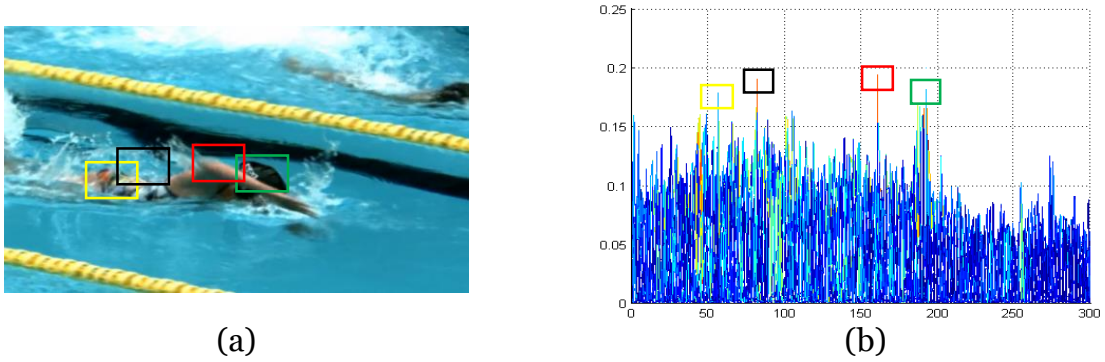


FIGURE 4.3 – Exemple d’une détection par une approche multipiste uni-descripteur basée sur la technique de corrélation NL – JTC. (a) Image cible contenant différentes détections possibles. (b) Plan de corrélation avec des marqueurs précisant les quatre plus hauts pics. Dans les deux images, le marqueur vert désigne la bonne détection

4.3.3 Approche de suivi multipiste multidescripteur

Le principe de l’approche multipiste uni-descripteur basée sur la corrélation est très intéressant car nous vérifions plusieurs détections potentielles. En revanche, nous pouvons optimiser encore plus cette approche en utilisant la richesse d’autres descripteurs comme le LBP et le HOG. Le processus de cette approche de suivi multipiste multidescripteur est illustré dans la figure 4.4. L’image de référence $i_{0,0}$ est comparée avec l’image cible selon trois techniques : Corrélation NL-JTC, LBP et HOG. Sachant que nous avons choisi de vérifier deux pics de corrélation, nous avons deux pistes pour la corrélation, une piste pour le LBP et une piste pour le HOG. Les images $i_{l,m}$ sont les détections potentielles de la tête du nageur dans chaque image cible, avec $l > 0$ le numéro de l’image de la séquence vidéo et $m > 0$ le numéro de la piste. Pour chaque piste basée sur la corrélation, la référence est actualisée selon un filtre composite de deux images : la première référence et la dernière cible détectée et validée par une comparaison des histogrammes de couleur.

Pour les deux autres pistes, la référence est actualisée par la dernière cible détectée et validée. En effet, chaque piste est considérée comme un processus de suivi indépendant durant n itérations, dans notre cas, nous avons fixé $n = 5$. Ensuite, après le choix de la piste valide, le processus multipiste est relancé avec la nouvelle référence. Le critère mis en place pour la validation d’une piste donnée consiste à analyser l’historique des m pistes, puis choisir celle ayant les meilleures détections. Ce choix est fait en calculant pour chaque piste la moyenne des coefficients de Bhattacharyya issus de la comparaison des histogrammes de couleur des cibles détectées avec la première référence. La piste ayant la plus grande moyenne sera validée et la nouvelle référence sera choisie parmi les n cibles détectées selon la valeur du coefficient de Bhattacharyya. Une fois cette

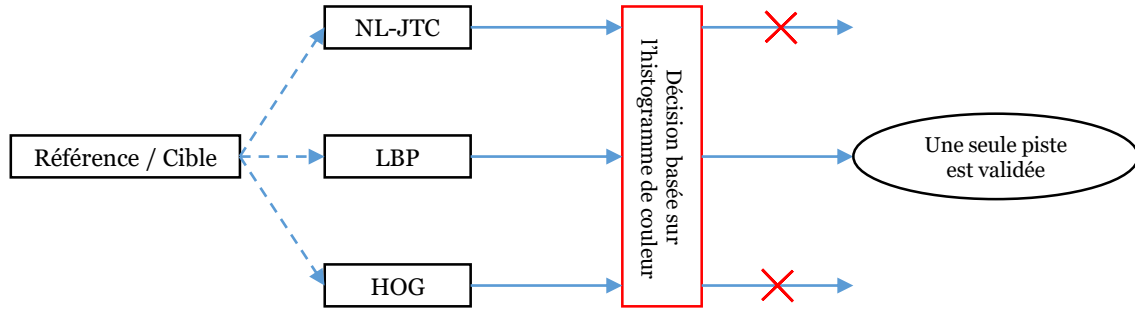


FIGURE 4.5 – *Limitation de l’approche multipiste : Fusion au niveau des décisions, une seule technique (piste) est validée alors que les autres sont rejetées. Exemple : si la piste LBP est validée alors les résultats des deux autres techniques seront perdus.*

4.4 Approche de suivi par fusion dynamique

Au même titre que l’approche multipiste qui consiste à faire une fusion au niveau des décisions, nous voulions proposer une approche de fusion de données plus robuste. Pour cela, nous nous sommes intéressés à la complémentarité qui existe entre la corrélation basée sur le JTC et l’histogramme de couleur. La corrélation a montré de bonnes performances en terme de précision pour la localisation. En revanche, elle reste sensible par rapport à la rotation et à la déformation de contours. D’autre part, la technique basée sur l’histogramme de couleur a montré une robustesse pour le suivi de nageur dans le cas de cible ayant des couleurs différentes du reste de la scène. Par contre, une confusion peut être faite entre les objets de même couleur en plus d’un manque de précision en terme de localisation. Pour pallier ces problèmes, et dans le but de concevoir un système de suivi de nageur robuste, nous proposons de fusionner ces deux méthodes.

4.4.1 Inspiré du principe de débruitage

Dans les travaux de Alfalou et al. [105], une fonction non-linéaire est appliquée au plan de corrélation afin de le débruiter, dans le même esprit nous proposons d’améliorer la détection en optimisant la décision. Dans ces travaux, la fonction non-linéaire utilisée augmente le pic de corrélation et atténue le bruit permettant de ce fait d’améliorer la valeur de leur critère de décision basé sur le PCE (*Peak to Correlation Energy*).

Cependant, cette optimisation est aveugle, i.e. elle ne prend pas en compte les informations liées à l’objet à suivre, et ne permet en aucun cas d’influencer la décision. Afin de pallier cette faiblesse, nous proposons ici une nouvelle technique permettant de fu-

sionner directement les plans de probabilités issues de différentes méthodes (corrélation, histogramme de couleur...). Cette fusion peut être vue comme un filtrage dynamique du plan de corrélation car la fonction utilisée, i.e. le plan de scores couleur, dépend de l’objet suivi et s’adapte à ses variations et à son évolution. Le plan dynamique contenant l’information de ressemblance entre la référence et la cible, en terme de couleur, renforce l’information de contour issue de la corrélation et améliore de ce fait la décision finale.

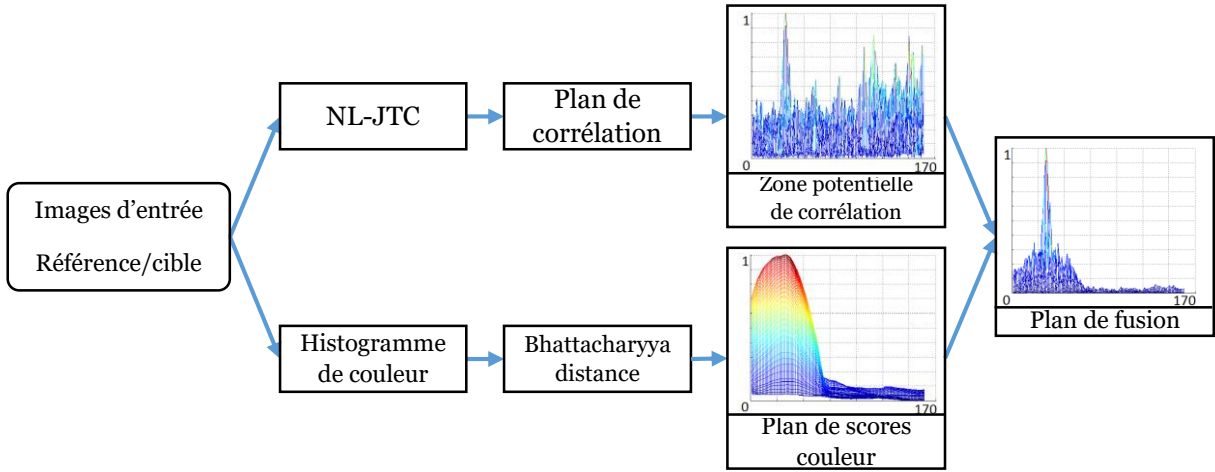
4.4.2 Principe de la fusion dynamique

Selon leurs particularités, les deux approches choisies ne peuvent pas être fusionnées directement. En effet, il est nécessaire que les plans de scores aient une taille identique ainsi qu’une dynamique commune, entre 0 et 1 par exemple. Partant du constat que le plan de corrélation contient deux zones identiques (en haut à gauche et en bas à droite) ainsi que deux autres zones non porteuses d’informations (en haut à droite et en bas à gauche), nous avons proposé (voir chapitre 3 section 3.3.3.1) d’extraire un sous plan, que nous avons appelé “zone potentielle” de corrélation, de même taille que le plan de scores couleur. Ainsi, seule l’étape de normalisation de la dynamique des plans restera à définir pour pouvoir appliquer notre fusion dynamique. Le fonctionnement de notre approche de fusion est résumé dans la figure 4.6.

Pour la corrélation, la référence et la cible sont placées dans le plan d’entrée, sur lequel, la technique NL-JTC est appliquée. Ensuite, la zone potentielle de corrélation est alors extraite. Pour l’histogramme de couleur, une fenêtre parcourt la cible en faisant des comparaisons avec la référence en utilisant la distance de Bhattacharyya pour comparer les deux histogrammes. Les scores obtenus forment alors le plan de scores couleur. Enfin, les deux plans sont alors fusionnés en un seul plan décrivant la ressemblance entre la référence et la cible à la fois en termes de contour, mais aussi de couleur.

4.4.3 Génération des plans de scores

Dans le chapitre 3 section 3.3.3.1, nous avons proposé d’extraire un sous-plan à partir du plan de corrélation complet. Cette zone est appelée zone potentielle de corrélation et elle représente la zone exacte de ressemblance entre l’image de référence et l’image cible. Chaque valeur $S_1(x, y)$ dans cette zone peut être considérée comme un score de ressemblance entre l’image de référence et une fenêtre correspondante ayant les coordonnées (x, y) dans l’image cible. D’autre part, Le plan des scores couleurs est généré en faisant des comparaisons entre la référence et une fenêtre glissante qui parcourt toute la cible. Pour chaque comparaison, un score $S_2(x, y)$ est calculé et il représente la valeur du coefficient Bhattacharyya de la comparaison entre les histogrammes de couleur de


 FIGURE 4.6 – *Processus de fusion du plan de la corrélation et du plan de scores couleur.*

la référence et la fenêtre ayant les coordonnées (x, y) dans l'image cible. Maintenant, nous avons deux plans de scores de même taille qui décrivent la ressemblance entre la référence et la cible en termes du contour et de la couleur. Enfin, afin de pouvoir effectuer notre algorithme de fusion dynamique, les plans doivent être normalisés pour qu'ils partagent un espace de valeurs commun.

4.4.4 Normalisation des scores

Les scores S_1 issus de la corrélation varient dans l'intervalle $[0, \infty[$. En revanche, les scores couleurs S_2 , qui sont à la base des coefficients de Bhattacharyya, ont des valeurs dans l'intervalle $[0, 1]$. Pour les deux techniques, les scores S_1 et S_2 sont des scores de ressemblance, à savoir que le degré de ressemblance croît avec ce score (0 indique qu'il n'y a pas de ressemblance alors que la valeur maximale indique qu'il y a une ressemblance totale). Avant de fusionner les deux plans de scores, il est donc nécessaire de les normaliser afin qu'ils partagent une dynamique commune. Pour cela, nous normalisons les scores de corrélation S_1 dans le même intervalle que les scores de couleur S_2 . Notre choix de l'intervalle $[0, 1]$ est essentiellement motivé par notre besoin de générer un score pertinent qui nous permet de mesurer et évaluer la ressemblance entre la référence et la cible. Pour normaliser les scores issus de la corrélation, nous appliquons la technique de normalisation Min-Max détaillée dans le chapitre 1 section 1.7.3. Ainsi, chaque valeur du plan de scores de corrélation est normalisée selon la formule suivante :

$$S_{1_{nor}}(x, y) = \frac{(S_1(x, y) - \min(S_1))}{(\max(S_1) - \min(S_1))} \quad (4.5)$$

Avec, $S_{1_{nor}}(x, y) \in [0, 1]$ le score de corrélation normalisé et (x, y) sa position dans le plan.

4.4.5 Fusion des plans

Après avoir normalisé le plan de scores de corrélation dans l'intervalle $[0, 1]$, les deux plans sont prêts à être fusionnés. Pour ce faire, nous procédons à la fusion par produit comme le montre la formule suivante :

$$S_{fus}(x, y) = S_{1_{nor}}(x, y) * S_2(x, y) \quad (4.6)$$

Cette fusion consiste à filtrer le plan de corrélation par le plan de scores de couleur. Ce dernier est considéré comme une fonction de filtrage dynamique qui filtre le plan de corrélation prenant en compte la ressemblance de couleurs entre la référence et la cible dans chaque image de la séquence vidéo, d'où l'appellation de l'approche : "fusion dynamique". Les figures 4.6 et 4.7 montrent le fonctionnement de la méthode ainsi que le plan résultant de la fusion.

4.4.6 Analyse du plan de fusion : précision et robustesse

La précision et la robustesse sont deux notions importantes dans les techniques de détection et de suivi. En effet, la première permet une localisation précise de la référence dans l'image cible, tandis que la seconde permet d'avoir une détection dans tous les cas de figures (rotation, bruit, occultations...). Le plus souvent, ces deux notions s'opposent et il est difficile de combiner précision et robustesse avec une seule méthode. En effet, pour la corrélation basée sur l'approche NL-JTC, seul un compromis entre ces deux notions peut être obtenu grâce au coefficient de non-linéarité k . Dans le cas de notre application au suivi de nageur, le meilleur compromis est obtenu avec $k \in [0.3, 0.4]$. Il permet d'avoir une localisation précise mais qui reste peu robuste face à la rotation et aux occultations (bruit dans le plan de corrélation), voir figure 4.7.a.

D'autre part, pour l'approche basée sur les histogrammes de couleur, le plan obtenu est plus homogène avec une zone de valeurs importantes à l'endroit de la ressemblance entre la référence et la cible et un bruit négligeable. De ce fait, la robustesse de cette technique est importante mais la précision de la localisation est insuffisante (largeur de la zone de ressemblance), voir la figure 4.7.b. Finalement, le plan obtenu grâce à notre approche de fusion montre un plan homogène avec un bruit fortement atténué et un pic élevé assurant une localisation précise de l'objet à détecter. Ce compromis entre robustesse et précision permet donc de tirer profit des informations issues des deux approches, à savoir les informations de contours et de couleur, voir figure 4.7.c.

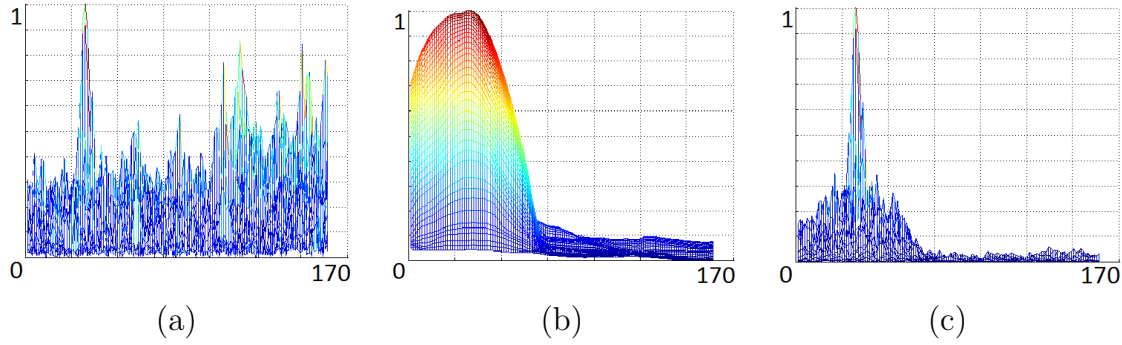


FIGURE 4.7 – Fusion de la zone potentielle du plan de corrélation avec le plan des scores couleurs. (a) – zone potentielle du plan de corrélation obtenu avec la technique NL-JTC ($k = 0.2$). (b) – plan de scores couleurs. (c) – plan obtenu grâce à notre technique de fusion dynamique.

4.4.7 Facteur de confiance

Dans le plan résultant de fusion dynamique, chaque valeur représente un score de détection. Ceci permet de mesurer la ressemblance, en termes de contour et de couleur, entre la référence et la fenêtre ayant la position (x, y) parcourant l'image cible. Afin de localiser la position de la tête du nageur à suivre (x_p, y_p) , il est suffisant de trouver la position du pic maximum dans le plan de fusion. Il est à noter que la hauteur de ce pic FC peut être considérée comme une mesure de confiance envers la décision. Celle-ci varie entre 0, dans le cas où il n'y a pas de ressemblance, et 1 dans le cas où il y a une ressemblance parfaite. Ces mesures sont calculées en utilisant les formules suivantes :

$$(x_p, y_p) = \underset{x, y}{\operatorname{argmax}} S_{fus}(x, y) \quad (4.7)$$

$$FC = S_{fus}(x_p, y_p) \quad (4.8)$$

4.4.8 Expérimentation sur des images synthétiques

Afin de valider et de montrer les avantages de notre technique de fusion, un ensemble de tests synthétiques a été établi. Comme le montre la figure 4.8, l'image cible créée contient trois formes de deux couleurs différentes (rectangle, triangle et cercle). Pour illustrer les résultats, différentes références ont été choisies afin de présenter le fonctionnement de la méthode et l'intérêt de notre approche par fusion dynamique. Le but de ces expérimentations est de montrer l'apport de notre approche de fusion dynamique vis-à-vis du corrélateur NL-JTC et des histogrammes de couleur. Pour cela,

nous présentons dans le tableau 4.1 la détection de trois formes, un cercle rouge, un triangle vert et une ellipse rouge (déformation du cercle rouge), dans une même image cible. Pour évaluer la pertinence de la détection pour chacune des trois approches, nous recherchons la forme correspondant au pic de hauteur maximale dans le plan de sortie (corrélation, couleur ou fusion dynamique). Afin de mesurer la confiance envers la détection, nous regardons la hauteur du pic maximal après avoir normalisé le plan dans l'intervalle $[0, 1]$.



FIGURE 4.8 – Image cible synthétique utilisée pour évaluer la pertinence de notre approche de fusion dynamique. Trois formes et deux couleurs différentes sont utilisées afin de tester la robustesse de notre approche.


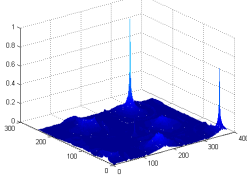
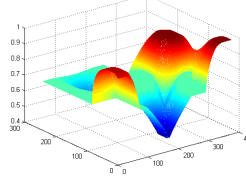
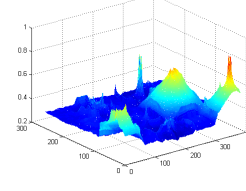

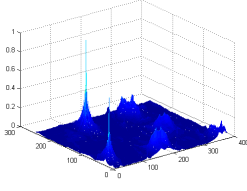
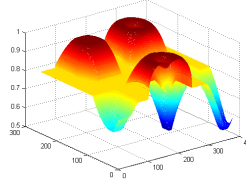
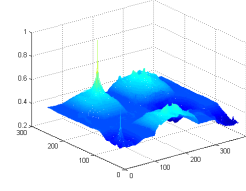

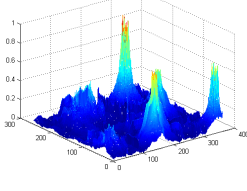
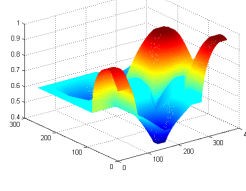
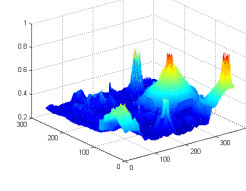
Références	Plan de corrélation	Plan de scores couleur	Plan de fusion
			
Cercle rouge	0.7 - fausse détection (Cercle vert détecté)	1 - bonne détection	0.84 - bonne détection
			
triangle vert	1 - bonne détection	1 - bonne détection	1 - bonne détection
			
ellipse rouge	0.7 fausse détection (Cercle vert détecté)	0.9 fausse détection (Rectangle rouge détecté)	0.85 bonne détection

TABLE 4.1 – *Evaluation des trois techniques de détection : NL-JTC, histogramme de couleur et fusion dynamique. Trois références sont recherchées dans l'image cible de la figure 4.8. La hauteur des pics permet d'évaluer la confiance envers la détection. Les meilleurs résultats apparaissent dans la dernière colonne qui correspond à la fusion dynamique.*

Les expérimentations présentées dans le tableau 4.1 mettent en évidence la pertinence de notre nouvelle approche de fusion dynamique. En particulier, il est intéressant de noter que :

- La corrélation montre de bons résultats dans le cas du triangle vert mais les références de couleurs rouges donnent de fausses détections. Ceci peut être expliqué par l'importance du contour (contraste) dans le cas vert/noir contrairement au cas rouge/noir.
- Le plan de scores couleur est presque identique pour les références rouges (cercle et ellipse). Cela vient du fait que cette méthode est basée uniquement sur la couleur et non sur la forme.
- Le résultat de la détection est significativement amélioré grâce à la fusion « dynamique ». On peut noter que le plan de fusion est considérablement amélioré et les faux pics atténués.

- La hauteur du pic, dans le cas de la fusion, peut être considérée comme une mesure de confiance vis-à-vis de la détection, voir la section 4.4.7. En effet, si la hauteur du pic vaut 1, alors la corrélation et l'histogramme de couleur ont détecté le même objet, et la confiance est maximale. En revanche, si les méthodes sont en désaccord, les pics des deux plans ne sont pas superposés donnant un pic de fusion inférieur à 1, et donc une confiance moins importante.

4.4.9 Discussion

Ces derniers tests ont montré l'intérêt de notre approche de fusion dynamique, vis-à-vis du corrélateur NL-JTC et de l'histogramme de couleur, pour la détection dans des images synthétiques. Nous appliquerons, dans le chapitre suivant, l'ensemble de notre méthode au cas réel du suivi des nageurs. Dans la suite de ce chapitre nous présenterons nos contributions concernant la résolution des problèmes d'occultations, partielles et totales, non résolus par cette approche. Pour cela nous introduirons le concept du suivi multizone qui est une amélioration de notre approche de suivi multipiste.

4.5 Approche de suivi multizone

Dans la partie précédente, nous avons retenu la tête du nageur comme la partie du corps à suivre. Ce choix est motivé par la rigidité et la visibilité de la tête durant la course par rapport aux autres parties du corps. Le problème qui reste difficile à résoudre est celui de l'occultation partielle ou totale de la cible (éventuellement la tête). Intuitivement, nous avons pensé à prédire la position de la tête du nageur en tenant compte de sa vitesse moyenne et de la dernière position validée. Cependant, la prédiction est insuffisante dans le cas où la tête est occultée pendant un nombre d'images important entraînant, dans ce cas, une réduction de la précision du suivi qui est un point important pour les traitements ultérieurs.

Pour améliorer cela, le maillot de bain peut être considéré comme une partie du corps du nageur à suivre, en plus de la tête. Malgré le fait qu'il soit moins visible que la tête, sa visibilité est généralement alternée avec celle de cette dernière. Motivé par ce constat, nous proposons donc de suivre les deux cibles en parallèle et de prédire la position de celle qui est invisible à partir de l'autre dans le cas des occultations. Selon le critère de confiance, nous distinguons les quatre scénarios suivants, aussi illustrés dans la figure 4.9 :

1. La tête est visible alors que le maillot de bain est occulté \Rightarrow La détection de la tête est valide et la position du maillot de bain est à prédire selon la position de la tête détectée.

2. La tête est occultée alors que le maillot de bain est visible \Rightarrow La détection du maillot de bain est valide et la position de la tête est à prédire selon la position du maillot de bain détecté.
3. Les deux cibles sont visibles \Rightarrow La détection des deux cibles est valide avec une précision optimisée en tenant compte de la relation entre les deux cibles.
4. Les deux cibles sont occultées \Rightarrow Une optimisation de la technique de prédiction présentée dans la section 4.2 est proposée. Celle-ci se base sur un critère de décision complexe qui sera présenté par la suite.

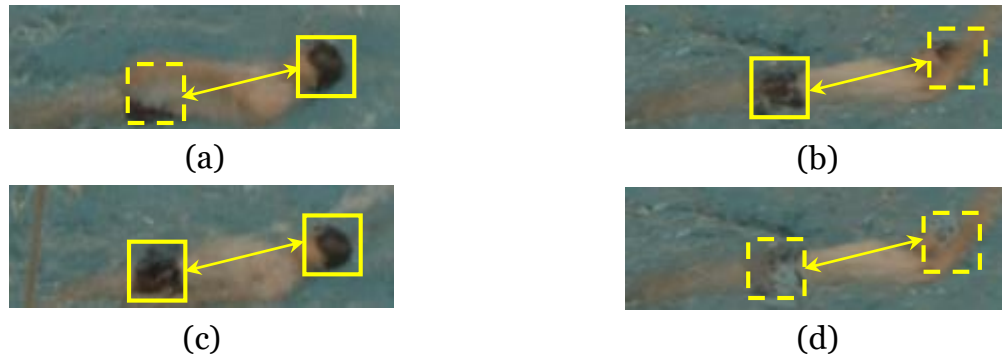


FIGURE 4.9 – Les quatre scénarios possibles pour le suivi basé sur l'approche multizone. (a) Détection de la tête valide, position du maillot de bain prédite. (b) Détection du maillot de bain valide, position de la tête prédite. (c) Détection des deux cibles valides. (d) Deux cibles prédites selon le critère de décision optimisé.

4.5.1 Inspiré par l'idée du multipiste

Dans le système de suivi multipiste que nous avons proposé, chaque piste représente un processus de suivi indépendant. Le but étant de vérifier plusieurs détections potentielles issues de différentes techniques (corrélacion optique NL-JTC, histogramme de couleur, LBP et HOG). Le choix de ces techniques est venu de la nécessité d'une description globale et riche de la référence à suivre. Partant de cette idée et dans le but d'optimiser le suivi, dans le cas où la partie du corps du nageur à suivre est occultée, nous proposons un nouveau système multipiste où chaque piste représente un processus de suivi d'une cible différente (tête, maillot de bain...). Ainsi, chaque piste pourra suivre une zone précise du corps du nageur tout en s'appuyant sur les détections des autres zones. Cette approche est appelée : multizone (*multi-related-target*).

4.5.2 Processus du suivi multizone

Deux informations sont nécessaires pour l'initialisation du processus de suivi multizone, la sélection des deux parties du corps à suivre comme références dans la première image ainsi que la relation géométrique entre les deux cibles exprimée par la distance euclidienne entre les objets. Dans la suite, nous choisissons la tête et le maillot de bain comme cibles, étant donnée la distance relativement stable qui les séparent et l'alternance de visibilité durant la course. Après avoir calculé la distance référentielle (d_r) entre la tête et le maillot de bain, nous appliquons la technique de détection par fusion dynamique séparément sur les deux références. Afin de ne négliger aucune détection, nous proposons de ne pas garder seulement la cible correspondante au plus grand pic dans le plan de fusion, mais de garder tous les pics supérieurs à un certain seuil (comme présenté dans notre approche multipiste uni-descripteur de la section 4.3).

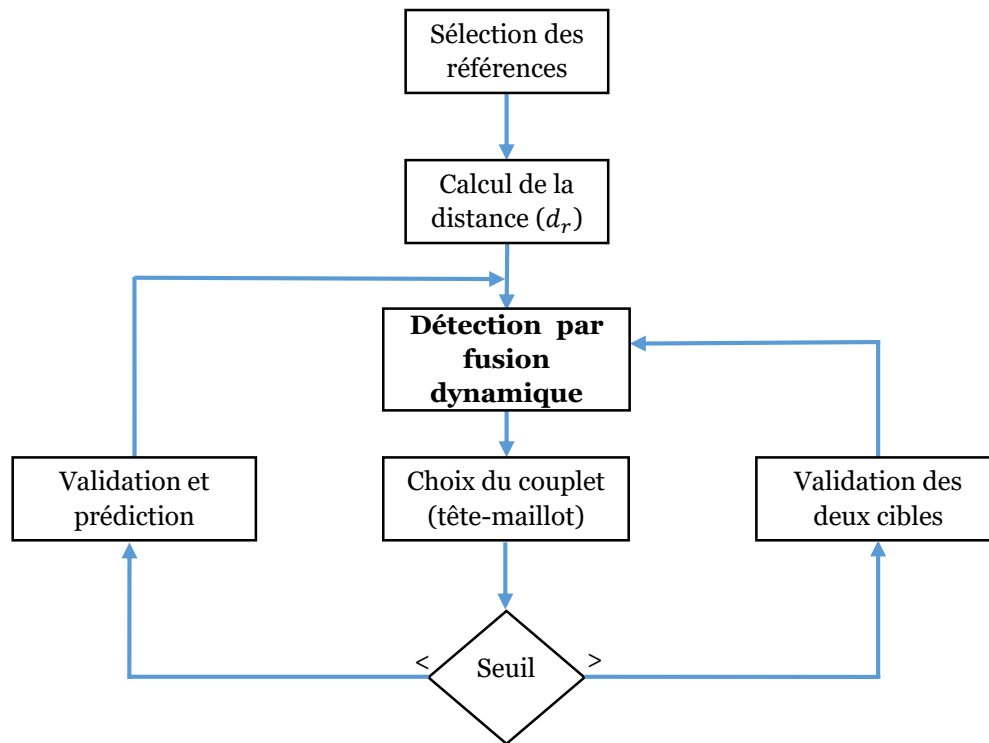


FIGURE 4.10 – Schéma synoptique de l'approche multizone. Premièrement, nous sélectionnons les références et nous calculons la distance entre elles. Ensuite, nous appliquons la technique de la fusion dynamique et nous choisissons le couplet (tête - maillot de bain) ayant la plus grande valeur de ressemblance. Si cette valeur est supérieure à un seuil empirique alors nous validons les deux cibles, sinon, nous validons la cible ayant la valeur la plus élevée et nous prédisons l'autre.

Ainsi, nous pourrions choisir a posteriori un couplet (tête - maillot de bain) parmi

les détections potentielles des deux cibles selon un critère de décision multi-paramètres. Ainsi, si ce critère est supérieur à un certain seuil empirique, la confiance envers le couplet sera élevée. Après la sélection du meilleur couplet, l'algorithme sera ré-exécuté avec les deux nouvelles cibles. En revanche, si le critère est inférieur au seuil, alors la confiance envers l'ensemble des deux détections sera faible. Dans ce cas, nous validerons la cible qui possèdera le facteur de confiance maximal et nous rejetterons l'autre cible pour la remplacer par une prédiction basée sur sa relation géométrique avec la cible validée.

4.5.3 Détections potentielles

Afin de retrouver toutes les détections potentielles, nous procédons à la recherche des maxima locaux dans le plan issu de la fusion dynamique $Plan_{fusion}$ qui est présenté dans la figure 4.11.a. Nous commençons par le débruitage du plan qui est significativement bruité dans le cas de notre étude en milieu aquatique. Pour cela, nous utilisons un filtre médian 2D de taille 3x3, le résultat est illustré dans la figure 4.11.b. Puis, nous appliquons un seuillage au $Plan_{fusion}$ afin d'obtenir le plan présenté dans la figure 4.11.c où la valeur du seuil est définie selon la formule :

$$Seuil = \max(\min(\max(Plan_{fusion}, dim_1)), \min(\max(Plan_{fusion}, dim_2))) \quad (4.9)$$

Avec dim_1 et dim_2 qui représentent respectivement les lignes et les colonnes de la matrice $Plan_{fusion}$. Ensuite, nous lisons encore une fois notre plan en utilisant un filtre gaussien (figure 5.d). Le but de cette étape est d'obtenir des pics dont l'allure est une gaussienne avec un seul pixel correspondant au maxima local.

Les maxima locaux retrouvés pour les deux cibles représentent les détections potentielles. Dans le but de détecter précisément le nageur, nous allons vérifier tous les couplets tête-maillot de bain possibles $N * M$, tel que N et M représentent respectivement le nombre de détections possibles pour la tête et le maillot de bain. Enfin, nous choisissons le meilleur couplet selon un critère de décision complexe que nous décrivons ci-après.

4.5.4 Critère de décision complexe

Afin d'évaluer chaque couplet tête-maillot de bain et choisir le meilleur, nous proposons un critère robuste constitué de différents sous-critères. En effet, ce principe est basé principalement sur la perception humaine pour le suivi des objets en mouvement. Par exemple, dans le cas des occultations partielles des nageurs par l'eau, nous avons tendance à suivre les parties visibles du corps du nageur et prédire le reste du corps

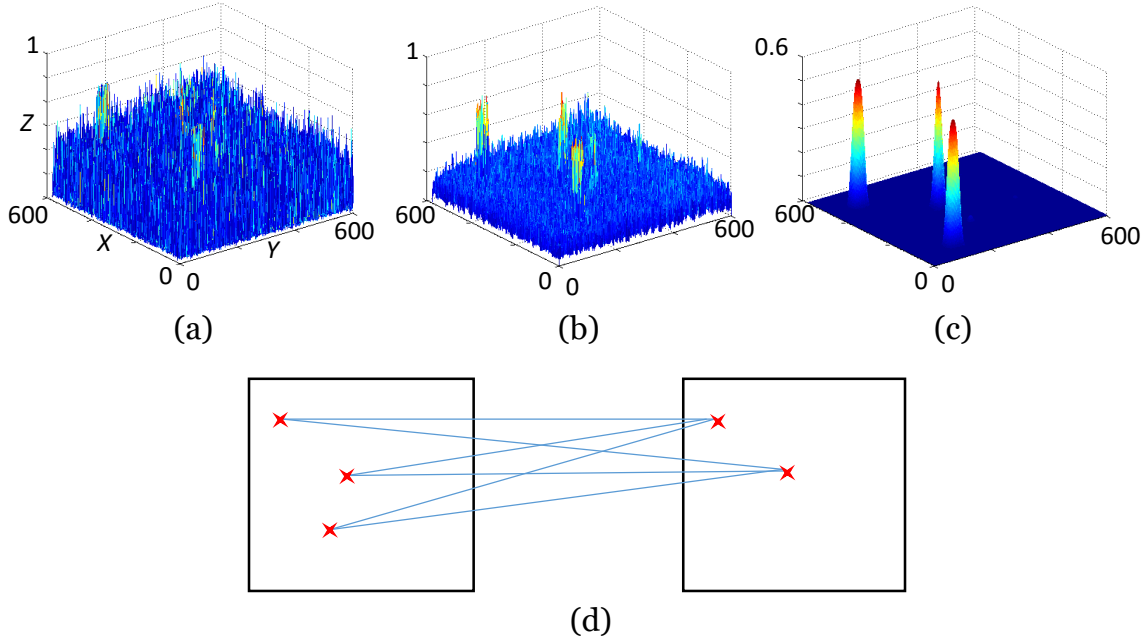


FIGURE 4.11 – Détections potentielles de la tête. (a) Plan de fusion dynamique. (b) Débruitage par le filtre médian. (c) Seuillage et filtrage gaussien. Le processus (a, b et c) est fait pour les deux cibles tête et maillot de bain. (d) Couplets de détections potentielles tête-maillot de bain.

occulté en se basant sur la relation géométrique entre eux mais aussi en utilisant la vitesse moyenne de l'ensemble. Motivé par ce constat, nous proposons un critère complexe de choix du meilleur couplet tête-maillot de bain. En particulier, ce critère prend en compte la valeur de confiance des deux cibles suivies, la distance inter-cibles et la vitesse de l'ensemble. Chacun de ces sous-critères est quantifié séparément pour chaque couplet, puis le choix est fait à partir d'une valeur calculée pour l'ensemble.

4.5.4.1 Facteurs de confiance

Le facteur de confiance est la valeur issue de l'application de la technique de fusion dynamique utilisée pour détecter les parties du nageur à suivre. Il correspond à la hauteur du pic d'une détection potentielle. Comme le montre la figure 4.11, un facteur de confiance est calculé pour chaque détection potentielle de chacune des deux zones. Pour cela, nous mettons en place deux vecteurs FC_t et FC_m qui représentent respectivement les facteurs de confiance des détections potentielles correspondant à la tête et au maillot de bain.

4.5.4.2 Distance inter-cibles

Cette valeur représente la distance euclidienne entre une détection potentielle de la tête et une autre du maillot de bain. Afin de calculer ces distances, nous avons besoin de connaître les coordonnées des détections potentielles que nous stockons dans les vecteurs X_t pour la tête et X_m pour le maillot de bain. Chaque distance sera comparée avec une distance de référence que nous calculons dans la phase d'initialisation du suivi. Au cours de cette comparaison, nous appliquons une fonction de filtrage gaussien centrée sur la valeur de la distance référentielle. Ceci dans le but de donner une valeur de 1 pour les distances égales à la distance référentielle et une valeur inférieure lorsque l'on s'éloigne de cette distance de référence. En particulier, cette valeur diminue en s'éloignant de la distance référentielle jusqu'à ce que nous arrivions à une valeur de 0 lorsque la différence entre les distances devient trop importante. La matrice $Dist$ est calculée comme le montre l'équation suivante :

$$Dist(i, j) = f(D(i, j)) \quad (4.10)$$

où i et j représentent respectivement l'indice de la détection potentielle de la tête et du maillot de bain. f est une fonction gaussienne, avec μ_f qui représente la distance de référence entre la tête et le maillot de bain. Notons que σ_f est fixé empiriquement après une étude statistique de la variation de la distance référentielle tête – maillot de bain (d_r) durant une course de nage libre, comme le montre la figure 4.12.

$$\sigma_f = \frac{MAX(d_r) - MIN(d_r)}{2} \quad (4.11)$$

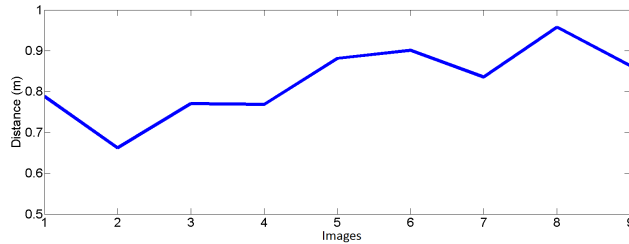


FIGURE 4.12 – Exemple de la variation de la distance tête maillot de bain durant une course nage libre 1500 - Championnat national de natation, Limoges 2015.

Finalement, la matrice D des distances entre les détections potentielles de la tête et du maillot de bain est calculée grâce à la formule suivante :

$$D = \begin{bmatrix} \|X_t(0) - X_m(0)\|_2 & \|X_t(0) - X_m(1)\|_2 & \dots & \|X_t(0) - X_m(M)\|_2 \\ \|X_t(1) - X_m(0)\|_2 & \|X_t(1) - X_m(1)\|_2 & \dots & \|X_t(1) - X_m(M)\|_2 \\ \vdots & \vdots & \ddots & \vdots \\ \|X_t(N) - X_m(0)\|_2 & \|X_t(N) - X_m(1)\|_2 & \dots & \|X_t(N) - X_m(M)\|_2 \end{bmatrix} \quad (4.12)$$

4.5.4.3 Ratio de la vitesse

Le calcul de la vitesse ne peut être introduit qu'après quelques instants (séquence de 10 images). Ceci est essentiel pour avoir une estimation fiable de l'avancement du nageur durant les premiers instants et de l'exploiter durant le reste de la course. Pour chaque couplet de détection potentielle tête – maillot de bain, nous calculons la vitesse instantanée de l'ensemble qui sera représentée par la vitesse de la tête. Cette vitesse instantanée sera comparée avec la vitesse moyenne dans le but de récupérer une mesure de $Vitesse(i) \in [0, 1]$. Pour cela nous appliquons une fonction gaussienne écrêtée comme décrit par la formule suivante :

$$Vitesse(i) = g(V(i)) \quad (4.13)$$

où $V(i)$ est la vitesse instantanée du nageur correspondant à la détection potentielle de la tête ayant l'indice i . Le vecteur V est calculé comme suit :

$$V = \begin{bmatrix} \frac{\|X_t(0) - X_{t-1}(0)\|_2}{\delta t} \\ \frac{\|X_t(1) - X_{t-1}(1)\|_2}{\delta t} \\ \vdots \\ \frac{\|X_t(N) - X_{t-1}(N)\|_2}{\delta t} \end{bmatrix} \quad (4.14)$$

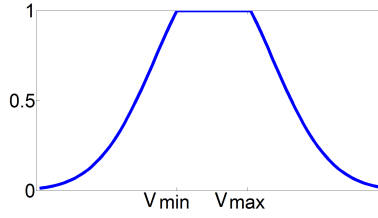
$g(\mu_g, \mu_g)$ est une fonction gaussienne écrêtée (figure 4.13), avec μ_g qui représente la vitesse moyenne du nageur V_{moy} calculée entre l'instant actuel t et l'instant 0 (début du suivi). σ_g est estimé par rapport à la variation de la vitesse instantanée du nageur, comme le montre la formule suivante :

$$\sigma_g = \frac{V_{max} - V_{min}}{2} \quad (4.15)$$

avec V_{max} et V_{min} qui représentent respectivement la vitesse maximale et la vitesse minimale qui sont estimées à $\pm 0.3m/s$ autour de la vitesse moyenne V_{moy} . La fonction gaussienne g est écrêtée autour des valeurs V_{max} et V_{min} pour favoriser tous les couplets qui ont une vitesse instantanée incluse dans cette marge.

4.5.4.4 Critère de décision finale

Le critère de décision final pour le choix du meilleur couplet tête – maillot de bain est calculé en faisant une sommation pondérée des quatre critères détaillés précédemment.


 FIGURE 4.13 – Fonction gaussienne écrêtée selon les valeurs de V_{max} et de V_{min} .

Deux critères sont favorisés, le facteur de confiance de la tête FC_t et le critère de distance inter-cibles $Dist$. Le premier car la tête du nageur a un taux de visibilité élevé par rapport au maillot de bain. Le second de par sa pertinence venant de la rigidité du corps humain qui contraint la variation de cette distance à de petites valeurs. Le critère final est calculé grâce à l'équation suivante :

$$Critere(i, j) = \alpha [FC_t(i) + Dist(i, j)] + (1 - \alpha) [FC_m(j) + Vitesse(i)] \quad (4.16)$$

avec $\alpha = \frac{2}{3}$ un coefficient de pondération qui permet de favoriser les critères les plus pertinents par rapport aux autres. On peut noter que ce critère final donne des valeurs dans l'intervalle $[0, 1]$ car tous les critères qui le composent sont normalisés dans ce même intervalle. Ceci permet de donner une estimation en pourcentage fiable et pertinente pour chaque couplet tête - maillot de bain. Finalement, ces estimations permettent de les comparer dans le but de choisir le meilleur couplet (I, J) qui peut être récupéré comme suit :

$$(I, J) = \underset{i, j}{\operatorname{argmax}} Critere(i, j) \quad (4.17)$$

La valeur de $Critere(I, J)$ est comparée avec un certain seuil (≥ 0.7) dans le but de valider la détection. Dans le cas positif, les positions des deux cibles détectées sont validées et le processus de suivi est relancé. Dans le cas contraire, nous validons la cible ayant le meilleur facteur de confiance et nous prédisons l'autre cible selon la distance de référence qui les lie. Le dernier est rencontré lorsque la valeur du critère $Critere(I, J)$ est très basse (< 0.5). Dans ce cas, nous considérons que les deux cibles sont occultées, et dans ce cas, leurs positions seront prédites globalement selon la technique de prédiction proposée dans la section 4.2. Le tableau 4.2 résume le processus de validation des cibles que nous adoptons.

Comparaison	Actions
$Critere(I, J) \geq 0.7$	Les deux cibles sont validées
$0.5 \leq Critere(I, J) < 0.7$	Si $FC_t \geq FC_m$, alors la position de la tête est validée et la position du maillot de bain est prédite.
	Sinon, la position du maillot de bain est validée et la position de la tête est prédite.
$Critere(I, J) < 0.5$	Les deux cibles sont prédites

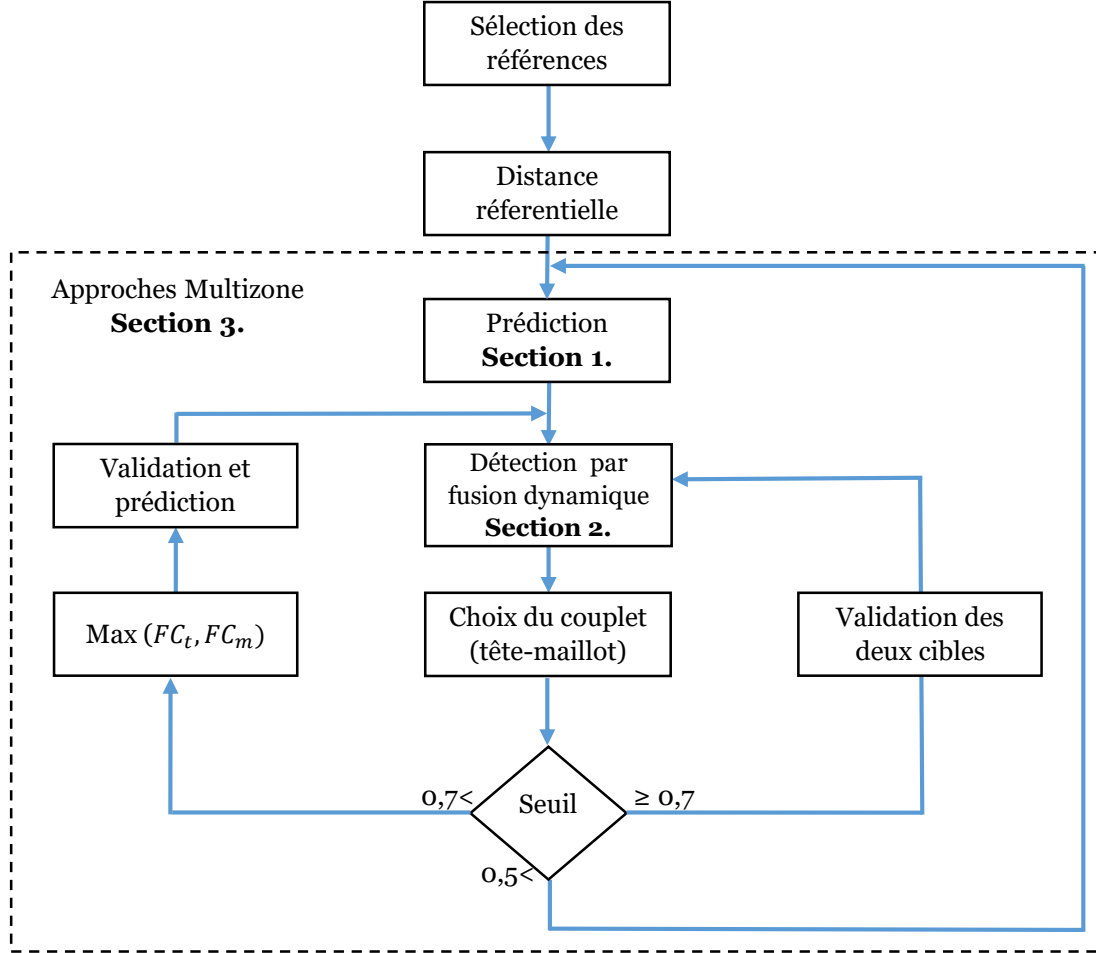
TABLE 4.2 – *Processus de validation des deux cibles (tête et maillot de bain) selon la valeur du Critere(I, J).*

FIGURE 4.14 – Schéma synoptique global du système optimisé de suivi des nageurs. Premièrement, nous sélectionnons les références de la tête et du maillot de bain. Puis nous prédisons leurs positions dans la prochaine image. Ensuite, nous appliquons l'approche de détection par fusion dynamique et nous récupérons les détections potentielles pour chacune des deux cibles. Enfin, nous appliquons l'approche multizone dans le but de choisir le meilleurs couplet tête - maillot de bain et nous relançons le suivi.

4.5.5 Discussion

Cette approche multizone permet de résoudre le cas de l'occultation de la tête du nageur à suivre. Pour cela nous avons proposé de suivre le maillot de bain en parallèle

de la tête. La relation géométrique entre ces deux zones, en plus des informations de ressemblance et de vitesse, permet de retrouver la zone occultée par rapport à celle qui est visible. De plus, cette approche permet d'ajuster la localisation et d'optimiser le suivi dans le cas d'occultation partielle d'une ou des deux zones. En effet, la visibilité de la tête et du maillot de bain n'est pas toujours alternée, nous pouvons trouver, dans de rares cas, une invisibilité quasi-totale des deux zones. Dans ce cas de figure, nous pouvons prédire la position en utilisant l'approche de prédiction, présentée dans la section 4.2. La figure 4.14 présente un schéma global de notre système de suivi de nageurs avec les différentes optimisations proposés.

4.6 Conclusion

Nous avons détaillé dans ce chapitre nos différentes contributions qui visent à surmonter les difficultés du suivi des nageurs. Tout d'abord, une prédiction de la prochaine position du nageur est proposée afin de maintenir la tête de nageur à suivre dans la région d'intérêt. Celle-ci s'appuie sur des informations liées principalement aux caractéristiques du milieu pour localiser globalement le nageur. Afin de détecter précisément le nageur, nous avons commencé par proposer une approche multipiste qui consiste à appliquer différentes techniques indépendamment durant quelques itérations puis à valider la meilleure piste et à relancer l'approche multipiste. Le choix de la piste est considéré comme une fusion de type "maximum" aux niveaux des décisions ce qui ne permet pas de profiter pleinement des pistes rejetées à chaque fois.

Afin de pallier ce problème et pour améliorer la fusion des données, nous avons proposé l'approche de fusion dynamique. Un avantage important de cette approche est qu'elle fusionne le plan de scores couleur avec le plan de corrélation. Ainsi, les informations de contour et de couleur sont combinées afin de générer un descripteur riche des images référence et cible. Cette approche est très efficace dans le cas de la visibilité de la tête du nageur à suivre. En revanche, dans les cas des occultations, les deux informations couleur et contour sont perdues. Dans ce cas, nous faisons appel à la prédiction qui localise approximativement la tête du nageur. Cependant, notre objectif étant de localiser précisément le nageur, nous avons proposé l'approche multizone qui consiste à suivre deux zones parallèlement : la tête du nageur et son maillot de bain. Cela nous permet de retrouver la zone occultée par rapport à celle qui est visible. Un autre avantage de cette approche est qu'elle améliore la précision du suivi dans le cas des occultations partielles en profitant du critère complexe que nous avons proposé.

Chapitre 5

Expérimentations et résultats

Sommaire

5.1	Introduction	119
5.2	Protocole expérimental	120
5.2.1	Choix du système de prise de vue	120
5.2.2	Système de prise de vue <i>8K</i>	125
5.2.3	Base de données	126
5.3	Résultats de suivi	127
5.3.1	Validation de la vérité terrain	128
5.3.2	Evaluation des prétraitements	133
5.3.3	Evaluation de l'approche de suivi multipiste	136
5.3.4	Evaluation de l'approche de suivi à base de la fusion dynamique	140
5.3.5	Evaluation de l'approche de suivi multizone	144
5.3.6	Discussion	146
5.4	Etude de performances de nageurs	147
5.4.1	Introduction à l'étude de performances de nageurs	147
5.4.2	Vitesse cyclique et intra-cyclique	147
5.4.3	Vitesse instantanée	150
5.4.4	Filtrage des courbes de vitesse instantanée	151
5.4.5	Exemples d'application réelle	153
5.5	Conclusion	156

5.1 Introduction

L'objectif principal de cette thèse consiste à développer un système fiable de suivi automatique de nageurs dans le but d'évaluer leurs performances. Pour ce faire, nous avons

proposé, dans les chapitres précédents un système de suivi optimisé avec différentes contributions selon les difficultés rencontrées. Ce chapitre est composé des trois parties suivantes :

Tout d’abord, nous commençons par la présentation du protocole expérimental. Dans cette partie, nous présentons les différents systèmes de prise de vue testés tout au long de la thèse. Ensuite, nous validons notre choix et nous détaillons la base de données de vidéos utilisées pour nos différents tests.

Dans la deuxième partie, nous évaluons notre système de suivi et nous présentons les différents résultats. Pour cela, nous validons notre vérité terrain (annotation manuelle) par rapport aux résultats d’une centrale inertielle spécifique au suivi des nageurs. Puis, nous évaluons chacune des approches proposées et nous finissons cette partie par une discussion concernant notre choix final pour un système de suivi de nageurs optimal.

Dans la dernière partie, nous présentons une étude préliminaire des performances des nageurs basée sur l’annotation manuelle et sur le suivi automatique. En effet, nous nous intéressons aux mesures de vitesse cyclique, intra-cyclique et instantanée. Celles-ci permettent aux entraîneurs d’analyser la stratégie adoptée par les athlètes durant les courses en étudiant l’effet de chaque mouvement sur sa vitesse et son rythme global. Enfin, nous présentons des exemples de résultats finaux fournis aux entraîneurs afin de montrer les objectifs atteints et les perspectives de nos travaux de thèse.

5.2 Protocole expérimental

5.2.1 Choix du système de prise de vue

Dans le contexte de la collaboration avec la Fédération Française de Natation, nous sommes amenés à développer un système de suivi automatique des nageurs pour les entraînements et les compétitions. Celui-ci doit permettre de les évaluer afin d’optimiser leur performance. Dans ce cadre, nous nous sommes orientés vers un système à base de caméras. Ce choix est motivé par les contraintes liées aux spécificités du projet :

- Sans capteurs ou équipements de géolocalisation portés par le nageur car ceci peut le gêner et limiter ses performances, de plus ils sont interdits en compétition par la Fédération Internationale de Natation FINA.
- Sans marqueur physique porté par le nageur à cause des restrictions imposées par la *FINA* concernant les tenues des athlètes.
- Avec intervention minimale de l’utilisateur afin de simplifier et d’accélérer la tâche d’estimation des performances.

Afin de répondre à ces besoins, nous proposons un système de prise de vue à base de caméras de haute définition. En effet, l’enregistrement des vidéos d’entraînements et

de compétitions est très important. Il doit permettre de revisionner les courses et les analyser qualitativement, à savoir, la qualité des mouvements effectués par le nageur. D'autre part, le système proposé doit permettre de suivre automatiquement le nageur à partir des séquences vidéo et fournir une analyse quantitative en termes de distance parcourue et de vitesse. Pour ce faire, nous avons évalué différents systèmes de prise de vue que nous détaillons dans la suite de cette section.

5.2.1.1 Caméra *IDS* en mouvement

Le premier système testé est composé d'une caméra industrielle *IDS UI-3370CP-C-HQ* présentée dans la figure 5.2. Cette caméra est équipée d'une interface USB 3.0 pour l'acquisition en temps réel des vidéos et le stockage sur un ordinateur ou sur des dispositifs externe. Ce système a été récupéré auprès de la Fédération Française de Natation afin d'être testé au championnat du monde de natation à Barcelone en Juillet 2013. L'objectif était de filmer les compétitions en zoomant sur les nageurs afin d'avoir une bonne résolution nécessaire à l'obtention de bons résultats pour le suivi automatique. Dans ce système de prise de vue il est nécessaire d'appliquer une rotation au pied de la caméra afin de cadrer au mieux le nageur lors de l'avancement de la course.

Ce système a plusieurs avantages, notamment, l'acquisition en temps réel qui permet la visualisation et le traitement de vidéos directement sur un ordinateur puissant. En effet, celui-ci permet d'obtenir des vidéos de haute résolution (2048x2048) avec un taux de 80 images par seconde. Cependant, cette caméra n'est adaptée qu'aux objectifs à focales fixe. Ceci est un inconvénient majeur de ce système, car lors des compétitions de haut niveau nous ne contrôlons pas totalement l'emplacement de nos caméras dans les gradins des piscines. Par conséquent, nous sommes emmenés à calculer la taille de la focale selon notre emplacement probable dans les gradins et selon le champs de vision que nous voulons filmer, comme le montre les équations suivantes :

$$\alpha = 2\arctan\left(\frac{c}{2f}\right) \frac{180}{\pi} \quad (5.1)$$

$$d = \frac{\frac{c}{2}}{\tan\left(\frac{\alpha}{2}\right)} \quad (5.2)$$

avec, c le diamètre du capteur, f la longueur de la focale et d la distance entre la caméra et la scène filmée. La figure 5.1 présente le schéma explicatif du fonctionnement de la caméra et des différents paramètres utilisés pour estimer la taille de la focale qui correspond à un emplacement donné dans les gradins.

Cependant, l'inconvénient majeur de ce système de prise de vue reste le mouvement continu de la caméra qui oblige à calibrer chaque image de la séquence vidéo. D'autant

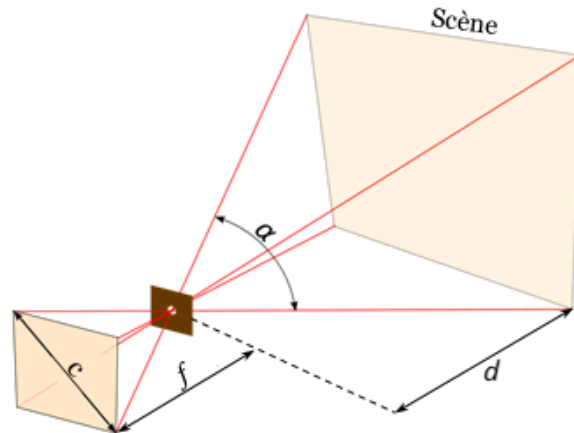


FIGURE 5.1 – Schéma explicatif du fonctionnement de la caméra et des différents paramètres permettant l'estimation de la longueur de la focale (f) par rapport à la distance (d) entre la caméra et la scène.

plus que cette tâche est rendu très complexe par le manque de marqueurs visibles dans les images tels que les changements de couleurs des lignes de couloirs utilisés comme repère pour le calibrage. Finalement, comme le montre la figure 5.2, la mise en place de ce système de prise de vue est difficile à cause des différents équipements utilisés : caméra IDS, objectifs, trépied, ordinateur, écran, multi-prises et câblage. Après les tests effectués en utilisant ce système, nous avons conclu que ce dernier n'était pas adapté aux besoins de notre projet et nous avons exploré d'autres pistes.

5.2.1.2 Caméra *Nikon D800* en statique

Les limitations du système de prise de vue utilisant la caméra *IDS* se résume à la difficulté de la mise en place, l'invariabilité de la focale de l'objectif et la rotation de la caméra. Afin de pallier ces problèmes, nous avons utilisé l'appareil photo *Nikon D800* dont nous disposons dans le laboratoire VISION de l'ISEN Brest. Cet appareil photo a été testé principalement durant les championnats régionaux de Bretagne à Quimper en 2014. Comme le montre la figure 5.3, cet appareil photo est facile à mettre en place car il ne nécessite qu'un trépied, ainsi, il est autonome et sa batterie est rechargeable. De plus, il dispose d'un objectif à focale variable permettant d'ajuster le champ de vision en fonction du bassin. Dans le but d'éviter la rotation de la caméra, nous réglons la focale pour obtenir un champ de vision couvrant tout le bassin. La prise de vue peut alors se faire avec une caméra statique (sans rotation) simplifiant de ce fait la calibration.

En utilisant ce système, le calibrage d'une seule image permet de calibrer toutes les séquences vidéo enregistrées avec la même configuration (positionnement de la caméra, focale, résolution, etc). Par conséquent, des mesures précises peuvent être faites sur les



FIGURE 5.2 – Mise en place du système de prise de vue utilisant la caméra industrielle IDS dans les gradins de la piscine en condition de compétition. Championnat du monde de natation - Barcelone, Juillet 2013



FIGURE 5.3 – Mise en place du système de prise de vue utilisant l'appareil photo Nikon D800 dans les gradins de la piscine en condition de compétition. Championnat régional de natation de Bretagne - Quimper, France 2014

vidéos, notamment la distance parcourue et la prédiction de la trajectoire du nageur. Le *Nikon D800* enregistre des vidéos qui peuvent atteindre un taux de 60 images par seconde en *Full HD* 1920x1080. Cependant, les tests préliminaires de suivi automatique de nageurs que nous avons effectués sur ses vidéos ont montré une insuffisance de la résolution ainsi que des artefacts liés à la compression appliquée sur les images. Ses limitations ont une influence directe sur la qualité de l'image, plus particulièrement, sur la taille et la forme de la cible à suivre. Afin de remédier à ces problèmes, il nous faut un système de prise de vue plus performant en terme de résolution et permettant l'enregistrement de vidéos peu compressées.

5.2.1.3 Caméra *Blackmagic 4K* en statique

Après ces tests effectués sur les différents systèmes de prise de vue, nous avons pu cibler les caractéristiques nécessaires pour enregistrer des séquences vidéo pour le problème de suivi de nageurs en entraînements et en compétitions. Les caractéristiques ciblées sont les suivantes :

- Très haute définition permettant d’avoir suffisamment de pixels décrivant le nageur à suivre (supérieure à 1080p).
- Taux suffisant d’images par seconde afin d’éviter le flou de mouvement (Supérieur à 20 images par seconde).
- Objectif à focale variable afin de permettre l’ajustement du champ de vision.
- Compression sans perte visuelle afin d’éviter la déformation du nageur à suivre par les artéfacts de compression.
- Simplicité de mise en place.

En faisant des recherches sur les produits existants en matériel optique, nous avons pu trouver une caméra en adéquation avec nos besoins. Nous parlons ici de la caméra *Blackmagic 4K* présentée dans la figure 5.4.



FIGURE 5.4 – Mise en place du système de prise de vue utilisant la caméra *Blackmagic 4K* dans les gradins de la piscine en condition de compétition. Championnat national de natation - Limoges, France 2015.

La caméra *Blackmagic 4K* enregistre des vidéos de très haute résolution 4K (3840x2160) et peut atteindre une fréquence de 30 images par seconde. Les vidéos sont stockées dans un disque dur SSD amovible qui peut être connecté à un ordinateur via une interface USB 3.0 pour des traitements ultérieurs. Différents niveaux de compression sont disponibles, de l’image brute du capteur dite *RAW* jusqu’au plus grand niveau de compression dit *ProRes Proxy* (avec perte visuelle). Après plusieurs tests, nous avons choisi la compression *ProRes 422 HQ* offrant une perte visuelle minimale quasiment indétectable à l’œil. Comme le montre la figure 5.4, cette caméra est simple à mettre en place. De

plus, elle est équipée d'un écran tactile de 5 pouces permettant la visualisation du flux vidéo ainsi que l'ajustement de la configuration : ouverture, vitesse d'obturation, ISO, compression, etc.

5.2.2 Système de prise de vue 8K

Le système de prise de vue utilisant la caméra *Blackmagic 4K* a été testé en entraînements dans une piscine de 25 mètres. Durant ces tests, nous avons enregistré des séquences vidéo avec un champ de vision couvrant tout le bassin. En utilisant cette configuration, la cible principale à suivre, qui est la tête du nageur, peut atteindre une taille de 40x30 pixels sur le premier couloir. Cette taille diminue en allant vers les couloirs les plus éloignés de la caméra. Avec la mise en place présentée dans la figure 5.4, la taille de la tête dans le dernier couloir est d'environ 25x18 pixels. Ceci est suffisant pour avoir une bonne description de la cible à suivre en termes de forme et de couleur.

Cependant, la taille de la tête du nageur filmé en utilisant la même configuration sur des bassins de 50 mètres sera divisée par deux. Par conséquent, le suivi devient plus difficile à cause du manque d'informations sur l'objet à suivre, ie. la tête du nageur. Afin de pallier ce problème, nous proposons un système de prise de vue 8K utilisant deux caméras *Blackmagic 4K*. Pour cela, chaque caméra couvre une moitié du bassin et les deux vidéos seront jointes en se basant sur le principe d'homographie. Ce dernier a pour objectif principal de calculer la transformation linéaire entre deux plans (images) selon l'équation 5.3. Cette transformation consiste en une matrice d'homographie H qui contient les informations concernant la rotation et la translation de la caméra entre les deux plans P_1 et P_2 .

$$P_1 = HP_2 \quad (5.3)$$

Dans notre cas, nous avons deux caméras de même type (*Blackmagic 4K*) qui filment les deux moitiés du bassin considérées comme des scènes co-planaires. L'objectif est donc d'estimer la transformation linéaire entre les images prises par les deux caméras afin de les projeter dans un même plan. Le résultat est une image 8K qui couvre l'intégralité du bassin comme le montre la figure 5.5.

Pour ce faire, nous appliquons implicitement le principe d'homographie en utilisant le calibrage pour corriger les perspectives et projeter les deux vidéos enregistrées par les deux caméras sur un plan image. Une zone de chevauchement entre les deux champs de vision est nécessaire afin d'assurer la reconstruction de l'image complète du bassin ainsi que pour synchroniser le passage des nageurs d'une vidéo à l'autre. La figure 5.5 montre le résultat de la reconstruction d'une image du bassin complet à partir de deux caméras après sa projection sur le plan métrique réel.

Pour notre étude, nous suivons le nageur à partir de la première vidéo jusqu'à son arrivée dans la zone de chevauchement qui se trouve à la moitié du bassin (autour de

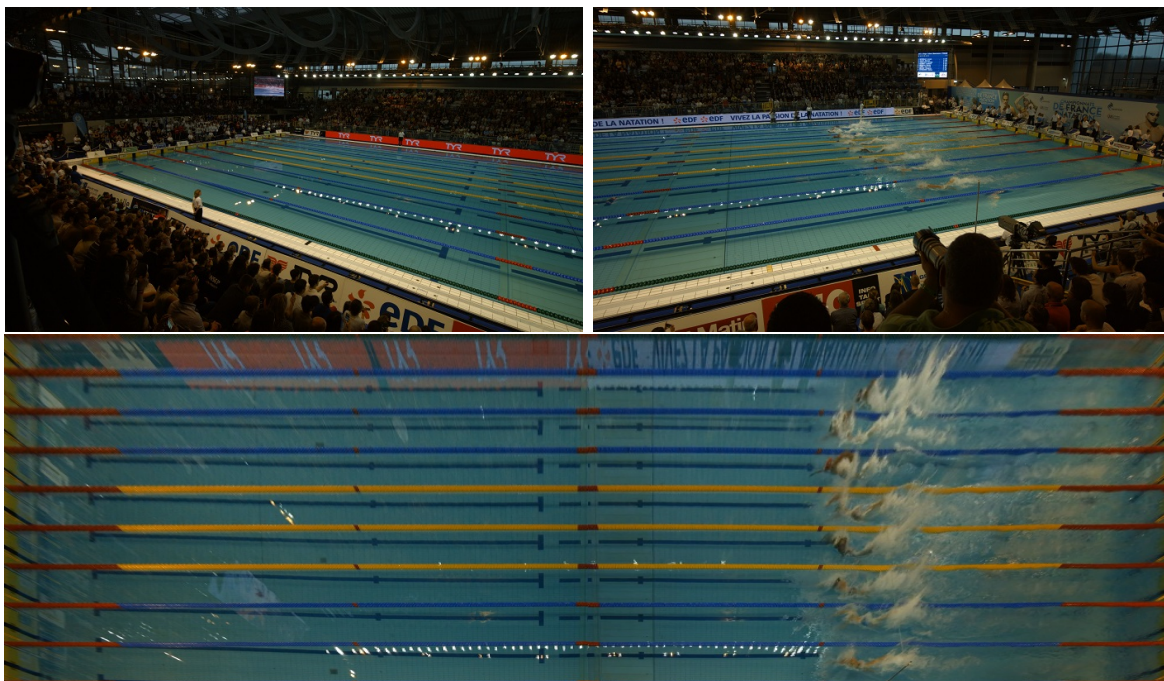


FIGURE 5.5 – *Reconstruction de l'image complète du bassin à partir de deux images 4K en se basant sur le principe d'homographie. Chacune des deux images couvre la moitié du bassin. En bas, l'image reconstruite après la correction de perspectives et l'association des deux images. Championnat national de natation - Limoges, France 2015.*

25 mètres). Ensuite, nous attendons son arrivée à la même position dans la deuxième vidéo afin de synchroniser et relancer le suivi dans la deuxième moitié du bassin. Ce système de prise de vue 8K a été testé sur les championnats nationaux de natation de Limoges 2015 et les championnats du monde de natation de Kazan 2015.

5.2.3 Base de données

Tout au long de l'évolution de notre choix du système de prise de vue, nous avons dû faire des tests de validation du système de suivi des nageurs. Pour ce faire, nous avons commencé par les championnats du monde de natation de Barcelone 2013, où nous avons utilisé le système de prise de vue utilisant la caméra *IDS* afin d'enregistrer des séquences vidéo d'entraînements et de compétitions. Ces séquences ont été enregistrées en utilisant une caméra en mouvement qui couvre un champ de vision de 6-12 mètres de hauteur et 6-12 mètres de longueur autour du nageur. En outre, les caractéristiques principales de ces séquences vidéo sont : une résolution de 2048x2048 pixels, une fréquence de 30 images par seconde et des durées qui varient en fonction des durées des courses.

Ensuite, nous avons utilisé l'appareil photo *Nikon D8000* pour enregistrer des vidéos

lors les championnats régionaux de Bretagne à Quimper en 2014. Ces séquences ont été enregistrées statiquement (sans rotation de la caméra) avec un champ de vision couvrant tout le bassin et avec les caractéristiques suivantes : une résolution *Full HD* 1920x1080 pixels, une fréquence de 25 images par seconde et des durées qui varient en fonction des courses. Finalement, d'autres compétitions ont été enregistrées en utilisant ces deux systèmes de prise de vue, notamment, les championnats nationaux de Rennes 2013, les championnats régionaux d'île de France de Paris 2013 et des entraînement à Brest.

Il est à noter que ces séquences vidéo ont été principalement utilisées pour les tests préliminaires du suivi automatique de nageurs et pour le choix du système de prise de vue. Ces séquences ont aussi été utilisées par l'équipe biomécanique de la Fédération Française de Natation pour d'autres analyses, notamment, le nombre de cycles, les temps de passage, etc.

Afin de tester le système de suivi automatique proposé dans les chapitres précédents, nous avons utilisé le système de prise de vue *8K* pour générer une base de données de vidéos de test. Cette dernière contient des vidéos d'entraînements et de compétitions nationales et internationales : des séances d'entraînement des nageurs du Club Nautique Brestois à la piscine Foch de Brest, les championnats de France de Limoges 2015 ainsi que les championnats du monde de Kazan 2015. Cette base de données contient une centaine de séquences vidéo avec tous les types de nages, les deux sexes et différents niveaux. Comme nous l'avons décrit dans la section 5.2.2, chaque course est enregistrée sur deux vidéos d'une résolution *4K* 3840x2160 pixels à une fréquence de 25 images par seconde et avec une compression *ProRes 422 HQ* offrant un débit de 110 MB/s.

5.3 Résultats de suivi

Notre objectif dans cette section consiste à évaluer notre système de suivi automatique de nageurs selon la logique que nous avons suivi au cours de son développement. Pour cela, nous allons commencer par évaluer l'approche de suivi permettant de déterminer la zone d'intérêt dans laquelle se trouve le nageur. Puis, nous présentons une étude comparative entre l'approche de suivi multipiste et les approches de suivi utilisant une seule méthode (NL-JTC, histogramme de couleur, LBP et HOG). Ensuite, nous évaluons notre approche de fusion dynamique en la comparant avec les deux techniques qui la constituent, NL-JTC et histogramme de couleur. Enfin, dans le but de montrer l'apport de l'approche multizone qui consiste à suivre la tête et le maillot de bain simultanément, nous la comparons avec le suivi de la tête toute seule utilisant la fusion dynamique.

5.3.1 Validation de la vérité terrain

5.3.1.1 CIREN

Récemment, l'entreprise ACTRIS Brest a développé un système d'étude de performances de nageurs basé sur la centrale inertielle CIREN (Centrale Inertielle pour la Recherche et l'Entraînement de Nageurs). Ce système est composé d'un capteur porté par le nageur permettant de mesurer et d'enregistrer ses mouvements. Ces données sont ensuite recueillies par un logiciel qui va les traiter et les synchroniser à la vidéo correspondante. Ceci permet aux entraîneurs d'étudier les performances des nageurs tout en visualisant la vidéo et les mesures en même temps. Ce système permet d'obtenir des mesures précises qui ont été validées par des entraîneurs qualifiés durant les entraînements de l'équipe de France. En revanche, ce système reste gênant pour le nageur durant les entraînements et est interdit durant les compétitions de haut niveau. Dans ce cadre, nous proposons dans la suite un système de mesure des performances basé sur l'utilisation de caméras qui ne nécessite d'aucun capteur porté.

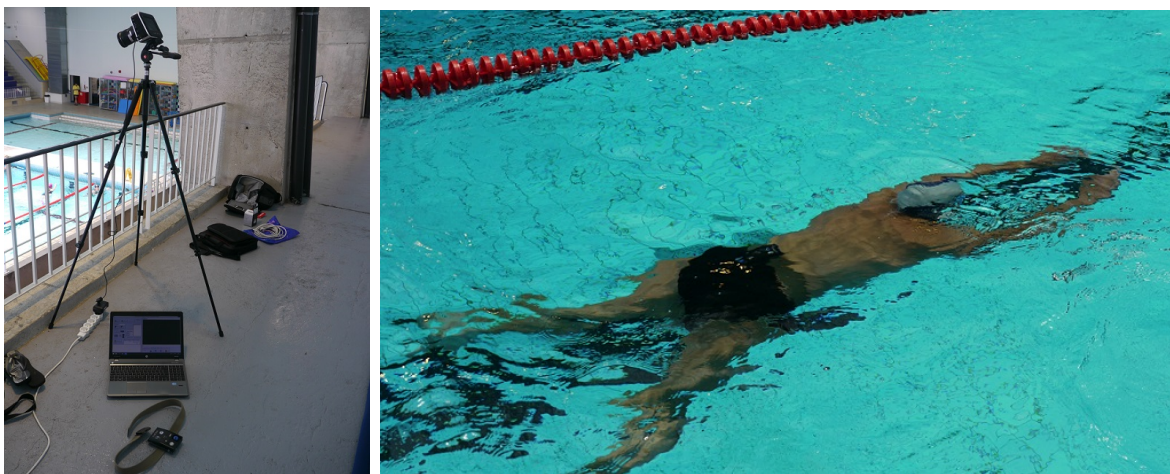


FIGURE 5.6 – Mise en place du système d'évaluation de nageurs utilisant la centrale inertielle CIREN. Tests effectués à la piscine Foch - Brest, France 2015.

5.3.1.2 Vérité terrain (Annotation manuelle)

L'évaluation de notre système de suivi automatique nécessite la connaissance des coordonnées de la position du nageur durant toute la séquence vidéo évaluée. Ces dernières nous serviront de vérité terrain pour valider celles obtenues par le système de suivi automatique. Pour ce faire, nous avons développé un prototype d'annotation manuelle qui consiste à calibrer la vidéo, zoomer sur le nageur et annoter la position de la cible

à suivre dans chaque image de la vidéo (voir l'annexe A pour plus de détails). Ces annotations sont enregistrées sous forme de vecteurs contenant dans chaque case les coordonnées de la cible et l'indice de l'image correspondante. Il est à noter que, pour chaque annotation, nous enregistrons les coordonnées pixels de la cible afin d'évaluer le système de suivi automatique. Nous enregistrons d'autre part ses coordonnées métriques afin de générer des courbes de vitesse qui seront utilisées comme vérité terrain et qui seront comparées avec celles issues du système de suivi automatique pour la validation de ce dernier.

Contrairement à la CIREN, cette annotation manuelle peut être utilisée pour le calcul de la vitesse instantanée des nageurs durant les entraînements comme durant les compétitions et cela est fait sans capteur porté par le nageur. Il nous reste donc à confirmer sa fiabilité et la précision des mesures fournies afin de la valider et de l'utiliser comme référence pour l'évaluation du système de suivi automatique.

5.3.1.3 Comparaison CIREN-annotation manuelle

Afin de valider le système d'évaluation de nageurs basé sur l'annotation manuelle, nous le comparons avec la centrale inertielle CIREN. Le processus d'annotation est relativement complexe, un calibrage est nécessaire pour chaque vidéo afin de calculer la correspondance pixel-mètre et de corriger les perspectives des images.

Nous avons constaté que l'annotation devient difficile dans le cas où le nageur à suivre est occulté (éclaboussures, bras, etc). Ceci oblige l'annotateur à interpoler et à estimer la position du nageur en se basant sur l'historique des annotations. A cause de l'ensemble des ces imprécisions, nous proposons de valider ce système d'annotation manuelle, grâce aux résultats de la centrale inertielle CIREN, afin d'utiliser ces annotations comme référence.

La figure 5.7 présente des courbes de vitesse instantanée d'un nageur amateur en nage libre. En analysant la courbe de vitesse calculée par la CIREN, en bas, nous remarquons un profil répétitif sous forme d'un avancement rapide suivi d'un avancement moins rapide (bras fort, bras faible). Ceci apparait bien sur la courbe présentant l'annotation manuelle, en haut dans la même figure, jusqu'à la 14^{ème} seconde. Ensuite, nous retrouvons les mêmes cycles mais avec une différence minime d'amplitude entre le bras fort et le bras faible. Il est à noter que les courbes de vitesse issues de l'annotation manuelles sont lissées par un filtrage gaussien dans le but de générer des mesures pertinentes proches des résultats de la CIREN. Cette partie sera détaillée dans la section 5.4.4.

Les droites horizontales figurant dans les deux courbes autour de la valeur 0.8 m/s correspondent à la vitesse moyenne calculée sur l'ensemble des valeurs de chaque courbe.

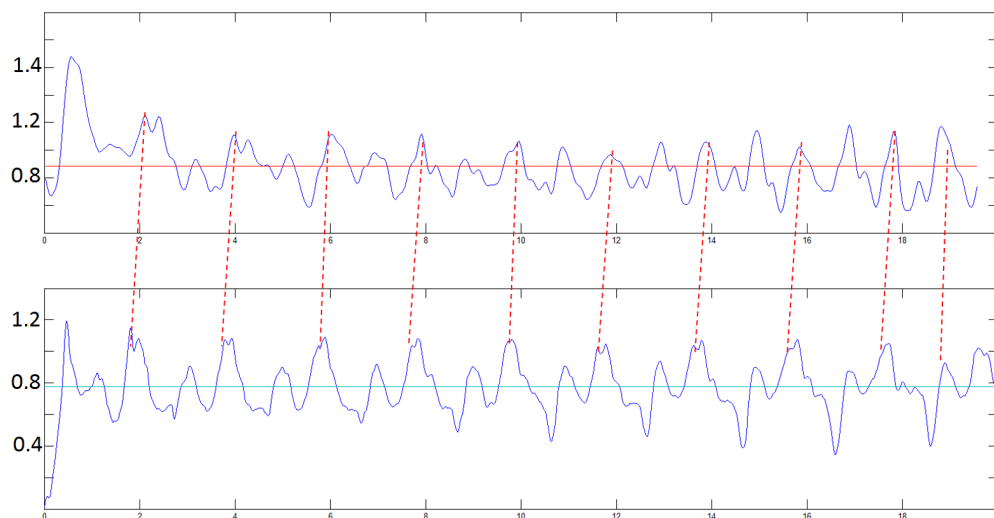


FIGURE 5.7 – *Comparaison des courbes de vitesse calculées à partir de l’annotation manuelle de la tête (haut) et de la CIREN (bas) - Entraînement nage libre - piscine Foch de Brest.*

Une différence légère est remarquée entre ces deux valeurs. Cela est la conséquence de la première partie des courbes qui correspond au plongeon qui ne fait pas partie de la phase de nage. Dans le cas où cette partie est éliminée, la vitesse moyenne tendra vers 0.74m/s pour les deux courbes. En se basant sur ses remarques, nous pouvons valider les résultats de l’annotation manuelle sur cette course.

D’autre part, ces résultats sont confirmés par les courbes de la figure 5.8 où nous constatons le même profil. Ces courbes représentent la vitesse instantanée d’un nageur amateur en nage brasse avec, de haut en bas, l’annotation manuelle de la tête, la CIREN et l’annotation manuelle du maillot de bain. Les mêmes cycles apparaissent sur les trois courbes avec des différences minimales. En ignorant la zone faisant référence au plongeon, la vitesse moyenne de chaque courbe tend vers 0.7m/s.

Cette étude comparative entre les courbes de vitesse instantanée issues de la centrale inertielle CIREN et de l’annotation manuelle permet de valider les résultats de cette dernière. Par conséquent, nous l’utiliserons, par la suite, comme référence pour valider les approches de suivi automatique. Avant cela, nous présentons quelques exemples de mesures d’erreurs afin de sensibiliser le lecteur aux intervalles d’erreur tolérés.

5.3.1.4 Mesures d’erreurs

Deux types d’erreurs sont possibles : les erreurs liées au calibrage et celles liées à la position du nageur. Le premier type concerne les positions des quatre points du

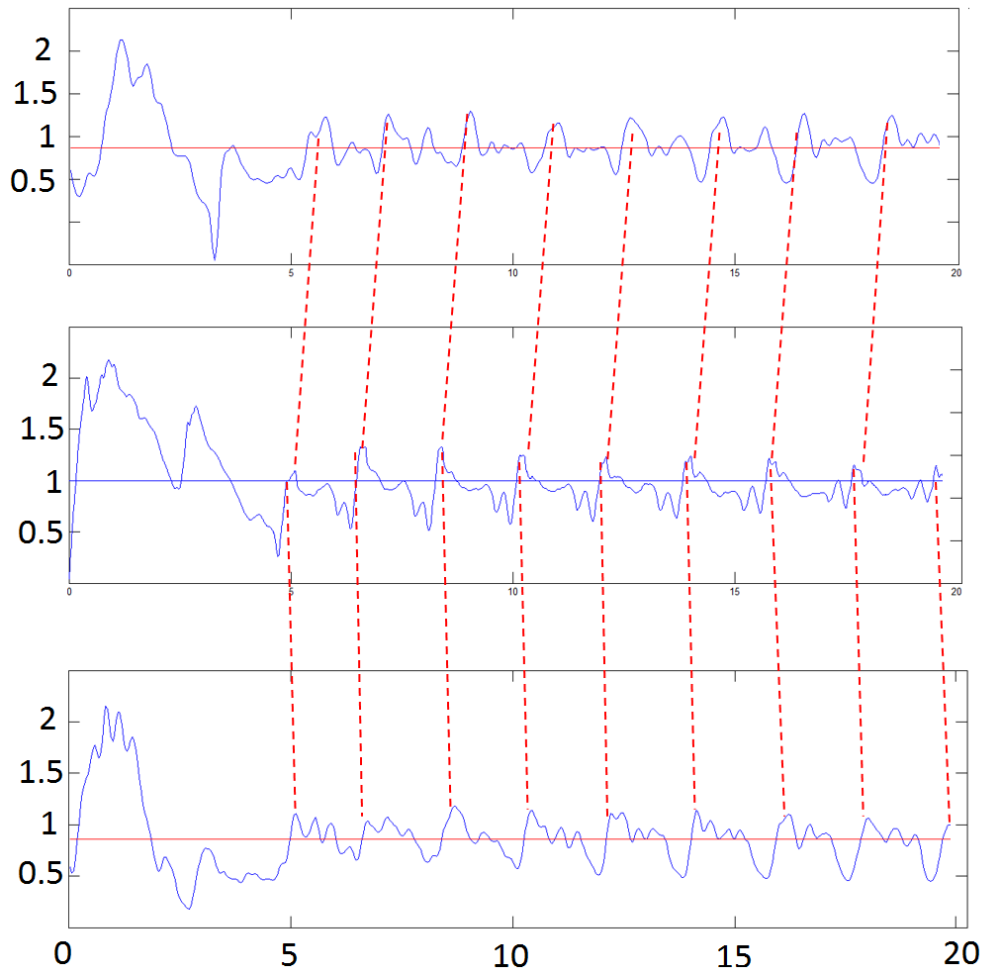


FIGURE 5.8 – Comparaison des courbes de vitesse calculées à partir de l’annotation manuelle de la tête (haut), de la CIREN (milieu) et de l’annotation manuelle du maillot de bain (bas) - Entraînement brasse - piscine Foch de Brest.

calibrage. Cette erreur est négligeable car nous utilisons un système de prise de vue statique. De ce fait, le calibrage de l’intégralité de la vidéo nécessite de calculer les paramètres de projection que pour une seule image. Ainsi, l’erreur de calibrage sera la même sur chaque image de la vidéo et n’impactera que très peu les mesures de vitesses. De plus, dans le cas d’une erreur de calibrage la même erreur sera appliquée à chaque image et n’aura que très peu d’influence sur l’estimation de la vitesse. En ce qui concerne le deuxième type d’erreur, celle-ci apparaît lors de l’estimation de la position du nageur dans le bassin représentée par la position annotée ou détectée de l’extrémité de sa tête.

Par exemple, dans la figure 5.9 :
Couloir 1

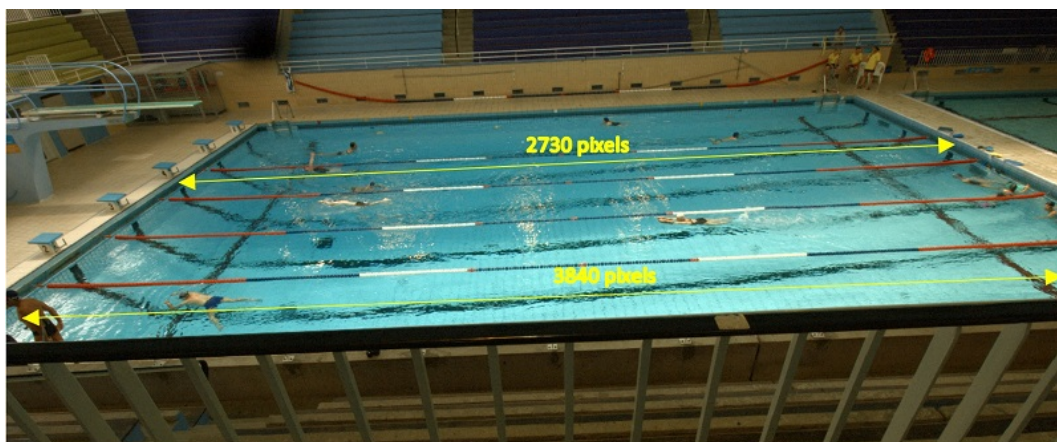


FIGURE 5.9 – *Estimation de la longueur des différents couloirs de nage en pixels - Bassin de 25 m*

- 1 pixel \rightarrow 5.7mm
- Si le nageur nage à une vitesse moyenne de 2m/s \rightarrow 80mm/image \rightarrow 14pixel/image
- L'erreur d'un pixel sur la position du nageur génère une erreur d'environ 7% sur sa vitesse

Couloir 4

- 1 pixel \rightarrow 9.2mm
- Si le nageur nage à une vitesse moyenne de 2m/s \rightarrow 80mm/image \rightarrow 8.7pixel/image
- L'erreur d'un pixel sur la position du nageur génère une erreur d'environ 11.5% sur sa vitesse.

5.3.1.5 Discussion

Ces mesures montrent clairement que le niveau de précision attendu du système de suivi de nageurs est élevé. En effet, une erreur de quelques pixels peut avoir des répercussions significatives sur les courbes de vitesse. Cependant, cette erreur n'est pas cumulative car si on se trompe de quelques pixels en annotant le bout de la tête du nageur, nous les récupérerons dans les images suivantes. Par conséquent, nous pouvons lisser ces erreurs et réduire leur effet sur le profil global de la vitesse instantanée du nageur. Pour cela, nous appliquerons un filtrage gaussien aux courbes que nous présentons par la suite dans la section 5.4.4. Malgré des erreurs possible, nous pouvons valider les résultats de la vérité terrain basée sur l'annotation manuelle afin de l'utiliser pour l'évaluation des approches de suivi automatique.

5.3.2 Evaluation des prétraitements

Afin de préparer la vidéo et faciliter le suivi, nous avons développé deux modules basés sur des prétraitements d'images : le module de localisation de la région d'intérêt et le module de détection automatique de la tête du nageur permettant d'initialiser le suivi. la suite de cette section présente l'évaluation de chacun de ces modules.

5.3.2.1 Evaluation de la localisation de la région d'intérêt

L'objectif de ce module est de détecter le mouvement du nageur dans le but de restreindre la région d'intérêt. Pour cela, nous avons proposé dans le chapitre 1 section 2.5.6.1, une méthode adaptée basée sur le principe *a contrario* [100].

L'idée de base consiste à générer un modèle de bruit standard correspondant au mouvement de l'eau. Ceci est effectué en calculant la différence entre des images représentant des moments où le couloir est vide. Une fois le modèle du bruit établi, nous calculons la différence entre chaque image de la séquence vidéo et l'image précédente. Ensuite, nous décomposons chaque image de référence en blocs. Chaque bloc est comparé avec le modèle de bruit pour déterminer sa nature : bruit ou mouvement du nageur. Enfin, la région d'intérêt contenant les blocs de mouvement validés est localisée.

Dans cette approche, nous avons constaté que notre méthode était incapable de détecter le mouvement du nageur dans le cas d'un mouvement minime. Pour pallier ce problème, nous avons proposé la technique de prédiction introduite dans le chapitre 4 section 4.2. Cette dernière se base sur un ensemble de critères pertinents : la direction de nage, la position référentielle validée et la vitesse moyenne.

Afin d'évaluer ce module de localisation de la région d'intérêt, nous l'avons testé sur 5 séquences vidéo de nage libre durant les championnats nationaux de Limoges 2015. Chaque séquence, contenant 400 images avec un taux de 25 images/s, commence par quelques images du couloir vide avant le plongeon du nageur. Ces quelques images nous permettent d'établir le modèle de bruit avec la méthode *a contrario*. La détection automatique du mouvement dans le couloir sera évaluée en utilisant le pourcentage de localisations réussies (nageur dans la région d'intérêt).

Séquences	Nageurs	Pourcentage de localisation	
		<i>a contrario</i> sans prédiction	<i>a contrario</i> avec prédiction
Nage libre 1	Nageur 1	95.25	100
	Nageur 2	97	100
	Nageur 3	95.75	99.5
Nage libre 2	Nageur 1	94.5	99.75
	Nageur 2	96	100

TABLE 5.1 – *Pourcentage de localisation de la région d'intérêt en utilisant la méthode *a contrario* avec et sans prédiction. Nage libre - Championnat national de natation, Limoges 2015.*

Le tableau 5.1 présente les résultats de la méthode *a contrario* appliquée à la localisation de la région d'intérêt selon les deux modes : avec et sans prédiction. Nous constatons que l'application de la méthode *a contrario* permet d'obtenir des pourcentages de localisation élevés qui varient entre 94.5% et 97%. Dans ce cadre, ces régions d'intérêt seront utilisées, par la suite, pour détecter précisément la tête du nageur. De ce fait, la localisation de la région d'intérêt est une étape cruciale. Dans le but d'optimiser les résultats et d'assurer l'existence de la tête dans la région d'intérêt, nous avons couplé la méthode *a contrario* avec une technique de prédiction adaptée (chapitre 4 section 4.2). Cette dernière nous a permis d'obtenir des pourcentages de localisation très élevés proches de 100%.

5.3.2.2 Evaluation de l'initialisation de la référence

Le suivi automatique des nageurs nécessite d'avoir une image de référence pertinente de la tête du nageur à suivre. Cependant, étant donné que la couleur du bonnet des nageurs peut changer, la seule information dont nous disposons est la forme de la tête. Partant de ce constat, nous avons proposé une nouvelle méthode que nous avons appelée *Scaled composite JTC* [1]. Cette dernière est expliquée en détails dans le chapitre 1 section 2.6.

La méthode *Scaled composite JTC* consiste à générer une base de données contenant des têtes de nageurs dans les différentes situations. Ensuite, le filtre composite [103] est appliqué sur n images choisies selon la situation (direction, âge, sexe, etc.). Afin de générer une image composite pertinente, nous avons fixé n à 3 car $n > 3$ génère une image composite bruitée et en choisissant $n < 3$ la description de l'image composite est insuffisante. Une fois l'image composite générée, elle sera re-dimensionnée selon le rapport taille de la tête - largeur du couloir. Enfin, la méthode *NL-JTC* est appliquée sur un plan d'entrée formé de l'image composite et d'une image cible extraite en utilisant la méthode *a contrario* évaluée dans la section précédente.

Dans le but de détecter la meilleure image de référence de la tête du nageur à suivre, nous appliquons la méthode *Scaled composite JTC* sur une petite séquence vidéo choisie au début de la course car la tête du nageur réapparaît rapidement après la sortie du plongeon. Parmi les cibles détectées, nous choisissons les trois meilleurs cibles afin de les utiliser comme références pour initialiser le suivi.

Afin d'évaluer ce module d'initialisation de la référence, nous l'avons testé sur 5 séquences vidéo (100 images) de nage libre durant les championnats nationaux - Limoges 2015. Ces séquences sont choisies au moment de la reprise de nage après le plongeon du début de course. Deux critères sont utilisés pour évaluer ce module : le pourcentage de détection sur les 100 images de chaque vidéo et le pourcentage d'initialisation réussie sur les 3 images choisies.

Séquences	Nageurs	Pourcentage de détection	Taux d'initialisation réussie (%)
Nage libre 1	Nageur 1	83.33	100
	Nageur 2	86.44	100
	Nageur 3	85.71	100
Nage libre 2	Nageur 1	78.31	100
	Nageur 2	82.14	100

TABLE 5.2 – *Evaluation de la méthode Scaled composite JTC pour l'initialisation de la référence. Nage libre - Championnat national de natation, Limoges 2015.*

Le tableau 5.2 montre des pourcentages de détection élevés entre 78.31% et 86.44% pour les 5 séquences testées bien que nous n'ayons utilisé que la méthode NL-JTC pour la détection. Ceci est expliqué par la visibilité de la tête du nageur sur cette période de reprise de nage où le nageur vient de sortir de l'eau après le premier plongeon. Les cas de fausses détections surviennent principalement à cause d'occultations partielles et totales de la tête du nageur à suivre. D'autre part, pour chacune des 5 séquences testées, nous choisissons les 3 meilleures détections selon leurs valeurs de PCE, comme le montre la figure 5.3. Dans cette figure, nous avons présenté les 10 meilleures détections pour la séquence Nage libre 1 du tableau 5.2. Parmi ces détections, nous choisissons les 3 meilleures ayant, respectivement, les valeurs de PCE suivantes : 0.92, 0.9 et 0.87. Ceci nous a permis d'obtenir 100% de réussite sur les images choisies pour l'initialisation du suivi.

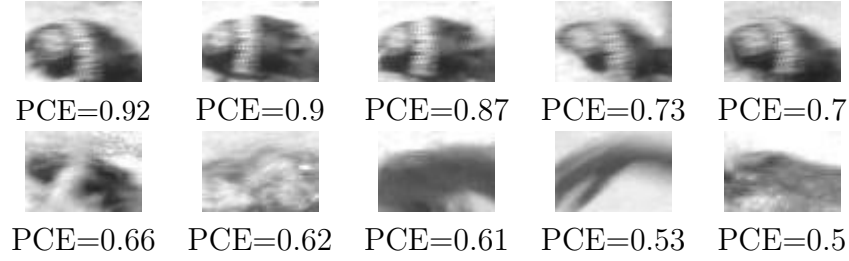


TABLE 5.3 – 10 meilleures détections pour la séquence *Nage libre 1 - Nageur 1. Championnat national de natation, Limoges 2015.*

Après la validation des deux modules de localisation de la région d'intérêt et d'initialisation de la référence, nous allons évaluer, par la suite, les approches que nous avons proposées pour le suivi automatique des nageurs. Pour cela, nous allons suivre l'ordre chronologique de la thèse à savoir, l'approche multipiste, puis l'approche basée sur la fusion dynamique et enfin l'approche multizone.

5.3.3 Evaluation de l'approche de suivi multipiste

Dans cette partie, nous évaluons les approches de suivi multipiste [3] présentées dans le chapitre 4, section 4.3 : l'approche uni-descripteur basée sur des pistes NL-JTC et l'approche multidésripteur basée sur des pistes NL-JTC, LBP et HOG. Pour cela, nous faisons une étude comparative avec les approches de base que nous avons adaptées pour le suivi de nageurs (chapitre 3) : NL-JTC, histogramme de couleurs, LBP et HOG. Notre étude porte sur la distance entre les positions des nageurs détectées en utilisant ces approches automatiques et la vérité terrain issue de l'annotation manuelle.

Nous évaluons ces approches par le pourcentage de suivi. Ce dernier est calculé en mesurant la distance entre les positions du nageur détectées, en utilisant les approches de suivi, et la vérité terrain. La décision est prise grâce à un seuil calculé relativement à la taille de la référence utilisée, comme le montre l'équation suivante :

$$seuil = \frac{\sqrt{Ref_x^2 + Ref_y^2}}{2} \quad (5.4)$$

avec Ref_x et Ref_y les dimensions de la référence.

Le seuil choisi est égal à la moitié de la diagonale afin d'autoriser, au minimum, un chevauchement d'un quart entre la cible détectée et celle qui est annotée. Finalement, nous utilisons les critères statistiques de moyenne et d'écart-type, afin d'évaluer sa stabilité tout au long de la séquence. Le tableau 5.4 présente les résultats des tests effectués sur deux séquences vidéo d'entraînement ayant les caractéristiques suivantes :

- Type de nage : nage-libre

- Vitesse moyenne : ~ 1 m/s (Vidéo 1) et ~ 1.5 m/s (Vidéo 2)
- Système de prise de vue : caméra IDS
- Résolution : $2048 * 2048$ pixels
- Nombre d'images : 150
- Fréquence : 30 images/seconde

Approches de suivi	Vidéos de test	Pourcentage (Pixels)	Moyenne (Pixels)	Ecart-type (%)
NL-JTC	Vidéo 1	88.00	25.7	17.02
	Vidéo 2	76.67	32.3	23.16
LBP	Vidéo 1	81.33	25.5	27.63
	Vidéo 2	83.00	24.62	25.5
HOG	Vidéo 1	84.67	28.63	17.76
	Vidéo 2	60.67	43.21	39.02
Histogramme de couleur	Vidéo 1	88.33	24.21	18.60
	Vidéo 2	86.67	26.06	19
Multipiste NL-JTC	Vidéo 1	94.00	21.70	12.98
	Vidéo 2	93.33	23.00	14.65
Multipiste multidescripteur	Vidéo 1	94.67	19.03	11.18
	Vidéo 2	93.67	20.37	11.18
Optimal	Vidéo 1	98.00	15.8	10.12
	Vidéo 2	97.33	16.34	10.83

TABLE 5.4 – Comparaison entre les résultats de l'approche de suivi multipiste et les différentes approches de suivi de base en termes de distance entre les positions détectées et la vérité terrain.

Le tableau 5.4 présente les résultats du suivi sous forme d'une étude comparative entre les approches multipistes (multipiste uni-descripteur et multidescripteur), les approches de base (NL-JTC, histogramme de couleur, LBP et HOG) et le suivi optimal que nous pouvons obtenir en prenant manuellement la meilleure détection entre celles issues des 4 approches classiques pour chaque image.

Les résultats montrent que les approches multipistes ont nettement amélioré les résultats des approches de base. Par exemple, pour la vidéo 1 où le nageur nage lentement à une vitesse moyenne de ~ 1 m/s, nous avons obtenu un pourcentage de 94% et 94.67% en appliquant, respectivement, les approches multipistes uni-descripteur et multidescripteur. Ceci est significativement meilleur que les approches de base qui varient entre 84.67% et 88.33%. En ce qui concerne la vidéo 2 où le nageur nage plus vite avec une vitesse moyenne de ~ 1.5 m/s, l'apport des approches multipistes est nettement visible. En effet, le pourcentage de suivi qui varie entre 60.67% et 86.67% pour les approches de base est amélioré jusqu'à 93.33% et 93.67% avec les approches multipistes.

D'autre part, nous constatons la même tendance pour les deux autres critères. Comme le montrent le tableau 5.4 et la figure 5.10, La moyenne des distances ainsi que l'écart-type diminuent significativement en appliquant les approches multipistes. La diminution de la moyenne des distances représente une optimisation de la précision du suivi tandis que la diminution de l'écart-type reflète la stabilité du suivi.

Ces optimisations apportées par les approches multipistes sont expliquées par leur principe de fusion des décisions. Ce dernier consiste à vérifier plusieurs pistes de suivi pour ensuite choisir la meilleure en terme de ressemblance de couleur. Concernant l'approche uni-descripteur, les pistes sont représentées par les pics de corrélation les plus élevés dans le plan de corrélation. Pour l'approche multidésripteur les pistes représentent : deux pics issus de la technique NL-JTC, la meilleure détection de la technique LBP et la meilleure détection de la technique HOG. Cette approche multidésripteur donne les meilleurs résultats comparés à toutes les autres approches de base ainsi qu'à l'approche uni-descripteur.

Enfin, nous concluons cette partie par une comparaison entre l'approche multipiste multidésripteur et le multipiste optimal. Ce dernier est conçu comme repère afin d'évaluer les approches multipiste. Il s'agit de choisir manuellement la meilleure des pistes dans chaque itération afin d'obtenir un suivi multipiste optimal. Comme le montrent le tableau 5.4 et la figure 5.10, l'approche multipiste multidésripteur reste moins performante que le multipiste optimal. Ceci est expliqué par la limitation de la fusion au niveau des décisions ainsi que l'utilisation de l'histogramme de couleur comme critère de choix entre les pistes qui favorise l'information couleur par rapport aux autres descripteurs. Pour pallier ces problèmes, nous avons proposé l'approche de fusion dynamique que nous évaluons dans la section suivante.

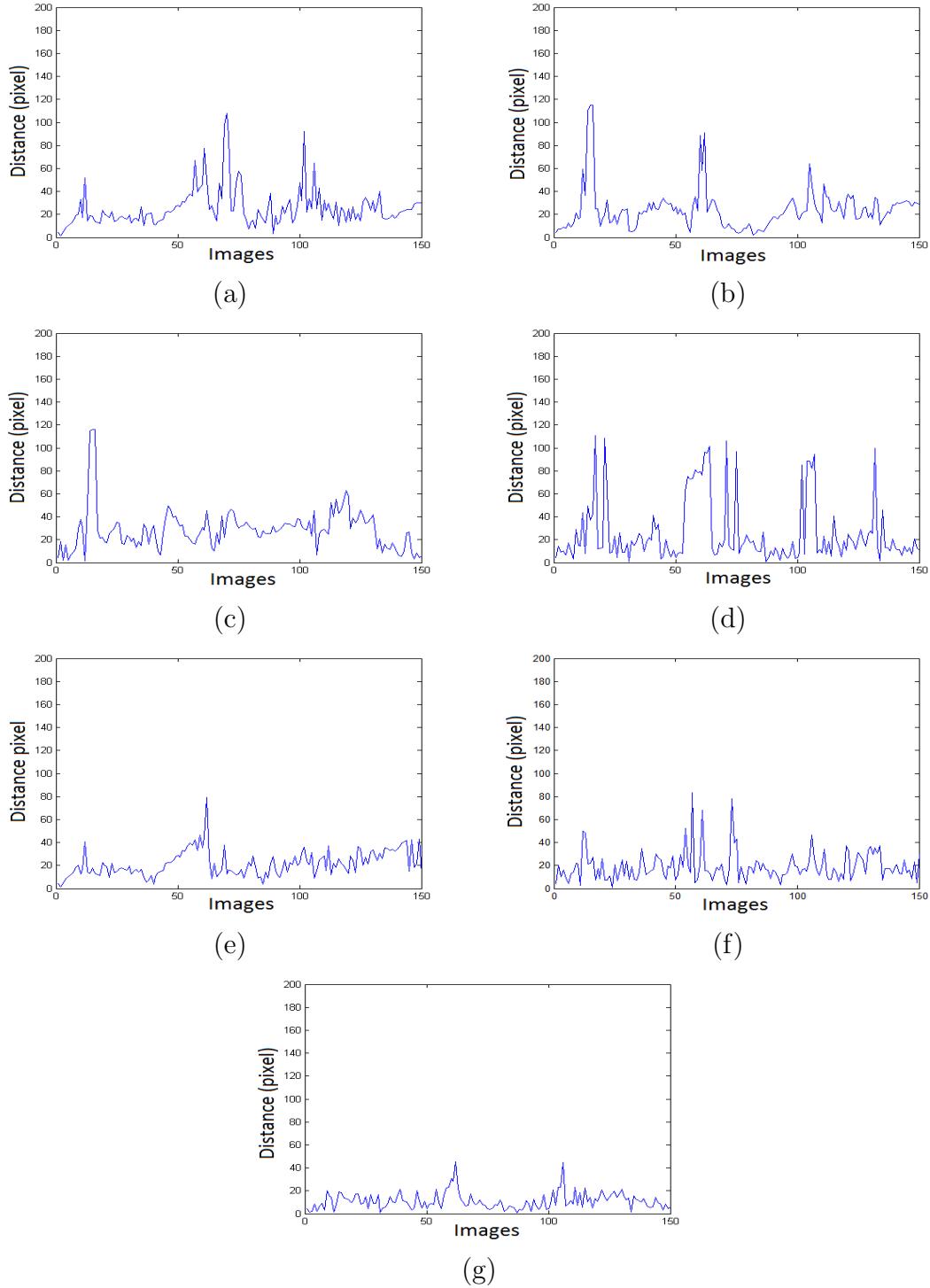


FIGURE 5.10 – Courbes de distance entre les positions détectées en utilisant les différentes approches de suivi et la vérité terrain. (a) : NL-JTC (b) : Histogramme de couleur (c) : HOG (d) : LBP uniforme rotation-invariant (e) : Multipiste NL-JTC (f) : Multipiste multidescripteur (g) : Minimum (optimal).

5.3.4 Evaluation de l'approche de suivi à base de la fusion dynamique

Dans le but de mieux fusionner les données, nous avons proposé de remonter dans le niveau de fusion. Par conséquent, nous remontons de la fusion des décisions, proposée dans l'approche multipiste, à la fusion des plans de scores, voir chapitre 1 section 1.7.2. Pour cela, nous fusionnons le plan de corrélation NL-JTC et le plan de scores de la technique basée sur l'histogramme de couleur. Ce choix vient de la complémentarité que nous avons remarquée entre ces deux techniques. En effet, La corrélation a montré une bonne performance en terme de précision pour la localisation mais elle reste sensible par rapport à la rotation et à la déformation des contours. D'autre part, la technique basée sur l'histogramme de couleur a montré une robustesse pour le suivi des nageurs mis à part dans le cas d'une scène contenant des objets de la même couleur que la cible. De plus, cette technique a montré un manque de précision en terme de localisation. Donc, afin de générer un descripteur riche nous avons proposé l'approche de fusion dynamique [4], détaillée dans le chapitre 4, section 4.4.

Afin d'évaluer notre approche de fusion dynamique, nous allons la comparer avec les approches suivantes : corrélation NL-JTC et histogramme de couleur. Ces tests sont effectués dans la piscine de Foch - Brest (25 m) sur deux séquences vidéo de deux nageurs professionnels du Club Nautique Brestois. La première séquence est un aller d'une course de nage libre alors que la seconde est un aller d'une course de dos. Les séquences vidéo sont enregistrées avec une résolution $4K$ (3840x2160 pixels) en 25 images par seconde sur une durée de 15 secondes chacune. Les paramètres utilisés pour les différentes expérimentations sont :

- Taille de la cible :

$$\begin{cases} l_c = 5 * l_r \\ h_c = 3 * h_r \end{cases} \quad (5.5)$$

où, $l_c \times h_c$ et $l_r \times h_r$ représentent, respectivement, la taille de la région d'intérêt et la taille de la référence. Ce choix est expliqué par le fait que le nageur se déplace horizontalement dans le couloir et qu'il ne se déplace que très légèrement sur la verticale.

- Valeur du degré de non-linéarité de la méthode NL-JTC : $k = 0.3$. Ceci permet d'obtenir un bon compromis entre bruit et finesse des pics dans le plan de corrélation.
- Nombre de composantes pour l'histogramme de couleur : $Nb_{bin} = 16$.
- Facteur de pondération pour la fusion par sommation pondérée : $\alpha = 0.4$. Ce choix se justifie par le fait que nous faisons un peu plus confiance à la technique utilisant l'histogramme de couleur. Ceci permet d'atténuer le bruit du plan de corrélation et donc d'avoir un plan de fusion moins bruité.
- Nombre de pixels estimé pour la zone présentant des artéfacts de corrélation

de bords à enlever : 5 pixels. Ces pixels sont enlevés tout autour de la zone potentielle de corrélation ainsi que sur le plan de scores couleurs afin de pouvoir les fusionner.

Dans le but de comparer quantitativement les différentes approches, nous proposons d'utiliser trois mesures : le pourcentage de suivi, le PCE et l'écart-type local.

1. Le pourcentage de suivi est obtenu en calculant la distance euclidienne entre les coordonnées des détections issues de chacune des trois approches et une vérité terrain préalablement établie manuellement pour chaque image des séquences vidéo.
2. Le PCE sert à mesurer l'énergie du pic par rapport à l'énergie de tout le plan.
3. L'écart-type local permet de mesurer le taux de bruit localement dans le plan afin d'estimer la précision et la robustesse de la méthode vis-à-vis de la localisation. En effet, Une petite valeur d'écart-type local implique une confiance plus importante envers la détection à cause de l'existence d'un bruit peu élevé dans le plan et vice versa. La valeur de l'écart-type local (*Local - STD*) que nous avons proposé est calculée par la formule suivante :

$$Local - STD = \frac{1}{L} \sum_{l=1}^L \sqrt{\frac{1}{n * m} \sum_{i=1, j=1}^{n, m} (x_l(i, j) - \bar{x})^2} \quad (5.6)$$

avec L le nombre de blocs dans le plan sachant que la taille de chaque bloc est $n \times m$. $x_l(i, j)$ représente la hauteur du point ayant les coordonnées (i, j) dans le bloc l . \bar{x} est la moyenne des valeurs du bloc.

En plus de ces trois critères, la hauteur du pic du plan de fusion peut être utilisée comme facteur de confiance. En effet, grâce à la normalisation du plan, celui-ci est compris dans un intervalle $[0, 1]$. Ainsi, un pic de fusion proche de 1 correspond à une fusion de deux pics d'intensité maximale localisés exactement au même endroit et donc une confiance très élevée. En revanche, si les pics sont situés à des endroits différents, le pic de fusion sera inférieur à 1 et la confiance sera moins importante. Nous n'utiliserons cependant pas ce critère d'évaluation car il n'existe pas pour les deux approches de base. Néanmoins, nous l'utilisons comme facteur de confiance pour valider la détection par l'approche de la fusion dynamique.

Afin de mieux comprendre les deux derniers critères d'évaluation : le PCE et l'écart-type, nous présentons dans la figure 5.11 et la figure 5.12 des courbes montrant des comparaisons entre les trois approches : la fusion dynamique, la corrélation NL-JTC et l'histogramme de couleur pour les deux critères cités précédemment.

Les courbes de la figure 5.11 montrent que l'approche proposée améliore significativement le critère PCE. Ceci vient du fait que le plan de fusion contient souvent un

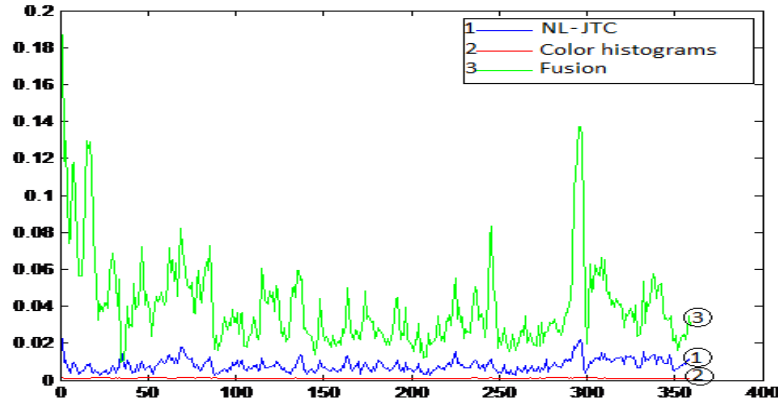


FIGURE 5.11 – Comparaison entre la corrélation NL-JTC, la technique de l'histogramme de couleur et la fusion dynamique en terme de PCE sur une séquence vidéo de course (dos) de 360 images. Entraînement - Piscine de Foch, Brest 2015.

pic fin et moins de bruit donc une décision à la fois précise et robuste. D'autre part, l'approche de corrélation génère souvent un plan bruité donc moins robustesse comme le montre la mesure du PCE. En ce qui concerne la dernière approche, basée sur la couleur, le PCE obtenu est très faible à cause de la largeur importante des pics qui induisent une détection robuste mais une mauvaise précision en terme de localisation.

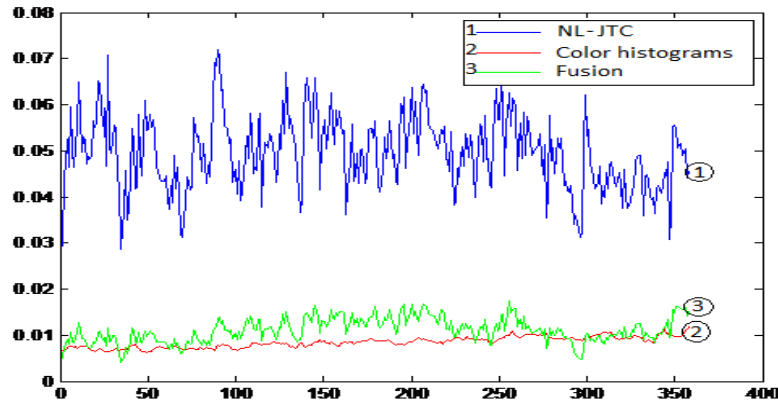


FIGURE 5.12 – Comparaison entre la corrélation NL-JTC, la technique de l'histogramme de couleur et la fusion dynamique en terme d'écart-type local sur la même séquence vidéo (course de dos) de 360 images. Entraînement - Piscine de Foch, Brest 2015.

Les courbes de la figure 5.12 montrent que l'approche de fusion dynamique ainsi que l'histogramme de couleur génèrent des plans peu bruités comparé à la corrélation

NL-JTC. Ceci est expliqué par les valeurs d'écart-type local qui sont faibles pour les deux premières approches impliquant moins de bruit et plus de robustesse. Cependant, pour la corrélation NL-JTC, nous constatons des valeurs plus élevées à cause du bruit dans le plan expliquant que cette approche est moins robuste par rapport comparée aux autres. Le tableau 5.5 résume les résultats des tests sur les deux vidéos présentées précédemment :

Séquences	Critères de comparaison	NL-JTC	Histogramme de couleur	Fusion dynamique
Dos	Pourcentage de suivi	68.73	60.00	73.09
	PCE (*10e-2)	2.00	0.20	6.00
	Ecart-type local (*10e-2)	4.97	0.86	1.13
Nage libre	Pourcentage de suivi	54.17	57.22	71.74
	PCE (*10e-2)	2.00	0.20	4.60
	Ecart-type local (*10e-2)	5.13	1.44	1.88

TABLE 5.5 – Résultats des comparaisons entre la corrélation NL-JTC, l'histogramme de couleur et la fusion dynamique en termes de pourcentage de suivi, PCE (valeur moyenne des PCE) et écart-type local (valeur moyenne des écart-types). Entraînement - Piscine de Foch, Brest 2015.

Le tableau 5.5 montre clairement que la fusion a significativement amélioré le pourcentage de suivi ainsi que le PCE et l'écart-type local. Pour la première vidéo (course de dos) le pourcentage de suivi est passé de 68.73% pour la corrélation NL-JTC et 60% pour l'histogramme de couleur à 73.09% pour la fusion dynamique. Même constat pour la deuxième vidéo (nage libre) où le pourcentage de suivi est amélioré de 54.17% pour la corrélation NL-JTC et 57.22% pour l'histogramme de couleur à 71.74% pour la fusion dynamique. Ceci est expliqué par le fait que cette dernière approche tire partie des avantages des deux autres techniques afin de générer une description riche et plus robuste de la cible à suivre (couleur + forme).

De plus, on peut noter que l'amélioration des valeurs de PCE reflètent une finesse plus importante des pics et donc un suivi plus précis. D'autre part, les valeurs de l'écart-type local montrent la distribution du bruit dans les plans. Leur amélioration implique une atténuation du bruit dans notre approche de fusion dynamique comparé à la corrélation. On note aussi que le plan de scores couleurs reste légèrement moins bruité que celui de la fusion.

Finalement, les deux figures 5.11 et 5.12 ainsi que le tableau 5.5 montrent l'apport de l'approche de fusion dynamique par rapport aux techniques de base non seulement en terme de pourcentage de suivi mais aussi en terme de confiance envers la décision.

Il est important de noter que les cas jugés comme mauvais suivis sont souvent liés à l'invisibilité de l'objet à suivre dans la région d'intérêt. En effet, la limitation du système

de suivi de nageur à base de fusion dynamique réside principalement dans l’occultation quasi-totale et totale de la tête à suivre. Afin de remédier à ce problème nous avons proposé l’approche multizone [5] que nous évaluons par la suite.

5.3.5 Evaluation de l’approche de suivi multizone

Dans le but d’optimiser le système de suivi basé sur la fusion dynamique, nous nous sommes inspirés du concept du multipiste. Dans ce contexte, nous avons choisi deux zones à suivre : la tête et le maillot de bain. La tête représente la zone la plus rigide et la plus visible du corps. Tandis que le maillot de bain, malgré sa plus faible rigidité, présente l’avantage d’être généralement visible lorsque la tête ne l’est pas. Par conséquent, notre idée a été de développer une approche multizone où chaque piste représente le suivi d’une zone (tête ou maillot de bain) et où l’on utilise l’approche de suivi basée sur la fusion dynamique. Il est à noter que chaque piste génère toutes les détections potentielles, et qu’ensuite le meilleur couple tête - maillot de bain est choisi selon des critères bien définis comme : la distance entre les deux zones, les facteurs de confiance ou encore la vitesse de l’ensemble.

L’objectif de l’approche multizone, que nous avons proposée dans le chapitre 4, section 4.5, consiste principalement à optimiser le suivi de nageurs dans le cas des occultations des cibles à suivre. Afin de montrer l’apport de cette approche, nous allons la comparer avec l’approche de suivi de la tête du nageur (tête seule) basée sur la fusion dynamique validée dans la section précédente avec une limitation signalée dans le cas des occultations.

Afin d’évaluer cette approche, nous avons effectué plusieurs tests sur des séquences vidéo de nage libre : Entraînements (piscine de 25 mètres, Foch, Brest) et Compétitions nationales (piscine de 50 mètres, Limoges, 2015). En ce qui concerne la prise de vue dans le bassin de 25 mètres, une seule caméra *Blackmagic 4K* est utilisée pour capturer tout le bassin. En revanche, pour une prise de vue de bonne résolution dans le bassin de 50 mètres, nous avons utilisé le système de prise de vue 8K (2 caméras *Blackmagic 4K* où chacune couvre la moitié du bassin).

Le tableau 5.6 présente une comparaison des performances du suivi des nageurs utilisant l’approche multizone (tête et maillot de bain) et le suivi de la tête seule par fusion dynamique.

N° séquence	Type de nage	Tête seule	Multizone
Séquence 1	Entraînement – 100 m nage libre	92.24	97.78
Séquence 2	Entraînement – 100 m nage libre	90.10	96.55
Séquence 3	Compétition – 1500 m nage libre	65.77	87.92
Séquence 4	Compétition – 1500 m nage libre	72.84	90.60
Séquence 5	Compétition – 1500 m nage libre	67.79	93.96

TABLE 5.6 – *Comparaison entre l’approche de suivi multizone et le suivi de la tête seule par fusion dynamique en terme de pourcentage de suivi. Entraînements : piscine de Foch - Brest. Compétitions : Championnat national de natation - Limoges 2015.*

Les résultats obtenus dans le tableau 5.6 montrent que l’utilisation de l’approche multizone améliore significativement le pourcentage de suivi. Ceci est validé sur 5 nageurs durant plusieurs séquences de nage libre de différents niveaux (entraînements, compétitions). En particulier, on note que cette approche utilisée sur des séquences d’entraînements améliore légèrement le pourcentage de suivi de 5% et 6%. En effet, dans ce cas, le nageur avance lentement avec une vitesse moyenne de ~ 1 m/s et, par conséquent, il y a moins d’éclaboussures. De ce fait, l’amélioration apportée par l’approche multizone ne concerne que les occultations partielles liées aux passages de la main ainsi que l’augmentation de la précision de la localisation.

Cependant, dans le cas des séquences de compétitions 1500 m nage libre, les athlètes nagent un peu plus vite à une vitesse moyenne de ~ 1.5 m/s. De ce fait, ils génèrent plus d’éclaboussures dégradant le suivi de la tête seule et en même temps, montre l’intérêt de l’approche proposée. En effet, le pourcentage de suivi est amélioré de 16%, 18% et 22% pour les différents nageurs. Ceci est expliqué par le fait que l’approche multizone permet d’éviter le décrochage du suivi dans les cas d’occultations de la tête en estimant sa position à partir du maillot de bain.

Les deux figures 5.13 et 5.14 présentent deux courbes de distances entre les positions de la tête détectée par chacune des approches de suivi et la vérité terrain.

Ces deux courbes présentent visuellement les résultats du tableau 5.6 concernant la séquence 1 - entraînement – 100 m nage libre et la séquence 3 - compétition – 1500 m nage libre. Nous remarquons dans les deux cas que la courbe correspondant à l’approche multizone est plus proche de la vérité terrain comparé au suivi de la tête seule par fusion dynamique. Ce constat est plus visible pour la figure 5.14 de compétition où le nageur avance rapidement générant des éclaboussures qui causent plusieurs occultations de la tête. Dans ce cas, l’approche multizone montre de bons résultats grâce à une meilleur estimation de la position du nageur durant les phase d’occultation.

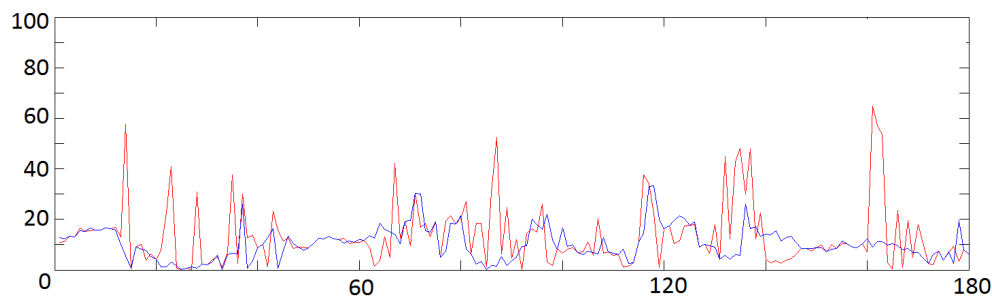


FIGURE 5.13 – *Evaluation de l’approche de suivi multizone (en bleu) et de suivi de la tête seule par fusion dynamique (en rouge) en terme de distance par rapport à la vérité terrain. Type de nage : 100 m nage libre – séquence 1 (180 images) - entraînement - piscine Foch de Brest.*

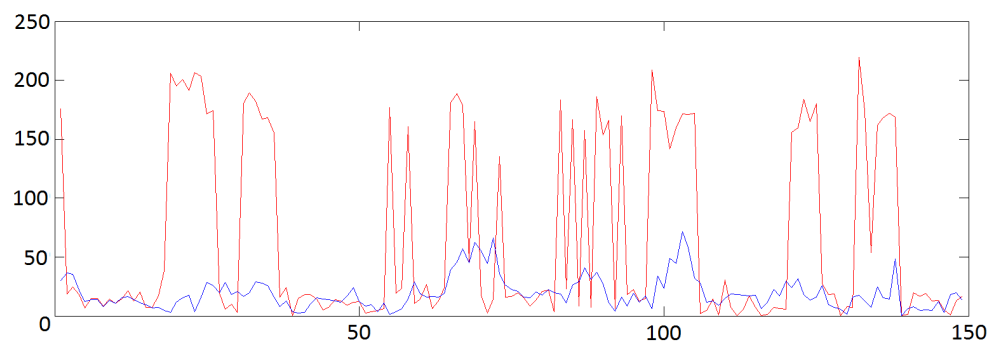


FIGURE 5.14 – *Evaluation de l’approche de suivi multizone (en bleu) et de suivi de la tête seule par fusion dynamique (en rouge) en terme de distance par rapport à la vérité terrain. Type de nage : 1500 nage libre – séquence 3 (150) - Championnat national de natation, Limoges 2015.*

5.3.6 Discussion

Dans cette partie, nous avons présenté une évaluation des approches proposées pour le suivi. Nous avons commencé par présenter les résultats liés au multipiste qui consiste en une fusion de décisions de plusieurs pistes de suivi. Ensuite, afin de l’optimiser, nous avons proposé et validé l’approche de fusion dynamique qui consiste à fusionner les plans de corrélation NL-JTC et d’histogramme de couleur. Enfin, dans le but de résoudre le problème des occultations, nous avons proposé et validé l’approche multizone qui consiste à suivre la tête et le maillot de bain en même temps associé à un critère complexe permettant de choisir le meilleur couple tête - maillot de bain afin d’éviter les décrochages de suivi causés par les occultations.

Il est à noter que le multizone est l’approche validée et retenue pour le suivi des nageurs. Par la suite, nous présentons une étude préliminaire des performances des

nageurs basée sur le suivi multizone et sur l’annotation manuelle des nageurs.

5.4 Etude de performances de nageurs

5.4.1 Introduction à l’étude de performances de nageurs

Durant les dernières décennies, de nombreuses recherches en biomécanique ont été effectuées [106–109]. Ceci dans le but d’analyser la stratégie de nage et de fournir des retours immédiats aux athlètes concernant leurs performances durant les entraînements et les compétitions.

Des études récentes [108, 109] ont montré que les nageurs les plus rapides adoptent la stratégie du rythme stable en évitant les variations inutiles. Cette stratégie permet l’utilisation optimale de l’énergie. Par exemple, dans les courses 200 m nage libre, on peut remarquer qu’un nageur de haut niveau nage un premier tour (1^{er} 50 m) avec un rythme rapide constant, puis les deux prochains tours (2^{ème} et 3^{ème} 50 m) avec un rythme constant mais moins rapide pour finir avec un dernier tour (4^{ème} 50 m) avec un rythme identique où légèrement plus rapide que les deux précédents.

D’autres études ont montré l’intérêt de la phase de plongeon et de retournement et leurs influences sur les performances finales de l’athlète [91, 110]. Cependant, la phase la plus importante reste la phase de nage située entre la phase de plongeon/retournement et le retournement. Durant cette phase, différents critères peuvent être analysés [106, 110] : SL (*Stroke Length*, longueur de la frappe), SR (*Stroke Rate*, taux de frappes), vitesse instantanée, vitesse inter-cycles, etc.

Dans le contexte de cette thèse, notre objectif consiste à analyser la vitesse du nageur durant la phase de nage. Pour cela, nous nous intéressons, par la suite, à la vitesse cyclique, intra-cyclique et instantanée calculées en utilisant l’annotation manuelle et le système de suivi automatique que nous avons développé. La vitesse cyclique sert à évaluer globalement le rythme de nage durant la course. La vitesse intra-cyclique permet, quant à elle, d’analyser la force motrice de la frappe de chaque bras du nageur et son effet sur la vitesse. Nous finirons par la vitesse instantanée qui donne une estimation détaillée de l’effet de chaque mouvement du nageur.

5.4.2 Vitesse cyclique et intra-cyclique

Pour la nage libre et le dos, le cycle de nage correspond à deux frappes de mains successives tandis que le demi-cycle correspond à une seule frappe. Pour la brasse et

le papillon, un cycle correspond à la frappe des deux mains et un demi-cycle consiste en l'une des deux phases suivante : en dehors de l'eau et sous l'eau. Ces mesures sont calculées, pour le moment, en utilisant l'annotation manuelle des moments clés. Ces derniers correspondent au moment où la main frappe l'eau. Cela veut dire, pour la nage libre par exemple, que nous annotons les frappes d'une seule main afin d'obtenir la vitesse de chaque cycle et nous annotons les frappes des deux mains afin d'obtenir la vitesse de chaque demi-cycle.

Dans les figure 5.15 et 5.16 nous présentons des courbes de vitesse cyclique et intra-cyclique d'un nageur durant un aller d'une course de nage libre 1500 m (Championnat national de natation, Limoges 2015).

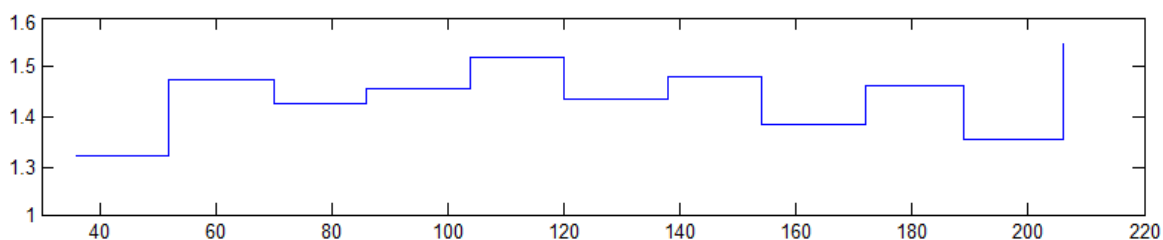


FIGURE 5.15 – *Vitesse intra-cyclique d'un nageur durant un aller 1500 m nage libre. Championnat national de natation, Limoges 2015.*

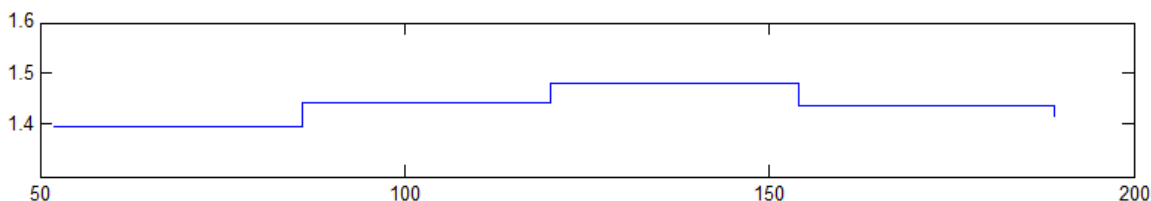


FIGURE 5.16 – *Vitesse cyclique du même nageur que la figure 5.15 durant la même séquence aller 1500 m nage libre. Championnat national de natation, Limoges 2015.*

Afin d'interpréter ces résultats, nous avons fait étudier avec des experts en biomécanique et des entraîneurs de l'équipe de France les deux courbes. La courbe de la figure 5.15 montre une variation intra-cyclique intéressante. En effet, nous remarquons dans la première partie (après le plongeon) des demi-cycles avec des vitesses relativement proches. Ensuite, à partir du milieu de l'aller nous remarquons des variations significatives entre les demi-cycle ce qui veut dire que le nageur a changé sa stratégie en appuyant sur un bras plus que sur l'autre.

En analysant la deuxième courbe de vitesse cyclique présentée dans la figure 5.16, nous obtenons une idée globale concernant la stratégie d'avancement du nageur. Ce

dernier nage avec un rythme relativement stable (une variation ≤ 0.1 m/s) avec une petite accélération (0.08 m/s) jusqu'au milieu puis une petite décélération (0.06 m/s) à la fin de l'aller avant le retournement. Ces résultats ont fait l'objet d'une étude biomécanique effectuée au niveau du département de la recherche de la Fédération Française de Natation [111] concernant les variations du rythme selon le niveau, l'âge et le sexe.

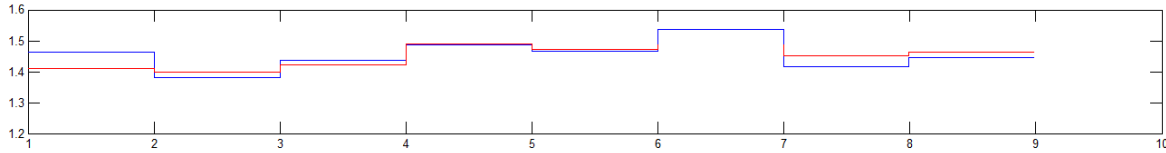


FIGURE 5.17 – Comparaison entre les courbes de vitesse intra-cyclique issues de l'annotation manuelle (en bleu) et du suivi automatique (en rouge) d'un nageur durant un aller 1500 m nage libre. Championnat national de natation, Limoges 2015.

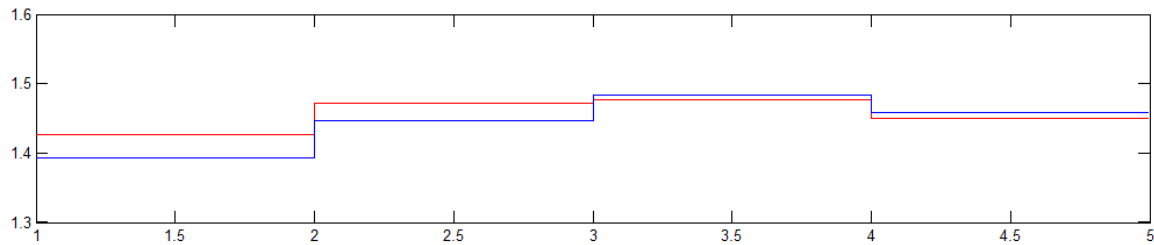


FIGURE 5.18 – Comparaison entre les courbes de vitesse cyclique issues de l'annotation manuelle (en bleu) et du suivi automatique (en rouge) d'un nageur durant un aller 1500 m nage libre. Championnat national de natation, Limoges 2015.

Les figures 5.17 et 5.18 présentent, respectivement, des courbes de vitesse intra-cyclique et cyclique d'un nageur pendant des entraînements nage libre. Ces courbes sont calculées en utilisant la vérité terrain (en bleu) et le suivi automatique (en rouge). Ces résultats montrent que nous avons obtenu le même profil avec de petites imprécisions. Le risque lié à l'utilisation du suivi automatique pour le calcul de ces mesures est que celles-ci sont basées sur l'utilisation de deux points pour le calcul de la vitesse de tout le cycle (2 points désignant le début et la fin du cycle). De ce fait, dans le cas où l'un des deux points est mal détecté, à cause d'une occultation par exemple, la vitesse cyclique pourra être faussée. Pour pallier ce problème, nous estimons la position du nageur pour l'image i en moyennant les positions détectées dans les images $i - 2 \rightarrow i + 2$. Ceci nous a permis de retrouver le même profil avec de petites erreurs dans les valeurs de vitesse des cycles. Actuellement, les études biomécaniques sont faites sur les courbes issues de la vérité terrain en raison de leur précision. Notons que la minimisation de la distance entre l'approche manuelle et l'approche automatique reste une perspective de recherche à court terme.

5.4.3 Vitesse instantanée

Le record mondial du 50m nage libre messieurs est battu par César Cielo au championnat du Brésil en 2009. Il a nagé à une vitesse moyenne de 2.39 m/s. En analysant les courbes de vitesse instantanée issues de la vérité terrain, nous avons remarqué qu'elle varie approximativement entre ± 0.4 m/s. Ceci veut dire que la marge de variation significative de la vitesse (V_s) varie autour de 0.8 m/s (marge entre le pic et le creux), comme le montre la figure 5.19. Ainsi, l'erreur tolérée pour la vitesse doit être estimé par rapport à cette valeur.

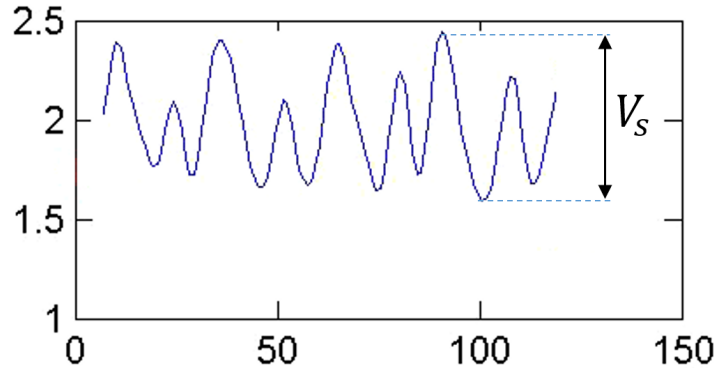


FIGURE 5.19 – Courbe de vitesse instantanée de César Cielo issue de la vérité terrain calculée sur une séquence de 120 images. V_s : variation significative de la vitesse instantanée. 50 m nage libre - Championnat du monde de natation, Barcelone 2013.

L'estimation de la vitesse instantanée se fait en calculant la distance parcourue par le nageur δX dans une très courte durée δt . Dans notre cas, nous considérons $\delta t = 1/T$ avec, T le taux d'images par seconde. Par conséquent, la vitesse instantanée V_i est calculée comme suit :

$$V_i = \frac{\delta X}{\delta t} \quad (5.7)$$

Dans cette partie, nous évaluons la précision des courbes de vitesse instantanée issues du suivi automatique basé sur l'approche multizone. Ceci se fait en les comparant avec celles issues de la vérité terrain. Pour ce faire, nous faisons une analyse statistique de la distance entre les courbes selon les deux critères suivants : la moyenne et l'écart-type. Le tableau 5.7 montre les valeurs de ces critères calculés sur les 5 séquences figurant dans le tableau 5.6.

N° séquence	Vitesse instantanée	
	Moyenne (m/s)	Ecart-type (m/s)
Séquence 1	0.62	0.48
Séquence 2	0.6	0.55
Séquence 3	0.63	0.64
Séquence 4	0.49	0.38
Séquence 5	0.44	0.35

TABLE 5.7 – *Evaluation de la vitesse instantanée issue du suivi multizone selon la moyenne et l'écart-type de sa distance par rapport à celle issue de la vérité terrain. Championnat national de natation, Limoges 2015.*

Ce tableau montre que les moyennes et les écarts-type de ces distances, pour les 5 séquences, varient respectivement entre (0.44 m/s et 0.62 m/s) et (0.35 m/s et 0.48 m/s). La valeur de la moyenne donne une idée globale sur les distances entre les deux courbes. Tandis que, la valeur de l'écart-type reflète la variation entre ces distances. Ces valeurs reflètent une erreur d'estimation de la vitesse relativement élevée par rapport à l'erreur admise maximale de $V_s \sim 0.6 \text{ m/s}$, malgré le fait que le suivi multizone ait été validé avec la vérité terrain.

Ces erreurs peuvent être expliquées par les petites imprécisions de quelques pixels au niveau du suivi automatique ainsi qu'au niveau de l'annotation manuelle de la tête. Ceci peut avoir des répercussions significatives après le passage des coordonnées pixels aux coordonnées métriques, comme nous l'avons signalé dans la section 5.3.1.4.

Cependant, ces imprécisions ne se cumulent pas quelque soit le système utilisé : vérité terrain ou suivi automatique. Pour le système basé sur la vérité terrain, une erreur minime d'annotation sera corrigée par la prochaine annotation. Même chose concernant notre système basé sur le suivi automatique qui est suffisamment robuste pour ne pas présenter de décrochages permanents. Pour cette raison, les courbes de vitesse brutes bruitées doivent être filtrées afin d'en extraire une information pertinente ressemblant à celle mesurée par la CIREN.

5.4.4 Filtrage des courbes de vitesse instantanée

Notre objectif dans cette partie est de générer des courbes de vitesse instantanée pertinentes afin de comprendre la stratégie de nage et l'influence des mouvements effectués par le nageur. Comme nous l'avons mentionné dans la section précédente, les courbes de vitesse instantanée brute sont très bruitées. Afin d'atténuer ce bruit et lisser ces courbes, nous appliquons un filtrage de type gaussien. Ce choix est fait dans le but de retrouver un profil de vitesse comparable à celui de la centrale inertielle CIREN

comme visible sur les figures 5.7 et 5.8 où nous remarquons des courbes contenant des pics gaussiens.

Par la suite, nous appliquons un filtrage gaussien sur les courbes de vitesse des deux séquences 1 et 3. Pour cela, nous utilisons des fenêtres gaussiennes de différentes tailles comme présenté dans le tableau 5.8.

N° séquence	Taille de la fenêtre	Multizone	
		Moyen (m/s)	Ecart-type (m/s)
Séquence 1	1	0.62	0.48
	15	0.22	0.25
	25	0.11	0.1
Séquence 3	1	0.63	0.64
	15	0.22	0.16
	25	0.15	0.12

TABLE 5.8 – Distance entre les courbes de vitesse instantanée issues du suivi automatique et de la vérité terrain après l'application du filtre gaussien avec des fenêtres de différentes tailles. Championnat national de natation, Limoges 2015.

Les résultats du tableau 5.8 montrent que l'application du filtre gaussien réduit la distance entre les courbes de vitesse instantanée issues du suivi automatique et de la vérité terrain pour les deux séquences 1 et 3. L'augmentation de la taille de la fenêtre engendre une diminution de la distance entre les courbes. La figure 5.20 montre que le profil des courbes de vitesse tend vers le profil fourni par la centrale inertielle CIREN. Il est à noter que nous retenons la fenêtre de taille de 25 pour filtrer les courbes de vitesse car ceci permet l'atténuation du bruit et l'obtention d'un profil qui reflète bien l'avancement du nageur.

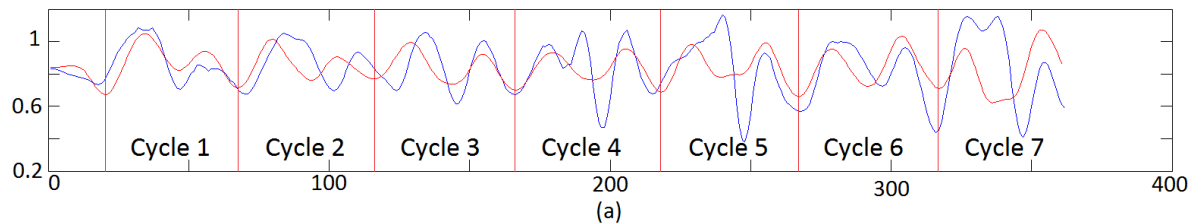


FIGURE 5.20 – Comparaison entre les courbes de vitesse instantanée issues du suivi automatique (en bleu) et de la vérité terrain (en rouge) après l'application du filtre gaussien avec une fenêtre de taille 25. Séquence 1 - championnat national de natation, Limoges 2015.

La figure 5.20 montre que la courbe de vitesse instantanée issue du suivi automatique génère un profil ressemblant à celui issu de la vérité terrain avec quelques erreurs. Dans

ces courbes de vitesse instantanée de nage libre, le cycle est représenté par deux pics où chacun correspond à une frappe d'un bras dans l'eau. Ces courbes montrent une bonne ressemblance pour les cycles 1, 2, 3 et 6 avec quelques erreurs minimales qui n'affectent pas le profil de nage. En revanche, nous remarquons une erreur d'un pic supplémentaire pour les cycles 4 et 7 ainsi qu'une déformation du premier pic du cycle 5. Ces erreurs sont expliquées par les erreurs du suivi automatique correspondant aux occultations totales.

5.4.5 Exemples d'application réelle

Notre application finale consiste à fournir aux entraîneurs et aux athlètes des retours sur la nage en indiquant la vitesse instantanée du nageur dans la courbe par un point interactif qui se déplace à chaque instant. Nous affichons aussi la région d'intérêt contenant seulement le nageur concerné. Ceci permet d'analyser le mouvement effectué par le nageur et en même temps son effet instantané sur l'avancement du nageur dans la courbe de vitesse instantanée.

Par la suite, nous présentons des figures contenant des cas d'application réelle analysés et validés par des entraîneurs professionnels. Ces figures présentent les 4 moments clés d'un cycle choisi aléatoirement dans la courbe de vitesse ainsi que les mouvements correspondants effectués par le nageur.

Pour les courses évaluées dans les figures 5.21, 5.22 et 5.23, nous remarquons une certaine ressemblance entre les profils de nage où un cycle correspond à deux creux et deux pics. Ceci est dû à l'alternance entre les frappes des bras dans les courses de types nage libre et dos. En effet, chaque creux représente une décélération causée par la frappe de la main sur l'eau. Cependant, chaque pic correspond à la phase où une des bras du nageur est sous l'eau ce qui provoque une accélération significative. L'amplitude de ces pics varient selon la puissance de frappe appliquée par le nageur.

Les figures 5.24 et 5.25 présentent des profils de nage de courses de types papillon et brasse. Pour le papillon, nous remarquons des cycles contenant deux pics d'amplitudes différentes. Le grand pic correspond au moment où la tête est en dehors de l'eau et la nageuse applique une frappe des deux bras pour aller sous l'eau (il s'agit ici d'une accélération de la tête plus que de la nageuse dans sa globalité). Le petit pic correspond au moment où les deux bras sont sous l'eau et la nageuse applique une force pour avancer. Ensuite, la nageuse se prépare pour sortir de l'eau ce qui la ralentit légèrement.

Pour la brasse, nous distinguons des cycles contenant un grand pic et un petit pic parfois plus large que le premier. Le grand pic correspond au moment où le nageur avance rapidement sous l'eau. Tandis que le petit pic correspond à toute la phase qui s'étend du ralentissement du nageur, avant sa sortie de l'eau, à son retour sous l'eau.

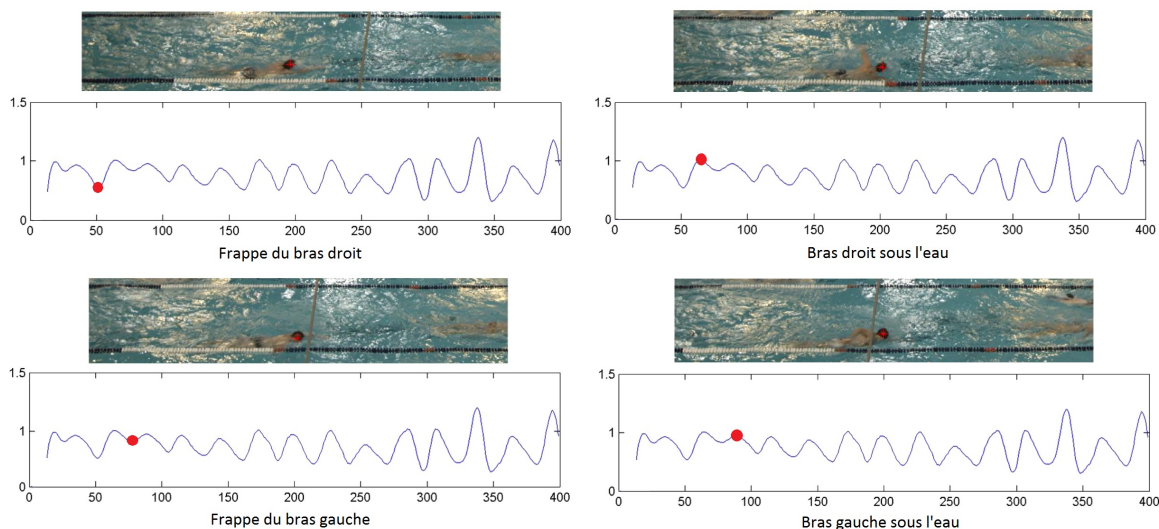


FIGURE 5.21 – Exemple de 4 moments clés extraits d’une vidéo d’analyse d’un nageur du Club Nautique Brestois durant l’entraînement. Suivi automatique - Piscine de Foch, Brest 2015.

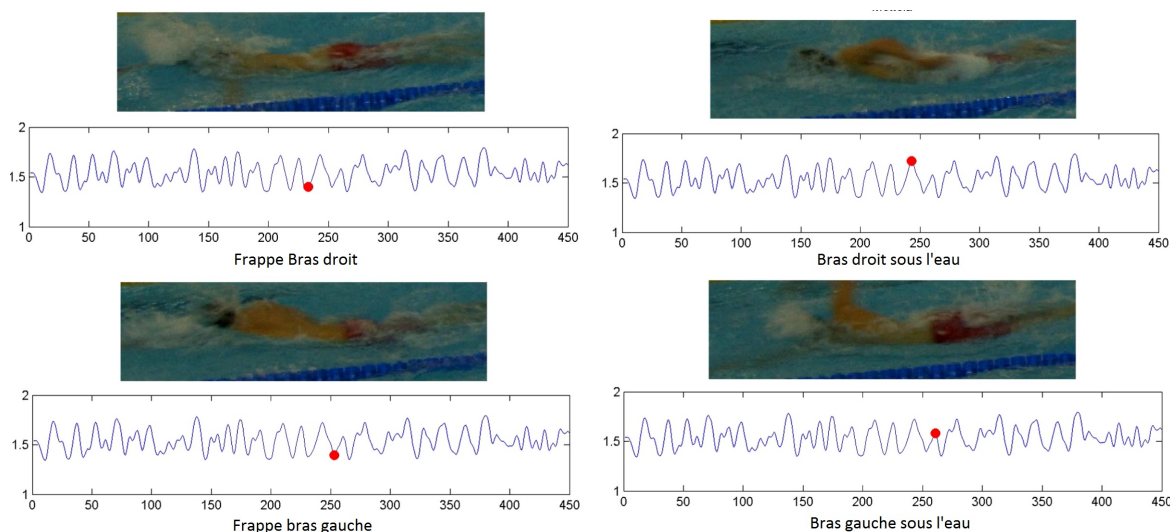


FIGURE 5.22 – Exemple de 4 moments clés extraits d’une vidéo d’analyse de Raymond Axel lors de la finale 800 m nage libre messieurs. Suivi automatique - Championnat national de natation, Limoges 2015.

Enfin, sur l’ensemble de ces figures, nous constatons une variation de la forme des cycles et de l’amplitude des pics. Ceci indique un changement de mouvement effectué qui peut être interprété par l’entraîneur en analysant l’image du nageur au même instant. Cette interprétation permet de déterminer les mouvements inutiles qui peuvent ralentir et fatiguer le nageur. Il est à noter que les figures 5.21, 5.22 et 5.23 correspondent à des courbes de vitesse issues du suivi automatique. En revanche, les courbes dans

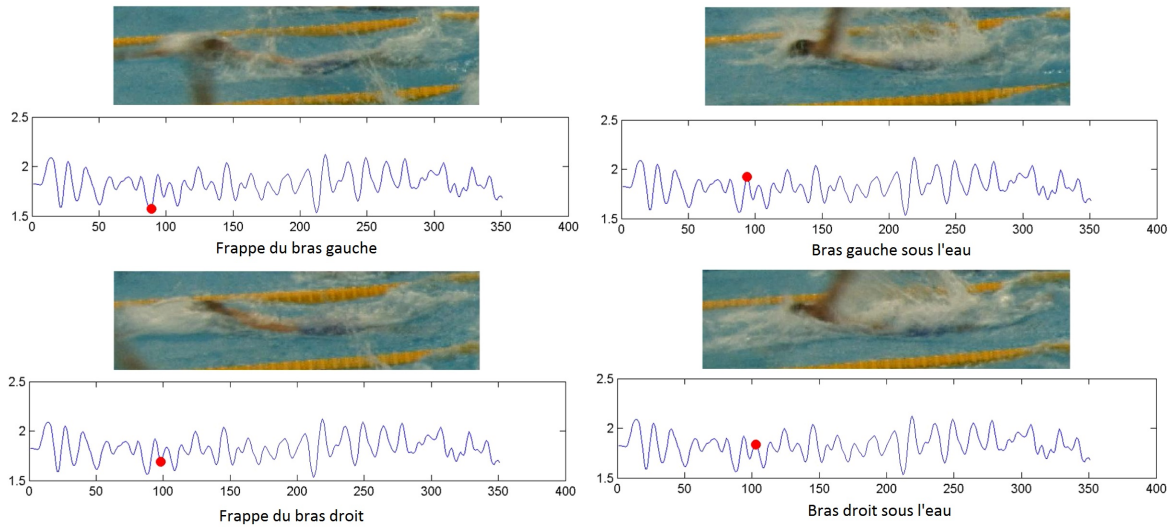


FIGURE 5.23 – Exemple de 4 moments clés extraits d’une vidéo d’analyse de Camille Lacourt lors de la finale 100 m dos messieurs. Suivi automatique - Championnat national de natation, Limoges 2015.

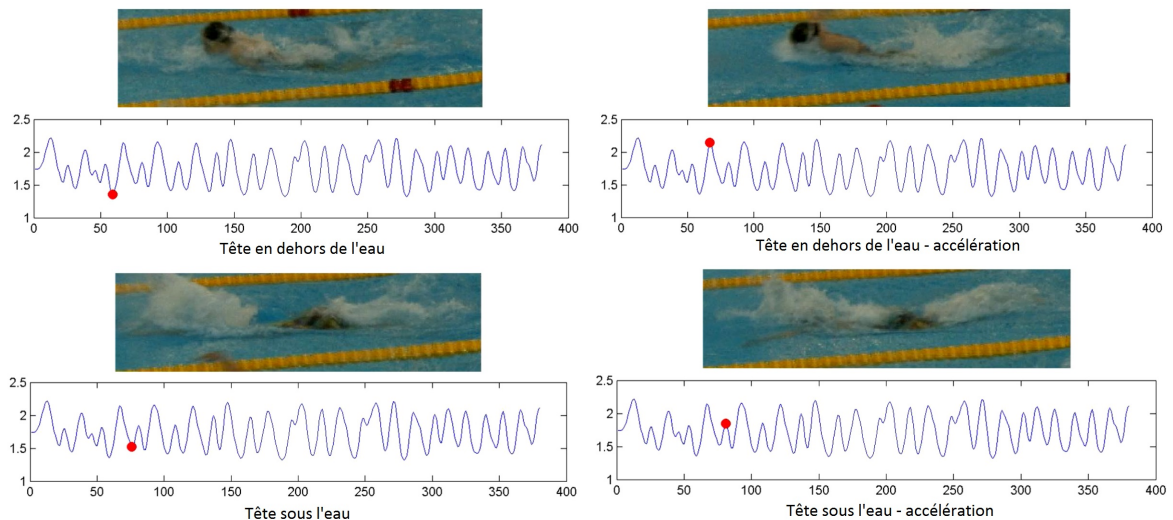


FIGURE 5.24 – Exemple de 4 moments clés extraits d’une vidéo d’analyse de Béryll Gastaldello lors de la finale 100 m papillon dames. Annotation manuelle - Championnat national de natation, Limoges 2015.

les figures 5.24 et 5.25 sont issues de l’annotation manuelle car celles issues du suivi automatique ont montré quelques erreurs spécifiquement dans le cas des vidéos de papillon ou de brasse. Ces dernières sont principalement liées au mouvement de la tête, en particulier, sa distance par rapport au plan de l’eau. Ceci génère de fausses variations de distance entre les deux zones (tête et maillot de bain) qui est supposée stable. Par conséquent, notre approche de suivi automatique multizone génère des erreurs aux

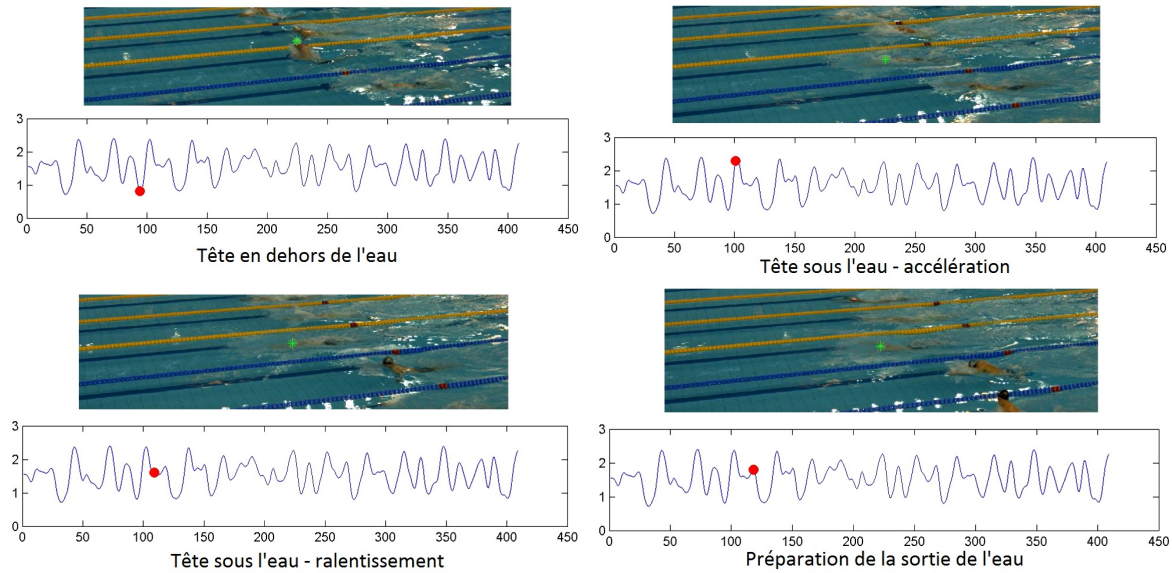


FIGURE 5.25 – Exemple de 4 moments clés extraits d’une vidéo d’analyse de Dahlia Thomas lors de la finale 200 m Brasse messieurs. Annotation manuelle - Championnat national de natation, Limoges 2015.

moments des occultations.

5.5 Conclusion

Dans ce chapitre, nous avons présenté le protocole expérimental utilisé pour les différents tests et nous avons proposé et validé un système de prise de vue 8K. Ce dernier est basé sur l’utilisation de deux caméras 4K positionnées dans les gradins de telle façon que chacune couvre la moitié du bassin. L’utilisation de ce système permet d’enregistrer des vidéos de bonne qualité facilitant la localisation du nageur.

L’adaptation des techniques de suivi simple (NL-JTC, histogramme de couleur, LBP et HOG) a montré des résultats insuffisants. Dans le but d’optimiser le suivi des nageurs, nous avons proposé plusieurs approches basées principalement sur la fusion de données et sur une logique décisionnelle complexe. Les tests effectués sur la première approche proposée, le multipiste, ont montré une amélioration intéressante au niveau du pourcentage de suivi. En revanche, ces résultats n’ont pas été suffisants pour l’étude des performances des nageurs. Par conséquent, nous avons proposé l’approche de fusion dynamique. Celle-ci consiste à fusionner le plan de corrélation avec le plan des scores couleurs qui permet de générer une description riche de la cible à suivre. Les tests effectués montrent une optimisation significative du suivi des nageurs.

Néanmoins, la limitation principale de cette approche reste les occultations de la cible à suivre. Afin de pallier ce problème, nous avons proposé l'approche multizone qui consiste à suivre deux zones en parallèle en utilisant l'approche de fusion dynamique : la tête et le maillot de bain. De plus un critère complexe a été proposé dans le but d'assurer un suivi fiable des nageurs. Cette approche a été évaluée par les différents tests effectués et a été validée et retenue pour le suivi et l'étude des performances des nageurs.

Nous avons conclu ce chapitre par une étude préliminaire des performances des nageurs. En particulier, nous nous sommes intéressés aux mesures de la vitesse cyclique, intra-cyclique et instantanée. Pour cela, nous nous sommes basés sur les résultats de l'annotation manuelle, validés pour tous les types de nage, et du suivi automatique qui ne sont validés que pour la nage libre et le dos. Pour la brasse et le papillon, les résultats du suivi automatique sont trop imprécis à cause du mouvement important du corps et particulièrement la distance importante entre la tête et le plan d'eau. Ceci reste dans nos perspectives de recherche que nous détaillons dans le prochain chapitre.

Conclusion et perspectives

Conclusion

Dans cette thèse, nous avons étudié la possibilité d'utiliser des capteurs de vision pour l'estimation et l'amélioration des performances des nageurs de haut niveau. Dans ce cadre, nous avons proposé une solution utilisant un suivi automatique basé sur des algorithmes de vision par ordinateur. Cette étude s'inscrit dans le cadre d'un projet proposé par la Fédération Française de Natation ayant pour but d'analyser les stratégies de nage adoptées par les nageurs afin d'aider les entraîneurs à cibler les faiblesses des sportifs et ainsi accroître leur niveau. Pour cela, notre recherche s'est organisée en trois grandes parties. Tout d'abord, nous avons commencé par le choix d'un système de prise de vue adapté aux besoins de notre projet. Ensuite, nous avons développé un système de suivi de nageurs optimisé permettant de surmonter les différentes difficultés liées à l'environnement aquatique et aux spécificités de la natation. Enfin, une étude préliminaire des performances des nageurs a été faite selon différentes mesures de vitesse : cyclique, intra-cyclique et instantanée.

Afin de choisir un système de prise de vue adapté, nous avons effectué plusieurs tests en utilisant différentes caméras : Caméra *IDS*, *Nikon D800* et *Blackmagic 4K*. Cette dernière a été retenue pour ses caractéristiques : très haute définition (3840 x 2160), 30 images par secondes, compression sans perte visuelle et facilité de mise en place. Pour les bassins de 25 mètres nous avons proposé d'utiliser un système mono-caméra placé au milieu des gradins. En revanche, pour les bassins de 50 mètres, ce système n'est pas suffisant car le suivi des nageurs nécessite une meilleure résolution. Dans ce cadre, nous avons proposé et validé un système de prise de vue *8K* qui se compose de deux caméras *4K*. Chacune de ces caméras couvre une moitié de bassin avec une zone de chevauchement au milieu afin de gérer le passage du nageur entre les champs de vision des deux caméras.

La partie principale de la thèse concerne le développement d'un système fiable et automatique de suivi des nageurs. Pour ce faire, nous avons commencé par le calibrage des vidéos qui représente une étape très importante car c'est elle qui permet de lier les

pixels de l'image à leurs valeurs métriques et donc le calcul de la vitesse. Ensuite, nous avons étudié, implémenté et adapté des techniques de suivi de la littérature, notamment celles basées sur la corrélation (JTC non linéaire), les histogrammes de couleur, les motifs binaires locaux (LBP) et les histogrammes de gradient orienté (HOG). Cependant, nous avons noté que ces techniques étaient parfois insuffisantes pour le milieu aquatique. En particulier, nous avons été confrontés à plusieurs difficultés, à savoir, l'orientation, la déformation de la tête et les occultations causées par les éclaboussures et les passages de la main. Afin de pallier ces problèmes, nous avons proposé des approches optimisées basées sur la fusion de données. En particulier, nous avons commencé par proposer un système multipiste qui est constitué de plusieurs pistes de suivi, où chacune représente l'une des techniques de suivi citées précédemment. Ensuite, parmi les décisions offertes par chaque piste, nous choisissons la meilleure selon un critère basé sur l'histogramme de couleur. Cette approche a significativement amélioré les résultats du suivi par rapport aux techniques de base, mais elle reste insuffisante pour l'étude des performances des nageurs.

Afin d'améliorer le suivi, nous avons proposé et validé une nouvelle approche nommée fusion dynamique qui consiste à fusionner le plan de corrélation (NL-JTC) et le plan des scores couleurs (histogramme de couleur) dans le but de générer une description riche de la cible à suivre (forme + couleur). Ce choix a été motivé par notre constat d'une certaine complémentarité entre ces deux techniques. En effet, la technique NL-JTC a montré une bonne précision pour la localisation mais reste sensible par rapport à la déformation des contours, tandis que l'histogramme de couleur a montré un suivi robuste mais moins précis, en termes de localisation. Avant cette fusion, nous avons proposé plusieurs optimisations pour les différentes techniques utilisées afin d'améliorer leur robustesse pour le suivi des nageurs et leur temps de calcul. D'autre part, nous avons adapté ces approches pour les préparer à la fusion. En particulier, nous avons proposé une méthode d'extraction de la zone potentielle de corrélation, la normalisation des deux approches par la méthode Min-Max et enfin leur fusion utilisant un critère de fusion pondérée. Notre approche de fusion dynamique a nettement optimisé les résultats de suivi mis à part dans les cas d'occultations quasi-totales et totales.

Afin de remédier à ce problème, nous avons proposée l'approche multizone. L'idée principale de celle-ci est de suivre, en même temps, deux parties du corps du nageur, à savoir la tête et le maillot de bain. Pour chaque zone, l'approche de fusion dynamique est appliquée mais notre décision ne se limite pas au pic du plan de fusion. En effet, nous avons proposé une technique pour la localisation de toutes les détections potentielles en effectuant une recherche des maximums locaux, dans le plan de fusion de chacune des deux zones. Puis, pour chaque couplet tête - maillot de bain, nous calculons la valeur d'un critère complexe composé des sous-critères suivants : facteur de confiance de chaque zone, distance euclidienne entre les deux zones ainsi que la vitesse de l'ensemble. Enfin, nous choisissons le couplet qui maximise la valeur de ce critère. Cette approche a été testée sur plusieurs séquences vidéos et a été validée et retenue pour le suivi des nageurs

et l'étude de leurs performances.

La dernière partie de la thèse est consacrée à l'étude préliminaire des performances des nageurs. Cette dernière est principalement basée sur l'analyse des courbes de vitesse cyclique, intra-cyclique et instantanée. Pour ce faire, nous nous sommes basés sur les positions des nageurs issues du suivi automatique. Les résultats obtenus, pour la nage libre et le dos, concordent avec la vérité terrain et permettent donc une utilisation de notre système en condition réelle. En particulier, nous obtenons une erreur moyenne de 0.1 m/s pour l'estimation de la vitesse instantanée. En ce qui concerne la brasse et le papillon, les résultats du suivi automatique ont montré quelques imprécisions causées par le mouvement important du corps du nageur et particulièrement par la distance entre sa tête et le plan d'eau.

Perspectives de recherche

Nous avons développé dans cette thèse un système automatique de suivi des nageurs permettant d'étudier leurs performances. Certaines limitations existent cependant dans notre système, notamment pour la brasse et le papillon. En effet, ces deux types de nage ont la particularité d'avoir de fortes amplitudes sur le mouvement de la tête (50 cm entre la tête hors et dans l'eau). Ce déplacement de la tête génère des imprécisions au niveau du suivi, liées principalement au passage des coordonnées pixels aux métriques engendré par la méthode de calibration choisie. En effet, celle-ci permet de calibrer le bassin selon uniquement deux dimensions qui négligent la position de la tête par rapport au plan d'eau falsifiant de ce fait les résultats de localisation.

Afin de pallier ce problème, il pourrait être intéressant d'étudier la possibilité d'appliquer la méthode de calibrage DLT dédiée aux plans $3D$ comme dans les travaux de Elipot et al. dans [24]. Ceux-ci consistent à utiliser deux caméras calibrées qui filment la même scène. En se basant sur le principe de stéréo-vision, il serait possible de déterminer avec précision les trois coordonnées de la tête afin d'optimiser la localisation de la tête du nageur pour la brasse et le papillon. Cependant, cette amélioration nécessite l'emploi d'un autre système de prise de vue pour la stéréo-vision ainsi qu'un suivi optimisé permettant de suivre encore plus précisément un point clé présent sur le nageur.

D'autre part, il aurait pu être intéressant d'approfondir les tests dans le but de développer une méthode de calibrage entièrement automatique. Dans notre système d'étude de performances basé sur le suivi, la seule étape effectuée manuellement est la sélection des quatre points clés du calibrage. Ces derniers correspondent aux changements de couleurs des lignes délimitant les couloirs. Etant donnée que les dimensions et les couleurs de ces lignes sont connues, il serait possible de développer une approche basée sur la détection de lignes et de couleurs pour effectuer cette étape. Ainsi le calibrage deviendrait automatique rendant de ce fait notre système d'étude des

performances tout-automatique.

Une autre piste intéressante pourrait être envisagée dans le but d'optimiser la précision de l'approche de suivi des nageurs. Il s'agit ici d'utiliser un système de prise de vue composé de plusieurs caméras calibrées et suspendues au dessus du bassin. Par exemple, pour un bassin de 50 mètres 10 caméras pourraient être utilisées afin de couvrir tout le bassin. Celui-ci serait alors divisé en 10 zones de 12 x 12 mètres (avec 2 mètres de chevauchement). Dans le but d'avoir une vision optimale, chaque caméra devrait être placée et fixée au centre de chaque zone à une hauteur permettant, avec une focale donnée, de filmer intégralement celle-ci. Les traitements pourraient être effectués séparément sur les séquences acquises des caméras concernées par le couloir de nage étudié. En effet, ce système offrirait de ce fait une vue du dessus du bassin d'une très haute résolution (10x la résolution d'une seule caméra). Ainsi, il serait possible de fortement limiter les occultations et d'avoir une meilleure visibilité du nageur permettant un suivi plus performant. Ce système pourrait être testé et validé sur des piscines d'entraînement pour ensuite être généralisé pour les piscines accueillant des événements de haut-niveau tels que les championnats nationaux et internationaux.

Dans d'autres contextes applicatifs, les différents modules du système développé pourraient être adaptés à d'autres applications de suivi, notamment les applications sportives : football, handball, surf, etc. Par exemple, dans le cas d'un match de football, nous pourrions mettre en place notre système de prise de vue $8K$ où chaque caméra couvrirait la moitié du stade. Ce système pourrait être étendu à une meilleure résolution avec l'utilisation d'un nombre plus important de caméras. Dans le même esprit que pour la natation, le terrain serait calibré afin de pouvoir effectuer des mesures de distance. Ensuite, les méthodes de détermination de la région d'intérêt et de fusion dynamique pourraient être utilisées afin de suivre les joueurs dans le stade. Ceci permettrait de mesurer les déplacements et la vitesse des joueurs pour évaluer la tactique adoptée. Cette étude pourrait se baser sur l'analyse du positionnement de chaque joueur ainsi que sur sa trajectoire.

Annexe A

Prototype d'évaluation des nageurs basé sur l'annotation manuelle

A.1 Introduction

Ce guide a pour but de présenter le prototype développé au sein du laboratoire Vision de l'ISEN Brest et l'entreprise Actris. Celui-ci est dédié au calibrage du bassin, l'annotation manuelle de nageurs dans une séquence vidéo en cliquant manuellement sur la position du nageur. Il est possible d'annoter chaque image de la vidéo ou simplement des cycles bien définis. Grâce à ces annotations il est possible de générer des courbes de vitesses. Celles-ci donnent une estimation précise de la vitesse instantanée ou cyclique qui pourra être fournies aux entraîneurs pour analyser les performances des nageurs.

A.2 Présentation de l'interface

L'interface principale de l'outil développé pour l'annotation est conçue de telle façon qu'elle soit simple et facile à utilisée. Celle-ci est illustrée dans la figure A.1 :

L'utilisation du prototype se fait à travers les taches suivantes :

1. Sélectionner la vidéo à analyser.
2. Calibrer la vidéo en cliquant sur 4 points dont on connaît les coordonnées métriques (figure A.2).
3. Annoter la position des nageurs (figure A.3).
4. Générer la courbe de vitesse et analyser les performances (figure A.4).



FIGURE A.1 – Interface graphique Matlab pour le calibrage, l'annotation des nageurs, le calcul de la vitesse et l'analyse des performances.

5. Aller au répertoire des vidéos d'analyse enregistrées.
6. Fermer l'interface.

A.3 Calibrage

Afin de calibrer le bassin, il faut cliquer sur le bouton "Calibrer". Ensuite, il est nécessaire de sélectionner 4 points (si possible assez éloignés les uns des autres dans le bassin), et de saisir leurs coordonnées métriques. Il à noter qu'il est possible de zoomer sur l'image pour avoir plus de précision. Enfin, une grille (carrés de taille 2.5x2.5 mètres) est affichée pour permettre la vérification du résultat de la calibration (figure A.2).

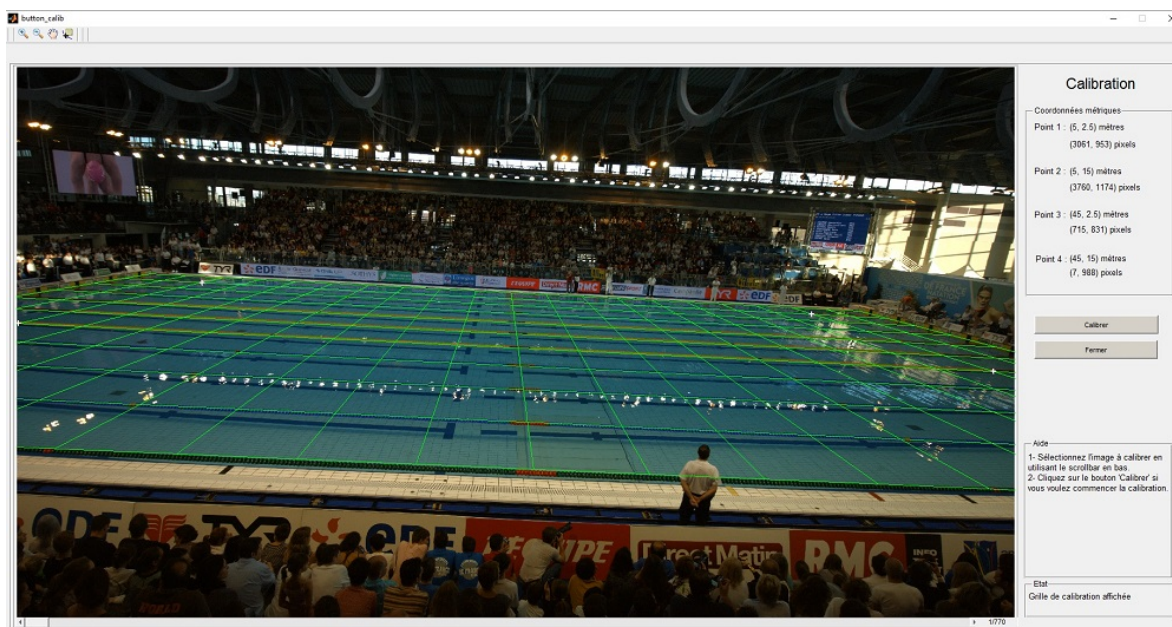


FIGURE A.2 – Sous-interface graphique dédiée au calibrage et la vérification de celle-ci.

A.4 Annotation

Le processus d'annotation est le suivant (figure A.3) :

1. Saisir le nom du nageur.
2. Avant de commencer l'annotation l'utilisateur est amené à parcourir la vidéo en utilisant les flèches de l'ascenseur ("scroll bar") en-dessous de la vidéo afin de choisir l'image de départ.

ANNEXE A. PROTOTYPE D'ÉVALUATION DES NAGEURS BASÉ SUR L'ANNOTATION MANUELLE

3. Lors de la première annotation, une droite de prédiction de la trajectoire est calculée. Celle-ci permet, comme le montre la figure 3, d'aider à l'annotation du reste de la vidéo. Attention, si le nageur s'éloigne de cette droite au cours de la course, il peut être utile de projeter la position du nageur sur cette droite afin de limiter les erreurs d'annotations (dans l'idéal, il faudrait que chaque annotation se trouve sur cette droite). Dans la figure 3, l'image du dessus correspond à la zone d'annotation zoomée tandis que celle du dessous donne une vision globale de l'avancement du processus d'annotation.
4. Les boutons flèches droite et gauche permettent d'avancer d'une seule image, tandis que les flèches haut et bas permettent d'avancer d'un pas (à choisir dans l'interface).
5. Le bouton "retour en arrière" est utilisé pour annuler une annotation.
6. Le bouton "r" permet d'indiquer que le nageur est arrivé au bout du bassin.
7. Le bouton "échappe" sert à indiquer la fin de l'annotation.
8. Une fois l'annotation terminée, les positions annotées seront sauvegardées afin de pouvoir les utiliser durant la prochaine étape : l'analyse.

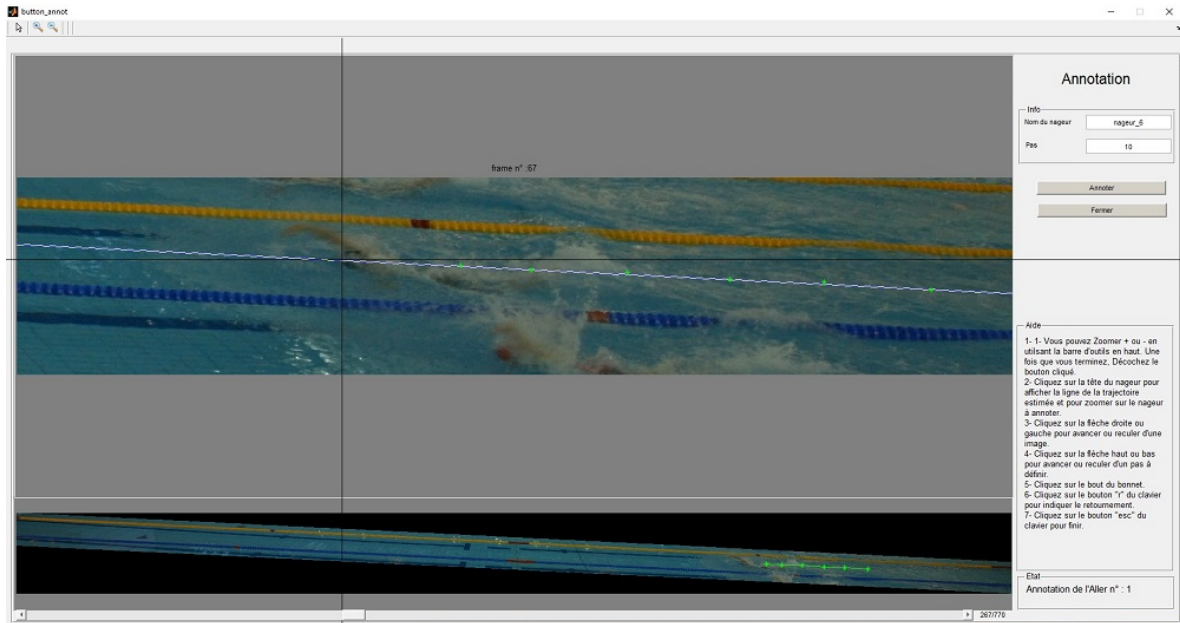


FIGURE A.3 – Sous-interface graphique dédiée à l'annotation de la position du nageur.

A.5 Analyse

La figure A.4 présente la sous interface graphique dédiée à l'analyse. Cette dernière fonctionne de la manière suivante :

1. Différentes fonctionnalités sont mises à disposition de l'utilisateur afin de faciliter l'analyse :
 - (a) Aller au cycle précis choisi par l'utilisateur.
 - (b) Avancer au reculer d'un cycle.
 - (c) Parcourir la vidéo d'analyse.
2. Pour exporter la vidéo avec l'analyse des performances il est nécessaire de cliquer sur le bouton "Génération d'une vidéo d'analyse". La vidéo ainsi exportée contient à chaque instant ($1/FPS$) l'image zoomée du nageur, la courbe indiquant la vitesse instantanée, le cycle et la distance parcourue.
3. Afin de consulter les vidéos d'analyse, il suffit de cliquer sur le bouton "Aller aux vidéos".

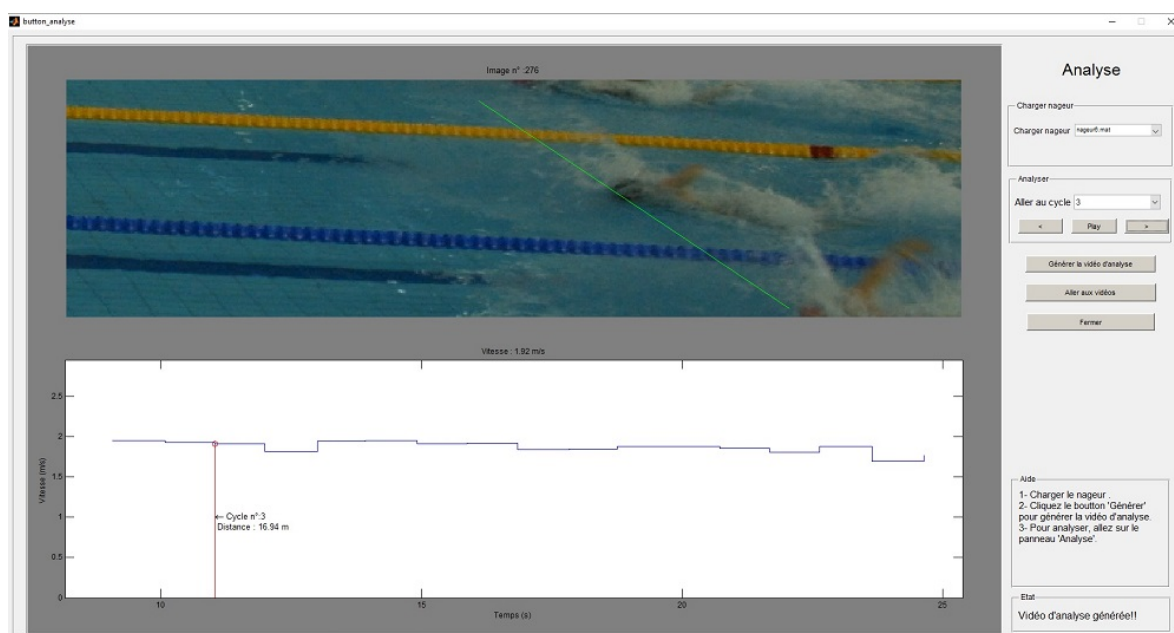


FIGURE A.4 – Sous-interface graphique dédiée à l'analyse d'une courbe de vitesse d'un nageur annoté.

Annexe B

Prototype d'évaluation des nageurs basé sur le suivi automatique

B.1 Introduction

Dans le but d'évaluer les approches proposées pour le suivi automatique, nous avons développé une interface graphique basé sur Matlab R2012b. Cette dernière permet d'effectuer toutes les étapes constituant notre système de suivi et d'étude des performances des nageurs. Comme le montre la figure B.1, notre interface graphique se compose de quatre parties assurant les différentes fonctionnalités : (1) acquisition, (2) suivi automatique, (3) statistiques et (4) performances. Il est à noter que cette interface a été simplifiée dans le but de faciliter son utilisation pour les utilisateurs.

B.2 Acquisition (*Video controller*)

Comme le montre la figure B.1.1, cette partie consiste principalement à la phase de préparation de la séquence vidéo pour les traitements ultérieurs. Dans ce qui suit, nous allons présenter les différentes fonctionnalités de chaque bouton de cette partie.

Select Video : ce bouton sert à charger la séquence vidéo de la course à partir du média de sauvegarde (disque dur, clé USB, etc).

Play : il permet, avec la barre de progression en dessous de la vidéo, de parcourir la séquence afin de choisir l'image à calibrer ou l'image de lancement du suivi.

Calibrate : il sert à afficher l'image choisie sur une autre fenêtre permettant de l'agrandir et d'annoter les quatre points clés nécessaires à l'extraction des paramètres de calibrage. Il est à noter que ces paramètres sont utilisés pour le calibrage de toute la

séquence vidéo.

Crop : il sert à initialiser la référence du suivi automatiquement ou manuellement selon les tests à effectuer.

B.3 Suivi automatique (*Tracking*)

La figure B.1.2 présente la partie permettant d'appliquer les différentes approches de suivi et d'afficher le résultat en même temps. Dans ce qui suit, nous allons présenter les différentes fonctionnalités de chaque bouton de cette partie.

Track : ce bouton sert à lancer le système de suivi et d'étude des performances des nageurs. Par défaut, l'approche utilisée est la fusion dynamique. En revanche, si la *check box* de l'approche multizone est cochée alors cette dernière sera utilisée comme approche de suivi. Aussi, si la *check box* composite est cochée alors le filtre composite sera appliqué pour l'approche de corrélation *NL-JTC*.

Ground truth : ceci sert à afficher une nouvelle fenêtre dédiée au calibrage, comme présenté dans la figure A.3 de l'annexe A.

Save : il permet d'enregistrer la vidéo d'analyse contenant la courbe de la vitesse instantanée ainsi que l'image du nageur concerné à chaque instant.

B.4 Statistiques (*Statistics*)

Dans cette partie, différents paramètres et statistiques sont affichés, comme présenté dans la figure B.1.3. Ceux-ci peuvent être divisés en trois ensembles : les paramètres de la séquence vidéo, les statistiques du suivi et les statistiques des performances du nageur.

Paramètres de la séquence vidéo : Nous trouvons ici les paramètres suivants : nom de la vidéo, format, résolution, durée de la séquence et durée de son chargement.

Statistiques du suivi : pour chaque image, le temps de calcul et le score du suivi sont affichés.

Statistiques des performances du nageur : il s'agit ici de la vitesse instantanée, de la vitesse moyenne et de la distance parcourue par le nageur.

B.5 Performances (*Evaluation*)

Comme le montre la figure B.1.4, les résultats sont affichés instantanément sous forme d'une courbe de vitesse instantanée avec l'image zoomée du nageur. Le bouton de

lissage (*Filtering*) est utilisé pour filtrer la courbe selon la taille de la fenêtre gaussienne introduite dans le champs correspondant. Il est à noter que ses résultats peuvent être enregistrée dans une vidéo d'analyse en utilisant le bouton *Save* comme nous l'avons déjà cité dans la section B.3.

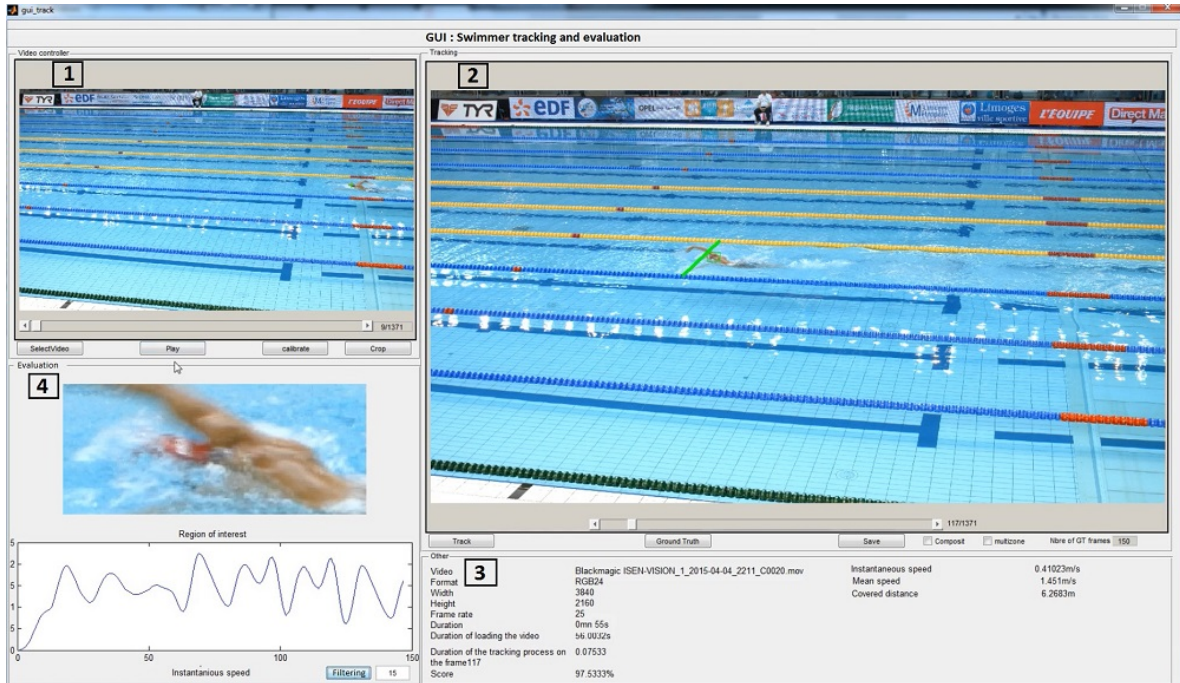


FIGURE B.1 – interface graphique Matlab dédiée au suivi et à l'étude des performances des nageurs.

Bibliographie

- [1] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard. All-automatic swimmer tracking system based on an optimized scaled composite jtc technique. volume 9845, pages 98450E–98450E–8, 2016.
- [2] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard. Optimized motion detection in calibrated videos based on a contrario priciple : application on swimmer tracking system. *en soumission*, 2016.
- [3] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard. A novel multi-tracking system for the evaluation of high-level swimmers performances. volume 9094, pages 90940A–90940A–7, 2014.
- [4] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard. Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques. *Optics Communications*, 356 :256–268, December 2015.
- [5] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard. Optimized swimmer tracking system based on a novel multi-related-targets approach. *Optics and Lasers in Engineering*, pages –, 2016.
- [6] Amit Adam, Ehud Rivlin, and Ilan Shimshoni. Robust fragments-based tracking using the integral histogram. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 798–805, Washington, DC, USA, 2006. IEEE Computer Society.
- [7] P. Katz, M Aron, and A Alfalou. Joint transform correlation for face tracking : elderly fall detection application. In *Proc. SPIE 8748*, page 87480I, 2013.
- [8] M.Z. Islam, Chi min Oh, and Chil-Woo Lee. Real time moving object tracking by particle filter. In *Computer Science and its Applications, 2008. CSA '08. International Symposium on*, pages 347–352, 2008.
- [9] Raquel Urtasun and Pascal Fua. 3d human body trackingacking using deterministic temporal motion models. Technical report, 2004.
- [10] Zhiwei Zhu, Kikuo Fujimura, and Qiang Ji. Real-time eye detection and tracking under various light conditions. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, pages 139–144, New York, NY, USA, 2002.

- [11] Zhengqiang Jiang, Du Q. Huynh, William Moran, Subhash Challa, and Nick Spadaccini. Multiple pedestrian tracking using colour and motion models. In *DICTA*, pages 328–334, 2010.
- [12] A. Alfalou, N. Ben Haj Yahia, and M. S. Alam. Face recognition using a non-zero-order correlation plane and a nonlinear joint transform correlator. In *Proc. Of SPIE 8498*, page 84981A, 2012.
- [13] I. Leonard, A. Alfalou, and M. S. Alam. Nonlinear fringe-adjusted jtc-based face tracking using an adaptive decision criterion. pages 81340M–81340M–6, 2011.
- [14] I. Leonard, A. Arnold-Bos, and A. Alfalou. Interest of correlation-based automatic target recognition in underwater optical images : theoretical justification and first results. pages 76780O–76780O–12, 2010.
- [15] A. Alfalou and C. Brosseau. *Face Recognition Book*, chapter Understanding Correlation Techniques for Face Recognition : From Basics to Applications, pages 354–380. INTECH, 2010.
- [16] Gao Junxiang, Zhou Tong, and Liu Yong. Face tracking using color histograms and moment invariants. In *Broadband Network Multimedia Technology, 2009. IC-BNMT '09. 2nd IEEE International Conference on*, pages 519–523, 2009.
- [17] Wenmiao Lu and Yap-Peng Tan. A color histogram based people tracking system. In *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, pages 137–140. IEEE, 2001.
- [18] Ming-Yi Ju, Chen-Sen Ouyang, and Hao-Shiu Chang. Mean shift tracking using fuzzy color histogram. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, pages 2904–2908, 2010.
- [19] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns : Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, (12) :2037–2041, 2006.
- [20] Yuchun Fang, Jie Luo, and Chengsheng Lou. Fusion of multi-directional rotation invariant uniform lbp features for face recognition. In *Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on*, pages 332–335, 2009.
- [21] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pages 886–893, 2005.
- [22] Stan Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 232–237, 1998.
- [23] Wang Chuan-xu and Li Zuo-yong. A new face tracking algorithm based on local binary pattern and skin color information. In *Computer Science and Computational Technology, 2008. ISCSCT '08. International Symposium on*, pages 657–660, 2008.

- [24] Marc Elipot. *Application of new techniques of camera calibration and the cinematic analysis of swimming*. PhD thesis, UNIVERSITE PARIS DESCARTES, 2010.
- [25] Y. I. AbdelAziz and H. M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Proc. of the Symposium on Close-Range Photogrammetry*, pages 1–18, Urbana, Illinois, 1971.
- [26] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 666–673 vol.1, 1999.
- [27] N. Kanopoulos, N. Vasanthavada, and R. L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-State Circuits*, (2) :358–367, 1988.
- [28] M. Piccardi. Background subtraction techniques : a review. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, pages 3099–3104, 2004.
- [29] Erik Pogalin, Andrew H. C. Thean, Jan Baan, Nellie Schipper, and Arnold W. M. Smeulders. Video-based training registration for swimmers. *Int. J. Comp. Sci. Sport*, (1), 2007.
- [30] Andrews Sobral and Antoine Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122 :4–21, 2014.
- [31] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Review and evaluation of commonly-implemented background subtraction algorithms. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec 2008.
- [32] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, volume 2, page 252 Vol. 2, 1999.
- [33] Eli Saber and A. Murat Tekalp. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. *Pattern Recognition Letters*, 1998.
- [34] M.J. Jones and J.M. Rehg. Statistical color models with application to skin detection. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, pages –280, 1999.
- [35] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, (1) :1–38, 1977.
- [36] B.D. Zarit, B.J. Super, and F.K.H. Quek. Comparison of five color models in skin pixel classification. In *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on*, pages 58–63, 1999.

- [37] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, pages 11–32, 1991.
- [38] J. Kovac, P. Peer, and F. Solina. Human skin color clustering for face detection. In *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, volume 2, pages 144–148 vol.2, Sept 2003.
- [39] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd : A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4) :722–732, 2010.
- [40] Yun-Seok Lee, Han-Suh Koo, and Chang-Sung Jeong. A straight line detection using principal component analysis. (14) :1744–1754, 2006.
- [41] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, (1) :11–15, 1972.
- [42] C. Schmid and A. Zisserman. Automatic line matching across views. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 666, Washington, DC, USA, 1997.
- [43] Mohamed HAMMAMI. *Modèle ded peau et application á la classification d'images et au filtrage des sites Web*. PhD thesis, ECOLE CENTRALE DE LYON, 2005.
- [44] V. Govindaraju, D.B. Sher, R.K. Srihari, and S.N. Srihari. Locating human faces in newspaper photographs. In *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on*, pages 549–554, 1989.
- [45] J B Waite, W J Welsh, Martlesham Heath, and Ipswich Ip Re. An application of active contour models to head boundary location. In *In Proceedings of the British Machine Vision Conference*, pages 407–412, 1990.
- [46] Ahmed M. Elgammal, David Harwood, and Larry S. Davis. Non-parametric model for background subtraction. In *Proceedings of the 6th European Conference on Computer Vision-Part II*, pages 751–767, London, UK, UK, 2000.
- [47] P. W. Power and J. A. Schoonees. Understanding background mixture models for foreground segmentation. In *Proceedings Image and Vision Computing New Zealand*, page 267, 2002.
- [48] Muhammad Sarfraz. *Computer-aided intelligent recognition techniques and applications*. Wiley Online Library, 2005.
- [49] R. M. Haralick, S. R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4) :532–550, July 1987.
- [50] Abdol-Reza Mansouri, Alfred S Malowany, and Martin D Levine. Line detection in digital pictures : A hypothesis prediction/verification paradigm. *Computer Vision, Graphics, and Image Processing*, 40(1) :95–114, 1987.
- [51] J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6) :679–698, June 1986.

- [52] D.S Guru, B.H Shekar, and P Nagabhushan. A simple and robust line detection algorithm based on small eigenvalue analysis. *Pattern Recognition Letters*, 25(1) :1 – 13, 2004.
- [53] Azriel Rosenfeld. Picture processing by computer. *ACM Comput. Surv.*, 1(3) :147–176, September 1969.
- [54] Paul VC Hough. Method and means for recognizing complex patterns. Technical report, 1962.
- [55] Bahram Javidi, Jun Wang, and Qing Tang. Multiple-object binary joint transform correlation using multiple-level threshold crossing. *Appl. Opt.*, 30(29) :4234–4244, Oct 1991.
- [56] M.S. Alam and C. N. Wai. Color pattern recognition using fringe-adjusted joint transform correlation. In *National Aerospace and Electronics Conference, Proceedings of the IEEE 2000*, pages 427–433, 2000.
- [57] Bahram Javidi. Nonlinear joint power spectrum based optical correlation. *Appl. Opt.*, (12) :2358–2367, 1989.
- [58] S. Jenicka and A. Suruliandi. Empirical evaluation of distance measures for supervised classification of remotely sensed image with modified multivariate local binary pattern. In *Emerging Trends in Electrical and Computer Technology (ICE-TECT), 2011 International Conference on*, pages 762–767, 2011.
- [59] S.T. Birchfield and Sriram Rangarajan. Spatiograms versus histograms for region-based tracking. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pages 1158–1163, 2005.
- [60] N. Morizet. *Reconnaissance Biométrique par Fusion Multimodale du Visage et de l Iris*. PhD thesis, Télécom ParisTech, 2009.
- [61] A. Ross and A. K. Jain. Multimodal biometrics : an overview. pages 1221–1224, 2004.
- [62] M. Faundez-Zanuy. Data fusion in biometrics. *IEEE Aerospace and Electronic Systems Magazine*, 20(1) :34–38, Jan 2005.
- [63] Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA, 1995.
- [64] M. Marron, J.C. Garcia, M. A. Sotelo, M. Cabello, D. Pizarro, F. Huerta, and J. Cerro. Comparing a kalman filter and a particle filter in a multiple objects tracking application. In *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, pages 1–6, 2007.
- [65] Zhe Chen. Bayesian filtering : From kalman filters to particle filters, and beyond. Technical report, McMaster University, 2003.
- [66] Elise Arnaud, Etienne Memin, and Bruno Cernuschi-Frias. Conditional filters for image sequence based tracking - application to point tracking. *IEEE Transactions on Image Processing*, (1), 2005.

- [67] P. Radecki and B. Hancey. Online building thermal parameter estimation via unscented kalman filtering. In *American Control Conference (ACC), Montreal, CA*, pages 3056–3062, 2012.
- [68] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, pages 107–113, 1993.
- [69] Paul A. Brasnett, Lyudmila Mihaylova, Nishan Canagarajah, and David Bull. Particle filtering with multiple cues for object tracking in video sequences. *Proceedings of SPIE*, pages 430+, 2005.
- [70] P. Brasnett, L. Mihaylova, N. Canagarajah, and D. Bull. Particle filtering with multiple cues for object tracking in video sequences. In *SPIE Proceedings*, pages 430–441. SPIE, 2005.
- [71] M. Sanjeev Arulampalam, Simon Maskell, and Neil Gordon. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, pages 174–188, 2002.
- [72] Mathias Fontmarty. *Vision et filtrage particulière pour le suivi tridimensionnel de mouvement humain. Applications a la Robotique*. PhD thesis, 2009.
- [73] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, (5) :564–575, 2003.
- [74] Dorin Comaniciu, Peter Meer, and Senior Member. Mean shift : A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 603–619, 2002.
- [75] Gael Jaffre and Alain Crouzil. Non-rigid object localization from color model using mean shift (regular paper). In *International Conference on Image Processing (ICIP), Barcelone, Espagne*, pages 317–320, 2003.
- [76] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, pages 137–154, 2004.
- [77] D.R. Magee. Tracking multiple vehicles using foreground, background and motion models. *Image and Vision Computing*, pages 143–155, 2001.
- [78] S. Gupte, O. Masoud, R.F.K. Martin, and N.P. Papanikolopoulos. Detection and classification of vehicles. *Intelligent Transportation Systems, IEEE Transactions on*, (1) :37–47, 2002.
- [79] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An hog-lbp human detector with partial occlusion handling. In *ICCV*, pages 32–39, 2009.
- [80] Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Rapid and robust human detection and tracking based on omega-shape features. In *ICIP*, pages 2545–2548. IEEE, 2009.
- [81] Yafeng Yin, Hong Man, Sachi Desai, and Haibo He. Multiple human detection and tracking by using multiple-stage hog detector and pfgpdm. pages 74800S–74800S–10, 2009.

- [82] Sang-Jean Lee, Sang-Bong Jung, Jang-Woo Kwon, and Seung-Hong Hong. Face detection and recognition using pca. In *TENCON 99. Proceedings of the IEEE Region 10 Conference*, pages 84–87, 1999.
- [83] P. Vadakkepat, P. Lim, L.c. De Silva, Liu Jing, and Li Li Ling. Multimodal approach to human-face detection and tracking. *Industrial Electronics, IEEE Transactions on*, (3) :1385–1393, 2008.
- [84] S. Spors and R. Rabenstein. A real-time face tracker for color video. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, pages 1493–1496, 2001.
- [85] S. Bhat S. David P. Kuchi, P. Gabbur. Human face detection and tracking using skin color modeling and connected component operators. *IETE J. Res. Special Issue on Visual Media Processing*, 2012.
- [86] D.J. Dailey, F.W. Cathey, and S. Pumrin. An algorithm to estimate mean traffic speed using uncalibrated cameras. *Intelligent Transportation Systems, IEEE Transactions on*, (2) :98–107, 2000.
- [87] Yonghong Long, Xiyu Xiao, Xiaohua Shu, and Shenglan Chen. Vehicle tracking method using background subtraction and meanshift algorithm. In *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on*, pages 1–4, 2010.
- [88] S. Avidan. Support vector tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (8) :1064–1072, 2004.
- [89] Andrew Miller, Arslan Basharat, Brandyn White, Jingen Liu, and Mubarak Shah. Person and vehicle tracking in surveillance video. In *Multimodal Technologies for Perception of Humans*, pages 174–178. Springer Berlin Heidelberg, 2008.
- [90] Wenshuo Gao, Xiaoguang Zhang, Lei Yang, and Huizhong Liu. An improved sobel edge detection. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, pages 67–71, 2010.
- [91] F Puel, J Morlier, M Avalos, M Mesnard, M Cid, and P Hellard. 3d kinematic and dynamic analysis of the front crawl tumble turn in elite male swimmers. *Journal of biomechanics*, (3) :510–515, 2012.
- [92] F Puel, J Morlier, M Avalos, M Mesnard, M Cid, and P Hellard. 3d kinematic and dynamic analysis of the front crawl tumble turn in elite male swimmers. *Journal of biomechanics*, (3) :510–515, 2012.
- [93] eyeviewsystems.
- [94] FINA. Fédération internationale de natation (fina), 2016.
- [95] S. Oprisescu, M. Ciuc, and V. Buzuloiu. Histogram and motion based intrusion detection and tracking algorithms for tof cameras. In *Signals, Circuits and Systems, 2009. ISSCS 2009. International Symposium on*, pages 1–4, 2009.
- [96] David Lowe. Visual recognition as probabilistic inference from spatial relations. *AI and the Eye*, 1990.

- [97] Agns Desolneux, Lionel Moisan, and Jean-Michel Morel. *From Gestalt Theory to Image Analysis : A Probabilistic Approach*. Springer Publishing Company, Incorporated, 1st edition, 2007.
- [98] Isabelle Abraham, Romain Abraham, Agnes Desolneux, and Sebastien Li-Thiao-Te. Significant edges in the case of a non-stationary Gaussian noise. *Pattern Recognition*, 40 :3277–3291, 2007.
- [99] A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by Helmholtz principle. 14(3) :271–284, 2001.
- [100] Pablo Musé, Frédéric Sur, Frédéric Cao, Yann Gousseau, and Jean-Michel Morel. An a contrario decision method for shape element recognition. *International Journal of Computer Vision*, 69(3) :295–315, 2006.
- [101] Lionel Moisan and Bérenger Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57(3) :201–218, 2004.
- [102] M. Sabet, R.A. Zoroofi, K.S. Niat, and M. Sabbaghian. Automated face tracking with self correction capability. In *Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference of*, pages 280–284, 2011.
- [103] J. L. de Bougrenet de la Tocnaye, E. Quémener, and Y. Pétillot. Composite versus multichannel binary phase-only filtering. *Appl. Opt.*, 36(26) :6646–6653, Sep 1997.
- [104] B. V. K. Vijaya Kumar. Tutorial survey of composite filter designs for optical correlators. *Appl. Opt.*, 31(23) :4773–4801, Aug 1992.
- [105] A. Alfalou, C. Brosseau, and W. Kaddah. Optimization of decision making for face recognition based on nonlinear correlation plane. *Optics Communications*, pages 22 – 27, 2015.
- [106] R. Arellano, P. Brown, J. Cappaert, and R. Nelson. Analysis of 50m, 100m 200m freestyle swimmers at the 1992 Olympic Games. *Journal of Applied Biomechanics*, 10 :189–199, 1994.
- [107] L. Seifert, L. Boulesteix, M. Carter, and D. Chollet. The spatial-temporal and coordinative structures in elite male 100-m front crawl swimmers. *Int J Sports Med*, 26 :286–293, 2005.
- [108] P. Hellard, J. Dekerle, M. Avalos, N. Caudal, M. Knopp, and C. Hausswirth. Kinematic measures and stroke rate variability in elite female 200-m swimmers in the four swimming techniques : Athens 2004 olympic semi-finalists and french national 2004 championship semi-finalists. *Journal of Sports Sciences*, 26 :35–46, 2008.
- [109] Kevin Thompson. *Pacing : Individual strategies for optimal performance*. Human Kinetics, 2014.
- [110] M. Elipot, P. Hellard, R. Taar, E. Boissire, J. L. Rey, S. Lecat, and N. Houel. Analysis of swimmers’ velocity during the underwater gliding motion following grab start. *Journal of Biomechanics*, 42(9) :1367–1370, June 2009.

- [111] FFN. Federation francaise de natation (ffn), 2016.

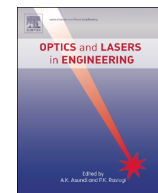
Publications

Articles de journaux

1. D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. hellard. *Optimized motion detection in calibrated videos based on a contrario priciple : application on swimmer tracking system*. En soumission.
2. D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. hellard. *Optimized swimmer tracking system based on a novel multi-related-targets approach*. *Optics and Lasers in Engineering*, 2016.
3. D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. hellard. *Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques*. *Optics Communications*, 356, 256-268, 2015.

Articles de conférences internationales

1. D. Benarab, T. Napoléon, A. Alfalou, A. Verney and P. Hellard. *All-automatic swimmer tracking system based on an optimized scaled composite JTC technique*. In *Proceedings of SPIE, Optical Pattern recognition XXVII*, 9845, 98450E-98450E-8, 2016 .
2. D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard, *A novel multi-tracking system for the evaluation of high-level swimmers performances*. *Proceedings of SPIE, Optical Pattern recognition XXV*, 9094, 90940A-90940A7, 2014.



Optimized swimmer tracking system based on a novel multi-related-targets approach

D. Benarab^{a,b}, T. Napoléon^a, A. Alfalou^{a,*}, A. Verney^b, P. Hellard^c

^a ISEN Brest, Département Vision, L@bISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France

^b ACTRIS Brest, 24 Rue Victor Grignard, Guipavas – BP 143, 29803 Brest Cedex 9, France

^c FFN (Fédération Française de Natation), 14 Rue Scandicci, 93508 Pantin Cedex, France

ARTICLE INFO

Article history:

Received 2 February 2016

Received in revised form

9 May 2016

Accepted 17 May 2016

Keywords:

Tracking system

Swimmer tracking

Dynamic fusion

Correlation

NL-JTC

Color histogram

Multi-related-targets

ABSTRACT

Robust tracking is a crucial step in automatic swimmer evaluation from video sequences. We designed a robust swimmer tracking system using a new multi-related-targets approach. The main idea is to consider the swimmer as a bloc of connected subtargets that advance at the same speed. If one of the subtargets is partially or totally occluded, it can be localized by knowing the position of the others. In this paper, we first introduce the two-dimensional direct linear transformation technique that we used to calibrate the videos. Then, we present the classical tracking approach based on dynamic fusion. Next, we highlight the main contribution of our work, which is the multi-related-targets tracking approach. This approach, the classical head-only approach and the ground truth are then compared, through testing on a database of high-level swimmers in training, national and international competitions (French National Championships, Limoges 2015, and World Championships, Kazan 2015). Tracking percentage and the accuracy of the instantaneous speed are evaluated and the findings show that our new approach is significantly more accurate than the classical approach.

© 2016 Published by Elsevier Ltd.

1. Introduction

Swimming is a very popular sport and the competition between national teams can be fierce. Unsurprisingly, swimming performance is much investigated [1–4] during training sessions and competitions, mainly through statistical and kinematic studies (study of the movements regardless of the causes that produce them). The current systems evaluate swimmers in terms of speed, breathing cycles, diving, turning and swimming techniques, etc. These data are used by coaches to assist the swimmers in analyzing their swimming strategies in order to improve their performance.

Most of these systems acquire data from sensors that the swimmers wear. These sensors provide accurate biomechanical data, but their use is prohibited in competition. Moreover, even in training sessions, wearing the sensors restricts movement. Other systems are based on the manual annotation of the swimmer's position in each video frame in order to estimate various measures. This type of systems yields accurate analysis without imposing constraints on the swimmers, but annotation is both time-consuming and labor-intensive. We thus sought to design an automatic swimmer tracking system with minimal user intervention and no physical markers.

For accurate analysis of swimmer performance, we first sought to optimize the tracking module. In the literature, the object tracking systems have been based on such detection techniques as the Non-Linear Joint Transform Correlator (NL-JTC) [5,6,19], color histograms [7,8], Local Binary Patterns (LBP) [9,10] or Histograms of Oriented Gradients (HOG) [11,12]. Starting from the principle that each of these techniques has advantages and limitations, Benarab et al. proposed a swimmer multi-tracking system [13] that combines several tracking approaches. The decision is made after several iterations and the best detected target is chosen according to a similarity criterion based on the color histogram and the history of detections.

This approach improved tracking performance compared with the mono-tracking approaches that use only one descriptor to characterize the target. However, the decision criterion is not optimal because it favors color information. In order to improve this criterion and better merge the data, Benarab et al. then proposed a dynamic fusion approach [3] motivated by the complementarity between the NL-JTC and color histograms technique. This approach significantly enhanced the tracking results in the case of complete or partial visibility of the target. Nevertheless, tracking is impossible in the case of total occlusion.

In order to solve this issue, we propose a multi-related-targets approach that tracks two targets in parallel (head and swimsuit). Among other things, this approach takes into account the distance

* Corresponding author.

between the two targets as well as the average swimming speed. This approach has two main advantages: it improves tracking accuracy and makes it possible to find even an entirely occluded target based on the other target when this latter is visible. In addition to our multi-related-targets approach, we propose to calibrate the video in order to optimize the prediction of the swimming trajectory. Last, we estimate the swimmer's instantaneous speed in order to study the performance.

In this paper, we start by a general overview of tracking techniques in Section 2. We then detail the steps of the swimmer tracking system based on the dynamic fusion technique in Section 3. In Section 4, we introduce the concept of the multi-related-targets approach and its complex decision criterion. Finally, in Section 5, we present the experimental protocol and the results in the form of a comparative study.

2. Methodology of tracking

2.1. JTC-based techniques

The joint transform correlator (JTC) is a correlation-based technique that consists of comparing two images in order to detect, localize or identify the reference in a target image. This method has proven its efficiency by the numerous publications in the literature [5,6,14]. Most often, this originally all-optical method is implemented numerically for reasons of simplicity and portability.

Different JTC variants have been proposed, the most common being the NL-JTC technique, which is based on a non-linear thresholding function applied to the joint spectrum. This allows for greater control of the accuracy and robustness of the technique by varying the sharpness of the correlation peaks. However, this technique is sensitive to contour deformation, blur and noise.

2.2. Histogram-based techniques

Histogram-based methods are designed to numerically describe the color, texture or contour characteristics of the target to be tracked. To do so, they encode the occurrence of particular information in the image in each component of a histogram. It should be noted that a single histogram can be calculated on the whole image, in which case we speak of global description, but it is also possible to calculate several histograms, one for each region of the image, and in this case we speak of local description. The comparison between the histograms corresponding to the similarity between the reference image and the target image is obtained through a calculation of distance (Bhattacharya [10], Chi squared [15], etc.).

The well-known techniques as reported in the literature are color histograms, HOG and local LBP. The essential notion behind the color histogram technique is that the target appearance can be characterized by the distribution of its color. This technique is robust with respect to the orientation and deformation of the object, but it is sensitive to changes in lighting conditions and the confusion between same-color objects. As for the HOG descriptor, an image can be described by the distribution of intensity gradients or edge directions. The LBP technique describes the target texture by calculating the relative gray level of the neighboring pixels. These latter two techniques are robust with respect to lighting change. However, in the case of swimmer tracking, the deformation of the target due to splashing will strongly affect the contours and make these methods less discriminative.

2.3. Dynamic fusion-based technique

The above-mentioned methods for automatic swimmer tracking are often insufficient due to the particularities of the swimming environment. In order to develop a robust system, Benarab et al. [3] proposed to fuse the NL-JTC and color histogram techniques, as they showed a high degree of complementarity. Indeed, the NL-JTC is accurate for localization but it is sensitive to contour deformation, while the color histogram approach is imprecise for localization but robust regarding the deformations of contour.

Therefore, the dynamic fusion technique was used in the design of a tracking system with accurate localization and robustness with respect to contour deformation. The fusion provides a richer description of the target by combining contour and color characteristics. This technique showed improved performance compared with the techniques mentioned above. However, it remains sensitive to strong occlusions and is insufficient in these situations.

3. Dynamic fusion-based swimmer tracking system

In order to propose a robust swimmer tracking system, we set up a system based on the following successive steps: swimming pool calibration, reference selection, trajectory prediction and swimmer recognition and tracking. It should be noted that the last three steps are repeatedly applied to all of the analyzed video.

The calibration module ensures the passage from pixel coordinates to metric coordinates and vice versa. The tracking process is then initialized in the step of selecting a reference, which is the part of the swimmer to be tracked: head, swimsuit, etc. Then, the trajectory prediction module is applied based on the calibration to restrict the region of interest. Next, the images are analyzed to detect and recognize the target using the dynamic fusion technique. Last, the reference is replaced by the detected target to relaunch the next iteration of the tracking process.

3.1. Two-dimensional direct linear transformation (2D DLT) calibration

Swimmer performance analysis from a video sequence consists of measuring and studying the swimming movements. To do this, we need to calculate the spatial coordinates of the target from the pixel coordinates in the image. This passage from pixel coordinates to metric coordinates is called calibration. It relies on the prior knowledge of the camera projection parameters, which vary depending on the camera settings and positioning. For this, the DLT approach was proposed by Abdel-Aziz and Karara [16]. In the case of calibrating a swimming pool, which is a 2D space, it is sufficient to apply a simplified 2D DLT equation:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1x_1 & -v_1y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -u_nx_n & -u_ny_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -v_nx_n & -v_ny_n \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_7 \\ L_8 \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_n \\ v_n \end{bmatrix} \quad (1)$$

where u and v are the pixel coordinates of a point in the image, x and y are the real metric coordinates of the same point. $L_{1..8}$ represent the calibration parameters which are the set of the unknowns of this system of equations. In order to calculate these eight parameters and solve this system of equations, it is required to select four points whose metric and pixel coordinates are known, as shown in Fig. 1. Once we calculate $L_{1..8}$, the following equation allows the passage from pixel to metric coordinates and vice versa.

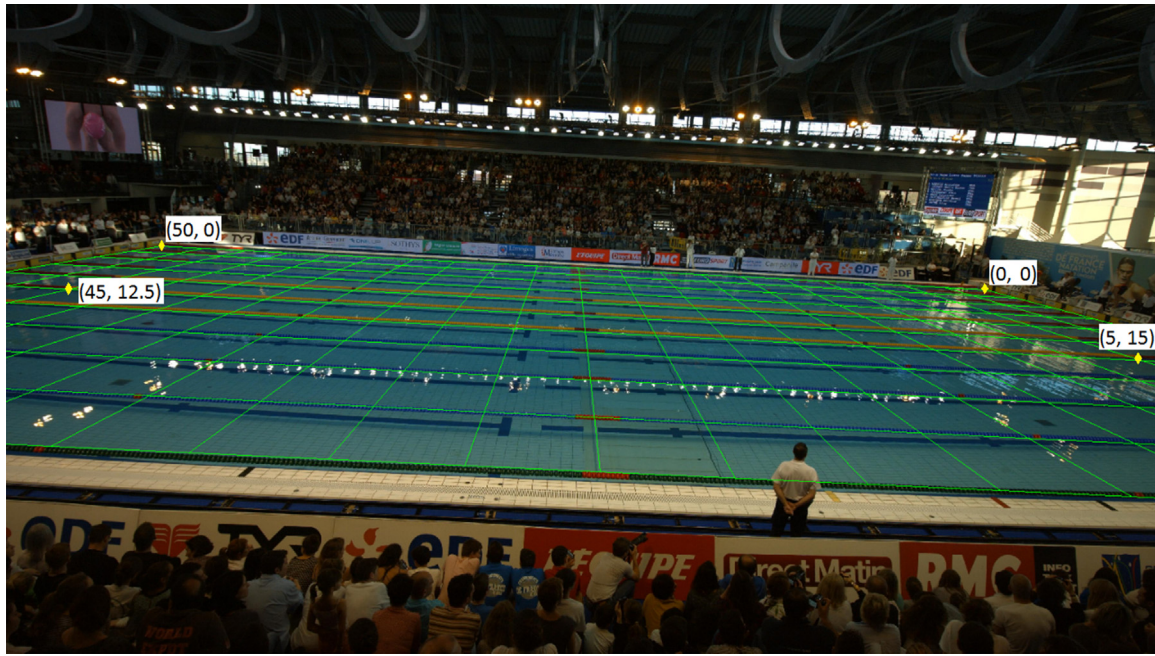


Fig. 1. Swimming pool calibration with the 2D DLT technique after a selection of 4 points. After the calibration, we superimpose a grid of a $50 \times 25 \text{ m}^2$, where the horizontal and the vertical distance between the lines is equal to 2.5 m. (French National Championships, Limoges, April 2015).

$$\begin{cases} u = \frac{L_1x + L_2y + L_3}{L_7x + L_8y + 1} \\ v = \frac{L_4x + L_5y + L_6}{L_7x + L_8y + 1} \end{cases} \quad (2)$$

The Olympic swimming pools that organize major events meet international standards on dimensions, width of swimming lanes and color separators. In Fig. 1, we selected four points for which we knew their metric coordinates. We then calibrated the swimming pool by applying the 2D DLT technique.

3.2. Reference selection

The swimmer's head is tracked because of its position in the water and its shape, easily discriminated from other parts of the body. We manually select the swimmer's head in the first frame that follows the first dive where the head is usually visible. This module can be done automatically using detection and supervised classification techniques. A well-defined database has to be generated by extracting a sufficient number of references from images in the post-dive phase in different competitions [3].

3.3. Trajectory prediction

In order to improve the tracking, we propose a trajectory prediction module. This latter consists of localizing and restricting the target region of interest, while taking into account the different properties of the swimming environment. This can be done thanks to the calibration, which establishes the correspondence between the pixel coordinates on the image and the real metric coordinates. To make a relevant prediction, we propose to consider the following information:

- Swimming pool dimensions

There are two types of Olympic swimming pools based on size: the large pool of $50 \times 25 \text{ m}^2$ and the small pool of $25 \times (10\text{--}20) \text{ m}$. Once the video is calibrated, we can find the metric coordinates of any pixel in the image.

- Lane and direction of swimming

Prior knowledge of the swimming lane number coupled with the calibration allows us to precisely extract the swimming lane from the image. As for the direction, it can be determined by knowing the side on which swimming begins.

- Mean speed

In the first ten images, the swimmer's mean speed is estimated approximately regarding the type of stroke [1–2 m/s]. Then, it is adjusted progressively along the tracking process.

- Referential position update.

The first validated referential swimmer position (X_r, Y_r) is the first reference selected in the first iteration. This position will be updated during tracking in the case where the decision criterion for the detection exceeds an empirical confidence level. Moreover, the actual region of interest position will be chosen according to this updated referential position.

Finally, by combining all these criteria, the region of interest $(W \times H)$ can be predicted at position (X_p, Y_p) as shown in the following formula:

$$\begin{cases} W = 5 * w \\ H = 3 * h \\ X_p = X_r + n * \text{Dist} - \frac{W}{2} \\ Y_p = Y_r - \frac{H}{2} \end{cases} \quad (3)$$

where w and h are respectively the width and length of the reference, X_r and Y_r are the coordinates of the last validated referential position, n is the number of frames from the last validated referential position, and Dist is the mean distance covered by the swimmer in a period of $\frac{1}{\text{FPS}}$ calculated using the mean speed.

3.4. Swimmer detection by dynamic fusion

The detection module is initialized by an input plane containing the reference image of the swimmer and the region of interest. In order to detect the target, two types of descriptors are analyzed, the contour and color, using respectively the NL-JTC and the color

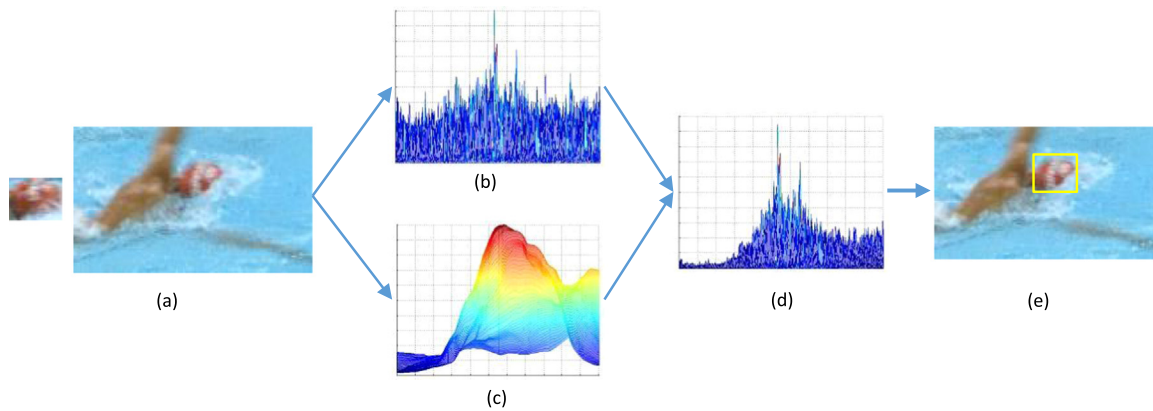


Fig. 2. Detection process using dynamic fusion approach proposed in [17]. (a) Reference and target images, (b) NL-JTC correlation plane, (c) color scores plane, (d) dynamic fusion plane, and (e) detection of the swimmer's head which is located in the same position as the highest peak of the dynamic fusion plane. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

histogram techniques, where each one generates a score plane with values in the interval $[0, 1]$. Next, the two planes are fused and the highest peak of the fusion plane represents the position of the detected target as shown in Fig. 2. It is important to note that the height of this peak is considered as a confidence factor for the detection.

3.5. Reference update

Generally, the reference needs to be updated by the new detected target at the end of each iteration to ensure its temporal coherence and relaunch the next iteration with an updated reference. However, the confidence in the detection is not always sufficient and we might update the reference by a false detected target. To determine the exact moment of the dropout, we propose to use the confidence factor of the dynamic fusion (FC) as a dropout detection criterion.

If the value of the confidence factor is greater than an empirical threshold, then the target is validated. Otherwise, we consider that the tracking is in a state of dropout (loss of the target to be tracked). In this case, the reference will not be updated until the next validated detection.

3.6. Temporal composite filter

We noted that when a single reference is used for tracking, it becomes difficult to track the swimmer over a relatively long period. Indeed, if we use the first reference image, the correlation decreases over time and if we ever actualize by the last detected target, it becomes difficult to recover the target after a dropout. To solve these problems, we propose to keep the first reference to ensure the robustness and the last detected target to ensure the

temporal coherence. This can be done by taking advantage of the composite filter principle [18] proposed for JTC techniques (Fig. 3).

The basic idea of the composite filter is to calculate a weighted sum of the chosen reference images as shown in the following equation:

$$REF(x, y) = \sum \alpha_i * ref_i(x, y) \quad (4)$$

where α_i represents the weight that can be used to promote the reference ref_i .

4. Multi-related-targets approach

The problem that remains difficult to solve is the partial and total occlusion of the target (e.g., the head). Therefore, we propose to predict the position of the swimmer's head by taking into account the swimmer's mean speed and the last validated referential position. However, the prediction is insufficient when the head is occluded for a long period. In this case, we lose in accuracy, which is important for further tasks. In addition to the head, the swimmer's suit can also be tracked. Although it is less visible than the head, its visibility generally alternates with that of the head. Therefore, we propose to track the two targets at the same time and predict the position of the occluded target with respect to the other. For this, we introduce the new multi-related-targets approach based on dynamic fusion.

4.1. Inspired by the multi-tracking approach

In the multi-tracking system proposed by Benarab et al. [13], each track represents an independent tracking process. The goal

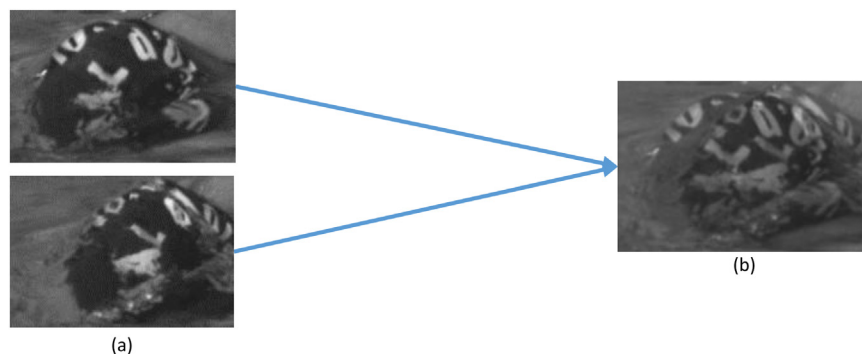


Fig. 3. Principle of the composite filter. (a) Reference images and (b) composite image.

behind this is to check several potential detections by applying different techniques (optical correlation NL-JTC, color histogram, LBP and HOG). This ensures a global description of the target to be tracked.

We were inspired by this idea and wanted to optimize the tracking in the case of occlusions. We therefore proposed a new multi-tracking system, with each track representing the tracking of a different target (head, swimsuit), and named it multi-related-targets tracking.

4.2. Principle of multi-related-targets tracking

Two types of information are required to initialize the multi-related-targets tracking approach. First, the body parts to be tracked are selected and used as references in the first frame. Second, the geometric relationship between them is determined and expressed as Euclidean distance. The angle information is discarded because of its high variability during the swim. In our case, we chose the head and the swimsuit as targets, given that the distance between them is relatively stable and their visibility during swimming alternates.

As shown in Fig. 4, we start by calculating the referential distance (d) between the head and the swimsuit. Then, we apply the dynamic fusion detection technique on the two references separately. The position of the detected target using this technique corresponds to the highest peak in the fusion plane, as proposed by Benarab et al. [17]. However, we propose to check all the peaks that represent potential detections of each target, as detailed in Section 4.3.

A pair (head, swimsuit) is chosen a posteriori from the potential detections of the two targets according to a multi-parametrical criterion. Moreover, if this criterion is greater than a certain threshold, then the confidence in this pair is high. After selecting the best pair (detailed in Section 4), tracking is relaunched with the new validated pair. However, if the value of this criterion is below the threshold, then the confidence in the chosen pair is low. In this case, we validate the target that has the highest confidence factor and reject the other target, then replace it with a prediction based on the geometric relationship with the validated target.

4.3. Potential detections

In order to find all potential detections, we conduct a local maxima search in the fusion plane shown in Fig. 5a. First, we start

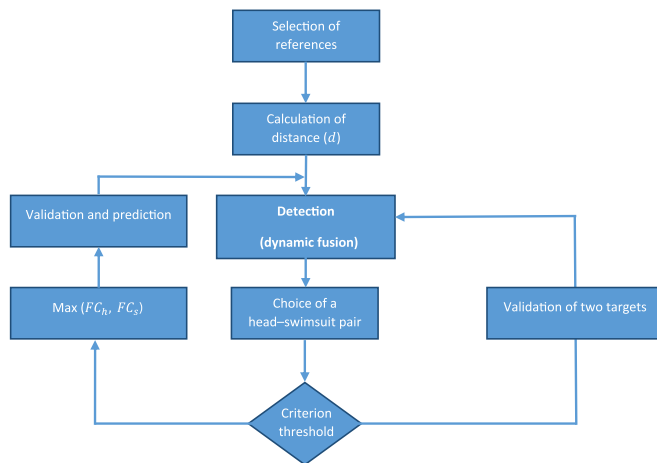


Fig. 4. Multi-related-targets approach. First, we select the references and calculate the distance between them. Then, we apply dynamic fusion technique and we choose the pair (head-swimsuit) with the highest value of the decision criterion. If this value is greater than the threshold then we validate the two targets, else, we validate the target with the highest criterion value and we predict the other.

by denoising the plane that seems significantly noisy. For this, we use a 2D median filter to obtain the plane shown in Fig. 5b. Then, we apply thresholding to the denoised plane. Next, we smooth our plane using a Gaussian filter (Fig. 5c). The purpose of this step is to have peaks with a Gaussian form with a single pixel corresponding to the local maximum for each Gaussian. Finally, all the potential pairs are matched (Fig. 5) and prepared for the next step which is decision.

4.4. Complex decision criterion

The local maxima found for the two targets correspond to the potential detections. In order to accurately detect the swimmer, we check all possible head-swimsuit pairs $M \times N$, where M and N are the number of potential detections for the head and the swimsuit respectively. We choose the best pair according to a complex decision criterion that consists of measuring the confidence in all the $M \times N$ potential detection pairs. It consists of four values calculated for each pair.

• Confidence factor

FC_h and FC_s represent the confidence factors of the head and the swimsuit potential detections, which are calculated by applying the dynamic fusion approach.

• Inter-target distance

This value is obtained by calculating the Euclidean distance between the potential detections of the head $X_{i(1..M)}$ and the swimsuit $Y_{j(1..N)}$. Therefore, the score matrix S_{Dist} is calculated as shown in the following equations:

$$D_{i,j} = \|X_i - Y_j\|_2 \quad (5)$$

$$S_{Dist(i,j)} = f(D_{i,j}) \quad (6)$$

where f_{μ_f, σ_f} is a Gaussian function with μ_f representing the referential distance between the head and the swimsuit calculated in the initialization phase. σ_f is set empirically after a statistical study of the variation of the distance between the head and the swimsuit during a crawl swim. σ_f is calculated using the following equation:

$$\sigma_f = \frac{1}{2}(\max(D) - \min(D)) \quad (7)$$

• Speed

This criterion is based on the speed calculation and it can be introduced after a few iterations (10 iterations). This is important as it gives an indication of the swimmer's progress in the first frames, which can then be used for the rest of the swim.

For each potential detection of the head-swimsuit pair, we calculate the instantaneous speed $V_i(t)$ of the whole block. This latter will be represented by the head speed during the period Δt .

$$V_i(t) = \frac{\|X_i(t) - X_i(t-1)\|_2}{\Delta t} \quad (8)$$

$$S_{V_i(t)} = g(V_i(t)) \quad (9)$$

where g_{μ_g, σ_g} is a clipped Gaussian function as shown in Fig. 6, with μ_g being the mean speed of the swimmer calculated between the current time t and the beginning. σ_g is estimated based on the variation in the swimmer's instantaneous speed, as shown in the following formula:

$$\sigma_g = \frac{1}{2}(V_{max} - V_{min}) \quad (10)$$

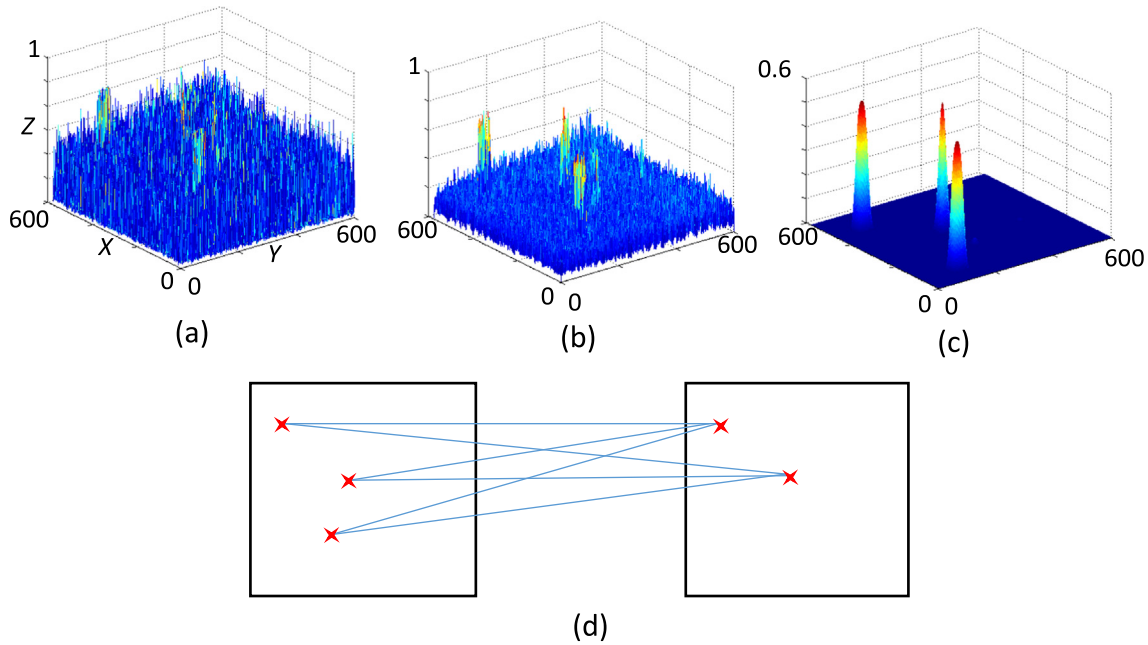


Fig. 5. Detection of local maxima in the fusion plane, where, (X,Y) denotes its size and Z denotes the output intensity. (a) Dynamic fusion plane, (b) denoising by the median filter, (c) Gaussian filtering and thresholding, and (d) associating head-swimsuit pair: on the left, potential detections of the head – on the right potential detections of the swimsuit.

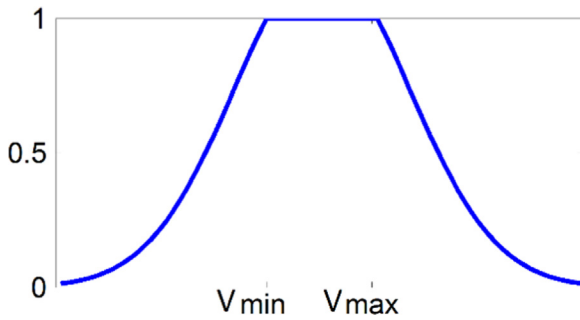


Fig. 6. Clipped Gaussian function around the values of V_{max} and V_{min} .

V_{max} and V_{min} represent the maximum and minimum speeds that are estimated to ± 0.3 m/s around the mean speed. The Gaussian function g is clipped around the V_{max} and V_{min} values to promote all the pairs that have an instantaneous speed included in that interval.

• Final criterion

Last, the final criterion for the selection of the best potential detected pair consists of a weighted sum of the four criteria introduced previously. Two criteria are favored: first, FC_h due to the high visibility rate of the head compared with the swimsuit, second, S_{Dist} due to its relevance. The multi-related-targets criterion $MRTC$ is calculated as shown in the following equation:

$$MRTC(i, j) = \frac{1}{3}(FC_h(i) + S_{Dist}(i, j)) + \frac{1}{6}(FC_s(j) + S_v(i, j)) \quad (11)$$

This criterion generates values in the range of $[0, 1]$ since all the criteria that compose it are within this range. This allows a percentage estimation that is relevant and reliable.

5. Experimental results

The objective of these tests is to show the enhancement gained by using the multi-related-targets tracking approach compared to

the classical technique. In this section, we present the experimental protocol and the results of the comparison studies of three systems. The first system is based on the proposed multi-related-targets approach. The second one is based on a mono-target approach, which will be called the classical head-only tracking approach. The last is the ground truth, which is the manual annotation of the swimmer's position in every frame of the video.

5.1. Experimental protocol

To evaluate the proposed approach, we conducted several tests on high-level swimmers during training (25-m pool, Foch, Brest), the National Championships (50-m pool, Limoges, April 2015) and World Championships (50-m pool, Kazan, July 2015).

For shooting, we used Blackmagic 4 K cameras whose characteristics are: 4 K (3840×2160 pixels), 25 frames/s. In the 25-m pool, a single camera was used to cover the entire pool. However, for shooting with high resolution in the 50-m pool, we used two cameras with each covering half of the pool. In this case, an overlap between fields of view is necessary to ensure the continuity of the tracking.

5.2. Tracking percentage

Tracking percentage is calculated by measuring the distance between the positions of the detected swimmer using the automatic tracking approaches and the ground truth. The decision is made according to a threshold calculated relative to the size of the reference that is used, as shown in the following equation:

$$Threshold = \frac{\sqrt{Ref_x^2 + Ref_y^2}}{2} \quad (12)$$

where Ref_x and Ref_y are the reference dimensions. In other words, the chosen threshold is equal to half the reference diagonal, which allows at least a quarter overlap between the detected target and the annotated one.

The results in Table 1 show that the multi-related-targets approach significantly improved the tracking percentage of several

Table 1

Comparison between the classical head-only tracking approach and the multi-related-targets approach in terms of tracking percentage.

Sequence	Type of swim	Classical head-only (%)	Multi-related-targets (%)
Sequence 1	Training – 100 m Crawl	92.24	97.78
Sequence 2	Training – 100 m Crawl	90.10	96.55
Sequence 3	Competition – 1500 m Crawl	65.77	87.92
Sequence 4	Competition – 1500 m Crawl	72.84	90.60
Sequence 5	Competition – 1500 m Crawl	67.79	93.96

swimmers in different sequences. In particular, applying this approach in the training sequences slightly enhanced the tracking percentage (5%, 6%) because, in this case, the swimmer moves slowly with a mean speed of 1 m/s. Consequently, there is less splashing and the enhancement due to the multi-related-targets approach concerns only the case of the occlusions associated with hand passages. In addition, the localization accuracy in the case of partial occlusions of the target to be tracked is enhanced as well.

However, in the case of 1500 m crawl competition sequences, the athletes swim relatively faster than training, at a mean speed of 1.5 m/s. They consequently generate more splashing, which degrades the classical head-only tracking approach and at the same time shows the interest of the proposed approach, for which the tracking percentage improved by 16%, 18% and 22% for different swimmers. The explanation is that the multi-related-targets approach avoids the dropout in the case of head occlusions by estimating its position with respect to the swimsuit.

The curves in Figs. 7 and 8 concern respectively the sequence 1 and 3 of Table 1. These curves show the distance between the positions of the tracked swimmer using both of the tracking approaches (Multi-related-targets and classical head-only) and the ground truth. This latter consists of a manual annotation of the swimmer's position in each frame of the video sequence. The smaller the distance, the more accurate the tracking approach.

5.3. Instantaneous speed

In order to estimate the instantaneous speed, we measure the distance covered by the swimmer in a period of $\frac{1}{FPS}$ where FPS is the number of frames per second. Two criteria are proposed to evaluate the speed: the mean and the standard deviation of the distance between the curves generated by applying the two automatic tracking approaches and those generated by the ground truth annotation.

This study is conducted by calculating the instantaneous speed of the same sequences. Table 2 shows that the multi-related-targets approach significantly improves the results compared with the classical head-only approach.

For instance, in sequence 1, the mean distance between the

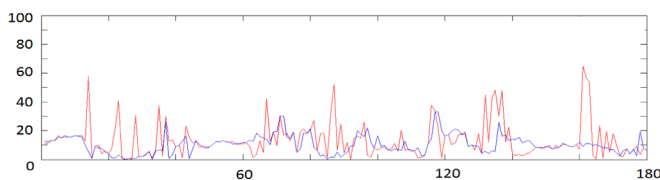


Fig. 7. Comparison between the proposed multi-related-targets approach (in blue) and the classical head-only approach [17] (in red). X: 180 frames (sequence 1), Y: distance between the swimmer's position using the two tracking approaches and the ground truth annotation. Type of swim: Crawl – training – Brest, France 2015. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

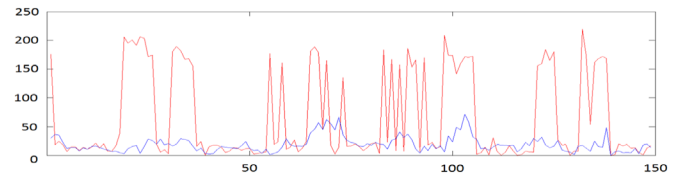


Fig. 8. Comparison between the proposed multi-related-targets approach (in blue) and the classical head-only approach (in red) [17]. X: 150 frames (sequence 3), Y: Distance between the swimmer's position using the two tracking approaches and the ground truth annotation. Type of swim: 1500 m crawl – French national championships, Limoges, April 2015. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Evaluation and comparison between the instantaneous speed generated by the automatic tracking approaches. The criteria are the mean and standard deviation of the distance between their instantaneous speed and the ground truth.

Sequence	Classical head-only		Multi-related-targets	
	Mean (m/s)	STD (m/s)	Mean (m/s)	STD (m/s)
Sequence 1	1.48	2.09	0.62	0.48
Sequence 2	1.61	2.44	0.6	0.55
Sequence 3	5.13	7.05	0.63	0.64
Sequence 4	2.48	3.62	0.49	0.38
Sequence 5	4.65	6.98	0.44	0.35

instantaneous speed of the classical approach and the ground truth equals 1.48 m/s. Using the proposed approach, we reduced the mean distance to 0.62 m/s. We note the same for the standard deviation, which improved and decreased from 2.09 m/s to 0.48 m/s. This is due to the optimization provided by the multi-related-targets approach, which consists of reducing the distance from the ground truth in terms of swimmer position in each frame as well as the instantaneous speed.

5.4. Instantaneous speed filtering

The world record for the men's 50 m crawl was beaten by César Cielo in the Brazil Championships in 2009. He swam with a mean speed of 2.39 m/s. By analyzing the curve of the instantaneous speed generated by the ground truth annotation of this swimmer, we notice that it varies approximately ± 0.3 m/s. It is important to note that the curves of the gross instantaneous speed are very noisy. Consequently, in order to extract the relevant information, we opt for applying Gaussian filtering on the curves with different size windows, as shown in Table 3.

The results in Table 3 illustrate the significant enhancement brought by applying Gaussian filtering on the instantaneous speed calculated for sequences 1 and 3. Moreover, we note that the distance between the instantaneous speed generated by the

Table 3

Effect of Gaussian filtering with different size windows on the instantaneous speed curves.

Sequence	Filtering window	Classical head-only		Multi-related-targets	
		Mean (m/s)	STD (m/s)	Mean (m/s)	STD (m/s)
Sequence 1	1	1.48	2.09	0.62	0.48
	10	0.36	0.42	0.22	0.25
	20	0.15	0.14	0.11	0.1
Sequence 3	1	5.13	7.05	0.63	0.64
	10	1.47	1.21	0.22	0.16
	20	0.66	0.45	0.15	0.12

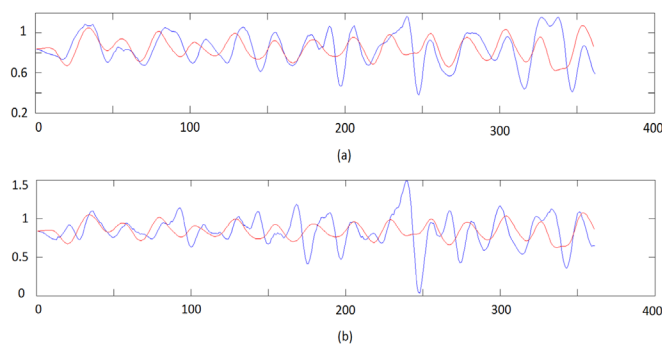


Fig. 9. Comparison between the automatic approaches and the ground truth in terms of filtered instantaneous speed curves. Size of the filtering window: 20. Red: ground truth. Blue: automatic approach. (a) Multi-related-targets and (b) classical head-only. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

automatic approaches and the ground truth is decreased with respect to the increase in the filtering window size. By analyzing the effect of the variation of the filtering window size on the ground truth instantaneous speed curves, we concluded that the optimal size is 20, where we note the relative match between swimmer advancement and the information provided by the generated curves. Indeed, the smaller windows generate noisy curves and the larger ones diminish toward the mean speed, which leads to a loss of relevant information. On the other hand, the multi-related-targets approach shows better performances than the classical head-only. This is illustrated clearly in the following curves of Fig. 9, where the proposed approach significantly matches the ground truth better than the classical approach.

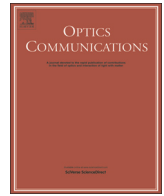
6. Conclusion

In this paper, a new multi-related-targets tracking approach is proposed to optimize swimmer tracking based on dynamic fusion detection. The main advantage of this approach is that it strengthens the robustness of the swimmer tracking system with respect to partial and total occlusions of the main target. In addition, this approach enhances the accuracy of localization by checking all the potential detections and choosing the optimal position according to a complex decision criterion. This criterion is composed of the detection confidence factor, the distance between the targets, and the swimmer's speed. A detailed comparative study evaluated the proposed approach in comparison with the classical approach and the ground truth. This evaluation included the tracking percentage, the instantaneous speed and filtering. Finally, the multi-related-targets approach clearly optimized

swimmer tracking in terms of accuracy as well as avoiding the occlusion cases, which led to an enhancement of $\sim 6\%$ in training and $\sim 22\%$ in competition. However, we still observe differences in the instantaneous speed curves compared with the ground truth.

References

- [1] Elipot M. Application of new techniques of camera calibration and the cinematic analysis of swimming [Ph.D. thesis]. Paris Descartes university; 2010.
- [2] Puel F, Morlier J, Avalos M, Mesnard M, Cid M, Hellard P. 3D kinematic and dynamic analysis of the front crawl tumble turn in elite male swimmers. *J. Biomech.* 2012;35:510–5.
- [3] Benarab D, Napoléon T, Alfalou A, Verney A, Hellard P. All-automatic swimmer tracking system based on an optimized scaled composite JTC technique. In: Proceedings of the SPIE 9845, Optical Pattern Recognition XXVII, 98450E; 2016.
- [4] Pogalin E, Thean AHC, Baan J, Schipper N, Smeulders AWM. Video-based training registration for swimmers. *Int. J. Comput. Sci. Sport* 2007;1.
- [5] Javidi B. Nonlinear joint power spectrum based optical correlation. *Appl. Opt.* 1989;12:2358–67.
- [6] Alfalou A, Brosseau C. Understanding correlation techniques for face recognition: from basics to applications. *Face Recognition Book*; 2010. p. 354–80.
- [7] Lu W, Tan Y-P. A color histogram based people tracking system. In: Proceedings of the IEEE international symposium on circuits and systems, ISCAS; 2001.
- [8] Junxiang G, Tong Z, Yong L. Face tracking using color histograms and moment invariants. In: Proceedings of the 2nd IEEE international conference on broadband network multimedia technology, IC-BNMT'09; 2009.
- [9] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 2006;12:2037–41.
- [10] Jenicka S, Suruliandi A. Empirical evaluation of distance measures for supervised classification of remotely sensed image with modified multivariate local binary pattern. In: Proceedings of the 2011 international conference on emerging trends in electrical and computer technology (ICETECT); 2011.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, CVPR; 2005.
- [12] Wang X, Han TX, Yan S. An HOG-LBP human detector with partial occlusion handling. In: Proceedings of the ICCV; 2009.
- [13] Benarab D, Napoléon T, Alfalou A, Verney A, Hellard P. A novel multitasking system for the evaluation of high-level swimmers performances. In: Proceedings of the SPIE 9094, Optical Pattern Recognition XXV, 90940A; 2014.
- [14] Alfalou A, Ben Haj Yahia N, Alam MS. Face recognition using a non-zero-order correlation plane and a nonlinear joint transform correlator. In: Proceedings of the SPIE 8498; 2012.
- [15] Gagunashvili N. Chi-square tests for comparing weighted histograms. *Nucl Instrum Methods Phys Res Sect A: Accel, Spectrom, Detect Assoc Equip* 2010; no. 2:287–96.
- [16] Aziz AYI, Karara HM. Direct linear transformation into object space coordinates in close-range photogrammetry. In: Proceedings of the symposium on close-range photogrammetry, Urbana, Illinois; 1971.
- [17] Benarab D, Napoléon T, Alfalou A, Verney A, Hellard P. Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques. *Opt Commun* 2015;356:256–68.
- [18] Alfalou A, Keryer G, De Bougrenet De La Tournaye J-L. Optical implementation of segmented composite filtering. *Appl Opt* 1999;38:6123–35.
- [19] Katz P, Aron M, Alfalou A. Joint transform correlation for face tracking: elderly fall detection application. In: Proceedings of the SPIE 8748; 2013.



Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques

D. Benarab^{a,b,*}, T. Napoléon^a, A. Alfalou^{a,**}, A. Verney^b, P. Hellard^c

^a ISEN Brest, Département Vision, L@bISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France

^b ACTRIS Brest, 24 Rue Victor Grignard - Guipavas - BP 143 - 29803 Brest Cedex 9, France

^c FFN (Fédération Française de Natation), 14 Rue Scandicci 93508 Pantin Cedex, France

ARTICLE INFO

Article history:

Received 9 June 2015

Accepted 21 July 2015

Keywords:

Correlation

NL-JTC

Tracking systems

Swimmer tracking

Color histogram

Normalization

Dynamic fusion

ABSTRACT

To design a robust swimmer tracking system, we took into account two well-known tracking techniques: the nonlinear joint transform correlation (NL-JTC) and the color histogram. The two techniques perform comparably well, yet they both have substantial limitations. Interestingly, they also seem to show some complementarity. The correlation technique yields accurate detection but is sensitive to rotation, scale and contour deformation, whereas the color histogram technique is robust for rotation and contour deformation but shows low accuracy and is highly sensitive to luminosity and confusing background colors. These observations suggested the possibility of a dynamic fusion of the correlation plane and the color scores map. Before this fusion, two steps are required. First is the extraction of a sub-plane of correlation that describes the similarity between the reference and target images. This sub-plane has the same size as the color scores map but they have different interval values. Thus, the second step is required which is the normalization of the planes in the same interval so they can be fused. In order to determine the benefits of this fusion technique, first, we tested it on a synthetic image containing different forms with different colors. We thus were able to optimize the correlation plane and color histogram techniques before applying our fusion technique to real videos of swimmers in international competitions. Last, a comparative study of the dynamic fusion technique and the two classical techniques was carried out to demonstrate the efficacy of the proposed technique. The criteria of comparison were the tracking percentage, the peak to correlation energy (PCE), which evaluated the sharpness of the peak (accuracy), and the local standard deviation (Local-STD), which assessed the noise in the planes (robustness).

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Research in the field of people tracking systems has resulted in a multitude of applications [4,1,6,21,12], notably in security, health-care and sports. For example, video surveillance systems are used to detect and record suspicious activities and enhance public safety [34,28]. In the health-care field, systems are used to track patients e.g. elderly to detect falls or other disturbing behaviors and to send an alarm message to the appropriate caregiver [23]. In sports, research into tracking devices has essentially been dedicated to the evaluation and improvement of athletic performances [29,18,26]. The French Swimming Federation (FFN, Fédération Française de Natation) [14] is one of the leaders in research on swimmer evaluation and the optimization of swimming

strategies in France [13,30]. A recent FFN project was to develop an automatic swimmer tracking system without physical markers or sensors. For that, we explored the possibility of using vision sensors and tracking the swimmers via video sequences. Swimmer tracking is a difficult task, in great part because of the aquatic environment: splashing, water movement and the movement of the swimmers through the water. In the literature, various object tracking systems have been proposed [4,12,23,31,17], such as those based on correlation techniques like the nonlinear joint transform correlator (NL-JTC) [3,23]. These systems take a plane containing a reference image and a target image as an input and provide a correlation plane that contains two peaks corresponding to the location of the tracked object as an output. Other tracking systems are based on color histograms [18,17], where the reference object is sought in the target scene through comparisons based on each color coded in the histogram. In the same vein, information on textures (local binary patterns: LBP) [10,31] or gradients (histogram of oriented gradient: HOG) [12,34] can be used to determine the similarity between two images.

* Corresponding author.

** Principal corresponding author.

E-mail addresses: djamel.benarab@isen-bretagne.fr (D. Benarab), ayman.al-falou@isen-bretagne.fr (A. Alfalou).

Unfortunately, most of these approaches are not robust enough to be used in an uncontrolled environment such as swimmer tracking in swimming pools with a range of limitations. In particular, correlation-based techniques, as well as the LBP and HOG, are sensitive to contour deformations and sometimes generate artifacts in the correlation plane, related to the correlation between the reference and target edges. On the other hand, color-based techniques can confuse objects having the same color and they are inaccurate in terms of localization. Last, the calculation time for techniques based on histograms is often long because an exhaustive search is needed to locate the reference image in the target scene. To overcome these problems and design a robust tracking system, we combined the correlation and color histogram techniques. Similar to Alfalou et al. [5], who applied a nonlinear function to the correlation plane in order to denoise it, we aimed to improve detection by optimizing the decision. Alfalou et al. [5] proposed a non-linear function that enhanced the correlation peak and reduced the noise, which improved the value of their decision criterion based on the Peak-To-Correlation energy (PCE) [4]. However, this optimization is blind in that it does not take into account any information related to the tracked object, nor does it allow the decision to be influenced in any way.

To overcome these shortcomings, we propose a new approach that directly combines the score planes of two methods: correlation and color histogram. This can be seen as a dynamic filtering of the correlation plane by a specific function – i.e. the color scores plane – that depends on the tracked object and adapts to its variations and evolution. This dynamic correlation plane containing the information on the similarity in color between the reference and the target enhances the contour information from the correlation and thereby improves the final decision. In addition to our main contribution based on the fusion of an optical method (correlation) and a numerical method (color histogram), we carefully analyzed the correlation planes and color scores and propose the following optimizations:

- Extraction of a potential correlation sub-plane that is the same size as the color scores plane to enable the dynamic fusion of the two planes.
- Optimization of the input plane by reducing its size in order to improve the calculation time for the correlation.
- Elimination of artifacts related to the correlation approach.
- Pretreatment of the input images with a Sobel filter to minimize the artifacts resulting from the blurred movement in the video sequences.
- Use of the integral images proposed by Viola and Jones [32] to improve the calculation time for the color histogram approach. In this paper, we begin by presenting the correlation and color histogram techniques in Section 2. In Section 3, we explain our technical choices for tracking and then present our dynamic fusion approach and show its advantages through experiments on a synthetic database. Several correlation plane optimizations are proposed and detailed in Section 4. The application of our fusion approach to swimmer tracking is evaluated on videos of

real competitions in Section 5. We conclude in Section 6 by highlighting our contributions and the perspectives offered by our work.

2. Tracking techniques

Numerous techniques have been proposed in the literature on object tracking [4,11,18,12,9,10]. Among the best known are correlation [4,24,23], color histogram [18,8,17], histogram of oriented gradient (HOG) [12,33] and local binary patterns (LBP) [10,2]. In this section, we focus on the first two techniques, which use different but important information: contour and color for the correlation and the color histogram technique respectively. We then consider the strengths and weaknesses of each method in order to optimally combine them in the process presented in Section 3.

2.1. Correlation approaches

Correlation is a classical technique, well known in the field of optical pattern detection and recognition. Two architectures have been proposed [4]: the Vanderlugt correlator (VLC), mainly used for identification, and the joint transform correlator (JTC), which is more robust for detecting and localizing a given object in a scene. The JTC architecture is preferred as part of a tracking system because of its ability to detect and localize at the same time, thereby improving the calculation time. However, this originally optical approach is physically difficult to set up, especially in our swimmer tracking application. For this reason, we chose to implement it numerically through an image processing algorithm.

The fundamental principle of the JTC architecture is the application of a Fourier transform to an input plane made up of a reference image and a target image – i.e., a target scene – to obtain a joint spectrum of these images. Then, an inverse Fourier transform is applied to the intensity of this spectrum in order to obtain the correlation plane containing the information on the reference/target similarity. The architecture of the JTC is presented in Fig. 1 and can be formulated as follows:

$$f(x, y) = c(x, y) + r(x - d, y - d) \quad (1)$$

where $f(x, y)$ in the spatial domain represents the input plane consisting of the target image c of the coordinates (x, y) and the reference image r of the coordinates $(x - d, y - d)$ placed at a distance d from the target. The classical joint spectrum $F(u, v)$ in the frequency domain is thus obtained by the Fourier transform applied to the input plane (Eq. (1)):

$$\begin{aligned} F(u, v) &= FT(c(x, y)) + FT(r(x - d, y - d)) \\ F(u, v) &= |C(u, v)| \exp[\phi_c(u, v)] \\ &\quad + |R(u, v)| \exp[\phi_r(u, v)] \exp[-j(ud + vd)] \end{aligned} \quad (2)$$

Then, the intensity of the classical joint spectrum is recovered using the following formula:

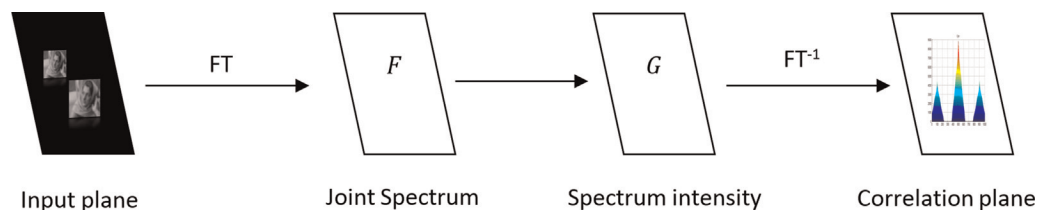


Fig. 1. C-JTC correlator architecture. The input plane contains both images reference and target. Then, the Fourier transform is applied to obtain the joint spectrum. Finally, the correlation plane is obtained by applying the inverse Fourier transform on the intensity of the joint spectrum. The obtained correlation plane contains two peaks of cross-correlation and an auto-correlation peak.

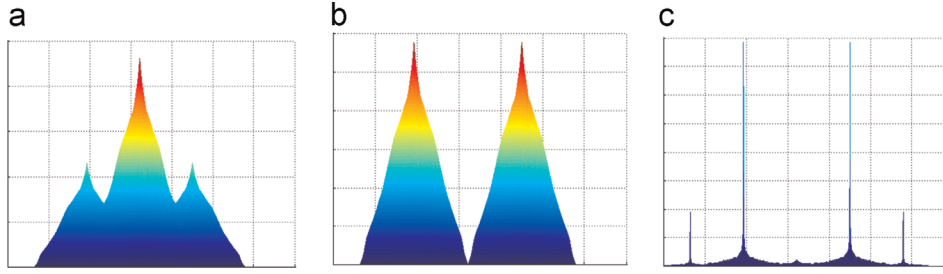


Fig. 2. Example of a same correlation plane obtained using different methods: (a) C-JTC, (b) NZ-C-JTC and (c) NL-NZ-JTC with $k=0$.

$$G_{CL-JTC}(u, v) = |F(u, v)|^2 = |C(u, v)|^2 + |R(u, v)|^2 + |C(u, v)| \exp[\phi_c(u, v)] |R(u, v)| \exp[-\phi_r(u, v) + j(ud + vd)] + |C(u, v)| \exp[-\phi_c(u, v)] |R(u, v)| \exp[\phi_r(u, v) - j(ud + vd)] \quad (3)$$

The correlation plane is obtained by applying the inverse Fourier transformation to the intensity of the classical joint spectrum G_{JTC} .

In the case of a similarity between the reference and the target, the correlation plane will contain three peaks: a very intense autocorrelation peak that represents the correlation of each image with itself and two peaks of cross-correlation between the reference and target images (Fig. 2a). In the rest of the paper, we will refer to these cross-correlation peaks by the correlation peaks. In the case of a tracking system, only the latter information is pertinent because it represents the degree of similarity between the two images as well as the position of the object in the target scene (through the direction and the length of the segment linking both peaks).

The autocorrelation peak is generated by the inverse Fourier transform of the term $|C(u, v)|^2 + |R(u, v)|^2$. To remove it, the non-zero-order C-JTC architecture (NZ-C-JTC) subtracts this term from the intensity of the classical joint spectrum (Eq. (3)) [25]. We then obtain the following:

$$G_{NZ-JTC}(u, v) = |F(u, v)|^2 - (|C(u, v)|^2 + |R(u, v)|^2) \quad (4)$$

Sometimes, this formulation is not sufficient to make the central autocorrelation peak fully disappear after applying the inverse Fourier transform. In this case, the center of the correlation plane can be filtered to remove the autocorrelation zone. One of the weaknesses of the NZ-C-JTC approach is its inability to accurately locate the object in the target due to the width of the correlation peaks. To overcome this problem, the non-linear JTC (NL-NZ-JTC) has been proposed [19] to control the width of the color histogram peaks by applying non-linear thresholding to the intensity of the joint spectrum:

$$G_{NL-JTC}(u, v) = |G_{NZ-JTC}(u, v)|^k \quad (5)$$

where k is the degree of non-linearity. In particular, the NZ-C-JTC architecture, whose accuracy is reduced because of the width of the peaks, can be obtained by using a degree $k=1$ (Fig. 2b). Conversely, a degree $k=0$ reduces the width of the peaks maximally (Fig. 2c), but the appearance of harmonics sometimes makes the method less robust in the case of a low-noise image. In our case, a compromise between accuracy and robustness can be found by using a k between 0.2 and 0.5.

For our swimmer tracking application, we propose using the NL-NZ-JTC architecture (Eq. (5)) because it simultaneously provides the detection and localization of the reference image in the target scene. In addition, as shown in Fig. 2c, localization can be both accurate and robust thanks to the application of a non-linearity coefficient. However, despite these advantages, this approach

has some disadvantages, especially high sensitivities to edge deformation, object rotation and a change in scaling. To overcome these limitations, we propose to use other information to enhance the performance of the correlation. In particular, the color information encoded in the form of histograms seems relevant because it is very different from the contour information. In addition, this technique is widely used in the literature [18,17] for object tracking in environments where the background color is different from the reference color.

2.2. Color histogram

Using the color histogram, the main information to describe an object is the color, specifically the distribution of its gray scale [8,18,17]. The basic idea is to represent the two images, reference and target, with a histogram where each component codes the number of pixels of a given intensity. The number of components, which corresponds to the intervals of intensities taken into account, is a parameter to be defined using a training set, for example. With this approach, the degree of similarity between the reference and target images is obtained by comparing their respective histograms using an adapted measure, like the *Bhattacharyya* distance [20] or χ^2 (Chi-Square) [15].

To detect a given object within a target scene, an exhaustive search can be conducted by comparing the reference image with a window of the same size that can cover the entire target. Here, the comparison is based not on the pixels themselves, but on the color histogram of each area. In order to improve the computation time of this approach, we propose an adaptation for the histograms of the integral images proposed by Viola and Jones [32]. This avoids recalculating the entire histogram when two areas have a large overlap. Last, if this approach is combined with the *Bhattacharyya* distance, the score for each comparison is between 0 and 1, which provides a plane, as shown in Fig. 3(a)–(c). Its size is given by

$$\text{size} = (\text{height_target} - \text{height_ref}) \times (\text{width_target} - \text{width_ref}) \quad (6)$$

The literature suggests that several detection and tracking systems based on color histogram are of interest [27,8], particularly when the background color and the color of the tracked object are very different. However, this technique is inaccurate for localizing the object, as can be seen in Fig. 3c, where we note very large peaks, and it cannot be used when the colors of the background and the object are too close.

3. Dynamic fusion

As shown in Section 2, the use of a single tracking method, whether correlation or color histogram, is insufficient for accurate detection. However, the two techniques are complementary. The correlation is precise in terms of localization but is too sensitive to contour deformations, whereas the color histogram approach is imprecise in terms of localization but robust regarding the contour

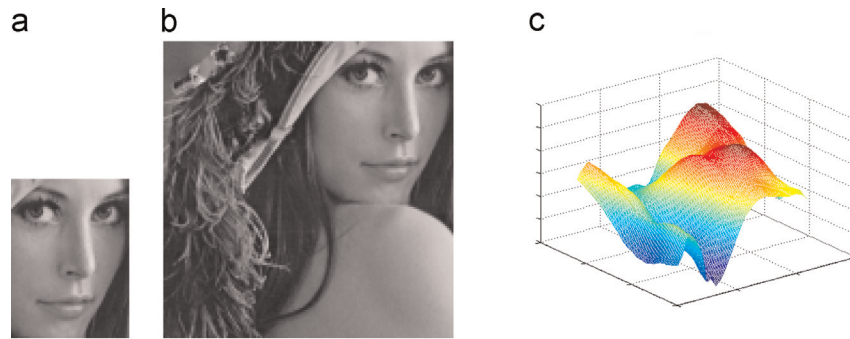


Fig. 3. Color histogram approach. (a) Reference image, (b) target image, and (c) the plane of color scores of similarity (color scores map).

deformations. We thus combined the two techniques to design a tracking system that would be accurate for localization and robust with respect to contour deformation.

3.1. Dynamic fusion concept

In order to optimize the correlation decision, Alfalou et al. [5] proposed a filtering approach to raise the energy of the correlation peak and therefore the value of the peak to correlation energy (PCE). This new approach applies denoising to the correlation plane by multiplying it by an empirically defined non-linear function to reduce the noise of the correlation plane and increase the peak energy. Therefore, it is called “blind” because it is not linked to the information extracted from the plane. The main advantages of this approach are the efficacy of the simulation and its ability to increase the discrimination of the used correlation approach. However, this technique has no influence on the decision because it does not change the position of the color histogram peaks.

As these weaknesses are closely related to the blind and empirical aspect of the chosen non-linear function, we propose a more robust approach based on the fusion of different types of information. We thus suggest a robust decision based on the advantages of each approach while keeping a close link with the input data. The fusion does not occur in the decision step but it directly concerns the probability planes of each method. In this paper, we use two approaches: the NL-NZ-JTC technique and color histogram approach. This new method can be seen as a filter for the correlation plane by a reliable “dynamic” function, namely the color scores plane. The fusion of the two planes will give a richer description of the tracked object by combining contour and color information.

Given their characteristics, the two approaches cannot be fused directly. The scores planes must have the same size and common dynamics between 0 and 1, for example. Based on the observation that the correlation plane contains two identical zones (top/left and bottom/right) and two other zones that contain no information (top/right and bottom/left), we propose to extract a sub-plane, which we call a “potential zone of correlation” (Section 3.3). This latter has the same size as the color scores plane. Thus, only the step of normalizing the plane dynamics remains to be defined in order to apply our dynamic fusion. The operation of our fusion approach is summarized in Fig. 4.

3.2. Input plane generation

The creation of the input plane for the JTC architecture is an important step because the position of the correlation peak will depend on the location of the target and reference images in this plane. The standard input plane proposed in the literature [4,23] is constructed as follows:

$$(c_x, c_y) = (\text{padding}, \text{padding})(r_x, r_y) = (\text{padding} + c_x, \text{padding} + c_y) \quad (7)$$

where the pairs (c_x, c_y) and (r_x, r_y) represent respectively the coordinates of the upper/left corners of the target and reference in the input plane. The padding is the offset between the target image and the edge of the input plane which prevents useful information – i.e. the color histogram peak – from overflowing the correlation plane. This value is set empirically to the height of the reference (Fig. 5).

Finally, the size of the input plane is given by

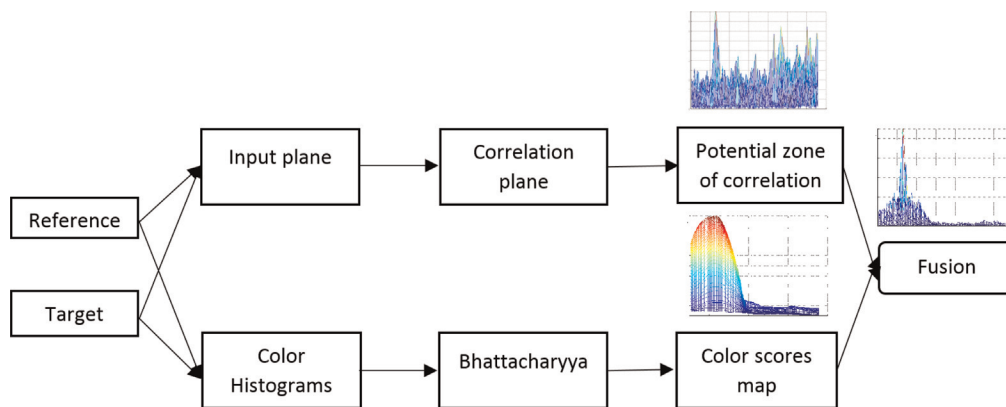


Fig. 4. Fusion process of the correlation plane and the color scores map. For the correlation the reference and target are placed in the input plane, on which the NL-NZ-JTC technique is applied. A sub-plane, called potential zone of correlation, is then extracted. For the color histograms, a window browse the target image making comparisons with the reference image using the Bhattacharyya distance of the two histograms. The obtained scores form the color scores map. Both planes are then fused into a single plane describing the similarity between the reference and the target in terms of both contour and color.

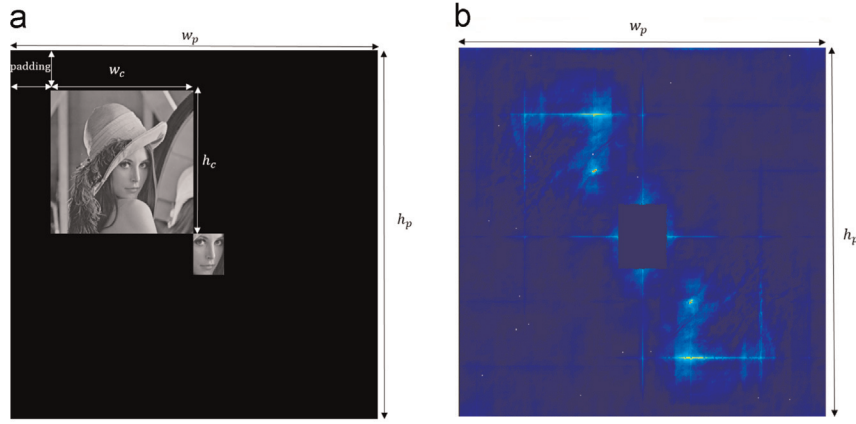


Fig. 5. Application of NL-NZ-JTC technique: (a) input plane where the images are placed diagonally and (b) correlation plane. Here, a filter is applied to the center of the correlation plane to remove the autocorrelation peak.

$$\begin{cases} w_p = 2*w_c + 2*padding \\ h_p = 2*h_c + 2*padding \end{cases} \quad (8)$$

Where (w_p, h_p) , (w_c, h_c) represent respectively the size of the input plane and the size of the target image. As explained in Section 3.1, the sizes of the correlation plane and the color scores map must be matched. To do this, we located and accurately extracted the potential zone of correlation, which can also be used to improve the localization and the computation time.

3.3. Extraction of the potential zone of correlation

The classic correlation plane, as presented in the previous section (Fig. 5), can be separated into two distinct areas: the first one is bearing information (color histogram peak) distributed between the upper/left and lower/right corners, the other not bearing information and distributed between the lower/left and upper/right corners, as clearly identifiable in Fig. 5b. On this basis, we propose restricting the correlation plane to a sub-plane containing only one correlation zone representing the similarity between the reference and the target. Thus, the "potential zone of the correlation" can be merged with our color scores map, thereby limiting the possibility of finding false peaks in the rest of the plane.

Therefore, our aim was to accurately locate this potential zone of correlation. To do so, we relied on three known relationships between the position of the two peaks of correlation and their position in the input plane:

- The distance between the peaks in the correlation plane is equal to twice the distance in the input plane between the center of the reference image and the center of the tracked object in the target scene.
- The slope of the line passing through the peaks is equal to the slope of the line passing through the center of the reference

image and the center of the tracked object.

- The correlation plane is symmetrical with respect to its center. Locating and extracting this potential zone of correlation means precisely locating its four corners. To illustrate this approach, we propose to apply NL-NZ-JTC on four input planes, each consisting of a target image of Lena, cut so that her face is placed exactly in one of the four corners (Fig. 6a–d) and Lena's face as reference (Fig. 6e).

In the case of the target in Fig. 6a, the distances de_x and de_y (Fig. 7a) between the center of the reference and the center of the same image in the target scene are given by

$$\begin{cases} de_x = w_r/2 + w_c - w_r/2 = w_c \\ de_y = h_r/2 + h_c - h_r/2 = h_c \end{cases} \quad (9)$$

where w_r, h_r and w_c, h_c represent respectively the width and the height of the reference and target images (Fig. 7a). According to Eq. (9), which establish the link between the input and output planes, we deduce the distances ds_x and ds_y between the peaks in the output plane (Fig. 7b), which are given by

$$\begin{cases} ds_x = 2de_x = 2w_c \\ ds_y = 2de_y = 2h_c \end{cases} \quad (10)$$

Given that the correlation plane is symmetric, according to Eq. (10), it is possible to calculate the coordinates of the correlation peak starting from the center of the plane (Fig. 7c). Based on Eq. (10), the coordinates $peak1_x$ and $peak1_y$ of the correlation peak, corresponding to the upper/left corner, are given by

$$\begin{cases} peak1_x = w_p/2 - (ds_x)/2 = w_p/2 - w_c \\ peak1_y = h_p/2 - (ds_y)/2 = h_p/2 - h_c \end{cases} \quad (11)$$

where w_p and h_p represent respectively the width and the height of the input plane. The coordinates of the other corners (2: upper/right, 3: lower/left and 4: lower/right) can be calculated in the

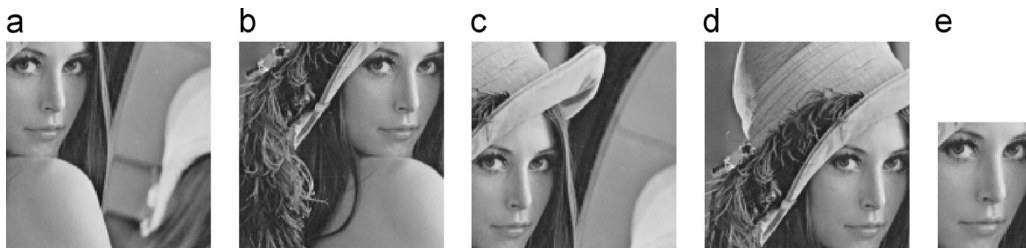


Fig. 6. Example of images that can be used to locate the potential zone of correlation. (a) Target image for the up/left corner, (b) target image for the up/right corner, (c) target image for the down/left corner, (d) target image for the down/right corner and (e) reference image.

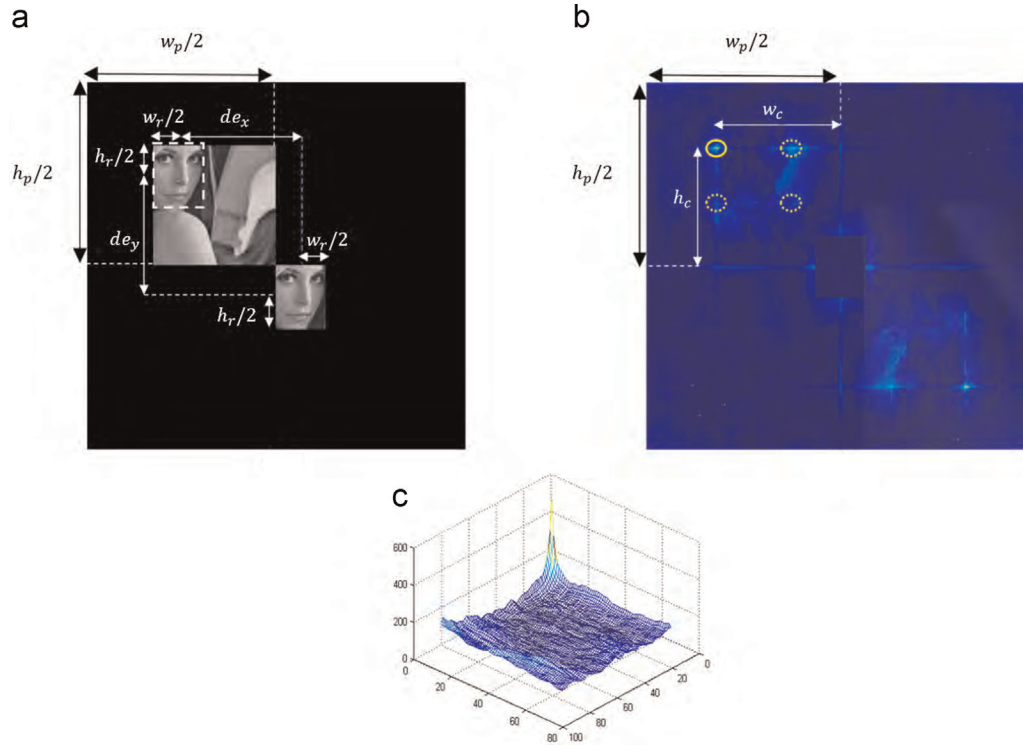


Fig. 7. Localization and extraction of the potential zone of correlation. (a) Input plane for the localization of the up/left corner. (b) Correlation plane where the continuous circle represents the correlation peak derived from the input plane (a) while the other circles represent correlation peaks for the other three input planes based on the targets shown in Fig. 6 (up/right, down/left and down/right corners). A filter is applied to the center of the correlation plane to remove the residual of the autocorrelation peak. (c) Potential zone of correlation extracted from the total correlation plane.

same way:

$$\begin{cases} (peak2_x, peak2_y) = (w_p/2 - w_r, h_p/2 - h_c) \\ (peak3_x, peak3_y) = (w_p/2 - w_c, h_p/2 - h_r) \\ (peak4_x, peak4_y) = (w_p/2 - w_r, h_p/2 - h_r) \end{cases} \quad (12)$$

Thus, PZC_w and PZC_h , the dimensions of the potential zone of correlation, can be calculated by

$$\begin{cases} PZC_x = w_c - w_r \\ PZC_y = h_c - h_r \end{cases} \quad (13)$$

Thanks to this extraction, detecting the object in the target scene involves simply locating the maximum peak of the potential zone of correlation. Indeed, the position of the object in the target image is given by the same coordinates. Finally, this potential zone of correlation extracted from the full correlation plane has the same size as the color scores plane. In addition, it can be seen as a similarity plane in the same way as that resulting from the comparison based on color histogram, the only difference being that the information processed is contour rather than colors. Indeed, in order to perform our dynamic fusion algorithm, the planes need to be normalized so that they share a common value space.

3.4. Planes normalization

After the determination and the extraction of the potential zone of correlation, the two planes have the same size and therefore can be used for our fusion algorithm. However, in order to set the weight given to each method, their dynamics need to be normalized. In particular, the correlation plane is in an interval $[0, V]$ (V : very high value), whereas the color scores plane consists of similarities within an interval $[0, 1]$ given by the value space generated by the Bhattacharyya distance. The planes therefore need to be

normalized to share common dynamics, for example $[0, 1]$.

Several normalization techniques are possible [16] but the most common is still the Z-score technique, which consists of projecting a plane over a space of targeted values by changing its mean (position) and its standard deviation (the scale). For this, the Z-score approach normalizes the value $p(i, j)$ of a data-set p by

$$p_{nor}(i, j) = \frac{p(i, j) - \mu}{\sigma} \quad (14)$$

where μ and σ are the mean and standard deviation of the data set to be normalized, respectively. However, with this method the values of the obtained planes will not be normalized in the interval $[0, 1]$ and a further shortcoming is that the method influences the distribution of the data.

For our fusion approach, we thus propose using the Min–Max technique that normalizes the data in the interval $[0, 1]$ while preserving their original distribution. The normalization of the values $p(i, j)$ is given by

$$p_{nor}(i, j) = \frac{p(i, j) - \min(p)}{\max(p) - \min(p)} \quad (15)$$

3.5. Fusion approaches

Thanks to the previous treatments, the obtained planes have the same size and share common dynamics in the range of $[0, 1]$. Thus, these planes can be considered as probability maps that can be combined by applying a fusion operator [16]. This in turn provides a new plane p_{fus} that synthesizes the information derived from probability maps p_1 and p_2 . The most common operators are

1. Min: $p_{fus}(i, j) = \min(p_1(i, j), p_2(i, j))$
2. Max: $p_{fus}(i, j) = \max(p_1(i, j), p_2(i, j))$
3. Sum: $p_{fus}(i, j) = p_1(i, j) + p_2(i, j)$

$$\begin{aligned}
&4. \text{Product: } p_{fus}(i, j) = p_1(i, j) * p_2(i, j) \\
&5. \text{Mean: } p_{fus}(i, j) = \frac{p_1(i, j) + p_2(i, j)}{2} \\
&6. \text{Weighted summation:} \\
&p_{fus}(i, j) = \alpha * p_1(i, j) + (1 - \alpha) * p_2(i, j)
\end{aligned} \tag{16}$$

where $\alpha \in [0, 1]$ is the weight.

In our fusion approach, we use the latter operator because it can combine the planes by assigning a weight to each one. This can be seen as a confidence factor that can be varied depending on the robustness of each method regarding the reference and target images. Finally, the weighted sum preserves the space values in the range $[0, 1]$, which facilitates the interpretation of the final decision (Section 5.3).

3.6. Accuracy and robustness


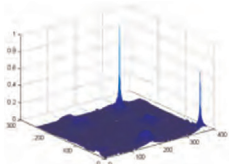
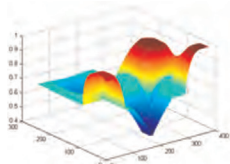
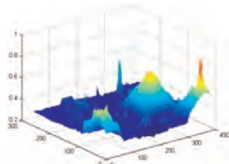

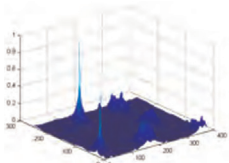
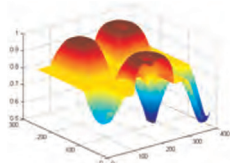
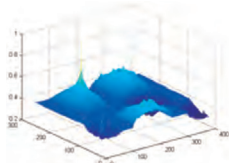

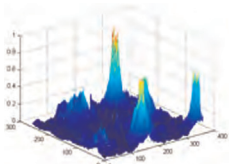
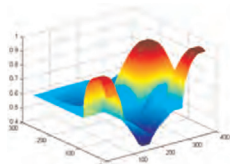
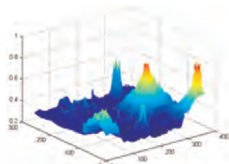
Accuracy and robustness are important concepts for detection and tracking techniques. The first allows the precise localization of the tracked object in the target scene, while the second provides

detection in all scenarios (rotation, noise, occlusions, etc.). Quite often, these two concepts are opposed and it is difficult to combine them with a single method. In fact, for the correlation based on the NL-NZ-JTC approach, only a compromise between these two concepts can be obtained using the non-linearity coefficient k . However, this necessarily depends on the images to be treated. In the case of our swimmer tracking application, the best compromise is obtained with $k \in [0.2, 0.4]$. This provides an accurate localization but is still lacking in robustness in situations of rotation and occlusions (noise in the correlation plane) (Fig. 8a).

Moreover, for the approach based on color histogram, the obtained plane is more homogeneous with a zone of high values at the location of reference/target similarity and negligible noise. Therefore, the robustness of this technique is high, but the accuracy of the localization is insufficient (large area of similarity) (Fig. 8b). Finally, the plane obtained after applying our fusion approach (Eq. (16)) shows a homogeneous plane with greatly attenuated noise and a high peak, thereby ensuring accurate localization of the tracked object. This compromise between robustness and accuracy makes it possible to take advantage of information derived from the two approaches: contour and color (Fig. 8c).

Table 1

Evaluation of the three detection techniques: NL-NZ-JTC, color histograms and dynamic fusion. Three forms are used as references to be detected in the target image in Fig. 9. Under each plane, the height of the peak correspondent to the object to be detected, the nature of the detection (good or false) and the object detected in the case of a false detection.

References	Correlation plane	Color scores map	Fusion plane
	 0.7 – false detection (green circle detected)	 1 – good detection	 0.84 – good detection
	 1 – good detection	 1 – good detection	 1 – good detection
	 0.7 false detection (green circle detected)	 0.9 false detection (red rectangle detected)	 0.85 good detection

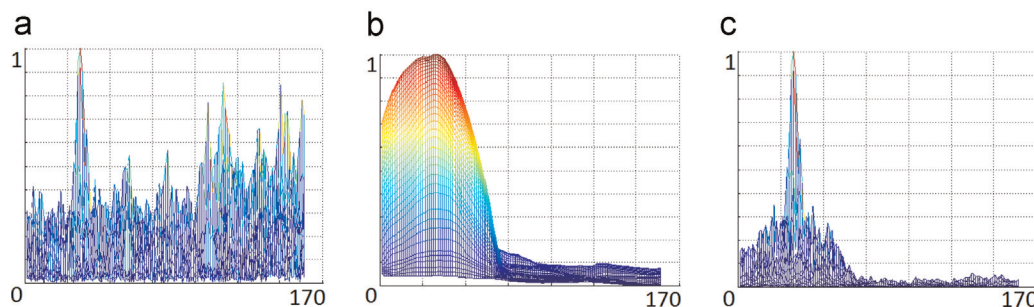


Fig. 8. Fusion of the correlation plane with the color scores map. (a) Correlation plane obtained using the NL-NZ-JTC technique ($k = 0.2$). (b) Color scores map. (c) Plane obtained using our fusion technique.

3.7. Experiments on syntactical images

In order to validate and demonstrate the advantages of our fusion technique, a set of synthetic tests was established. The created target image contains three forms with two different colors, and the reference is a form selected from among them (Fig. 9). The purpose of these experiments is to show the contribution of our dynamic fusion approach as compared with the NL-NZ-JTC and color histogram techniques. For this, Table 1 presents the detection of three forms, a red circle, a green triangle and a red ellipse (deformation of the red circle) in the same target image. To evaluate the relevance of the detection for each one of the three approaches, we look for the shape corresponding to the maximum peak in the output plane (correlation, color or fusion). To measure the level of detection confidence, we look at the height of the maximum peak after normalizing the plane between [0, 1].

The experiments presented in Table 1 show the relevance of our new dynamic fusion approach. In particular, the following should be noted:

1. The correlation shows good results in the case of the green triangle but the red color references give false detections. This can be explained by the importance of the contour (contrast) in the case of green/black as opposed to red/black.
2. The color scores plane is almost identical for the red references (circle, ellipse). This is because this method is based only on color and not on form.
3. The detection result is significantly improved thanks to the dynamic fusion. Note especially that the planes generated by the fusion approach are considerably improved and the false peaks mitigated.
4. The peak in the case of the fusion can be considered as a measure of confidence for the detection (Section 5.3). Indeed, if



Fig. 9. Synthetic target image, contains three forms and two different colors, used to evaluate the pertinence of our dynamic fusion approach. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

the height of the peak is equal to 1, then the correlation and color histogram detected the same object, and the confidence is maximum. However, if the methods are in disagreement, the peaks of the two planes are not superposed, giving a peak of fusion plane below 1, which means lower confidence.

Finally, we have shown the interest of our approach as compared with the NL-NZ-JTC technique and color histogram for detection using synthetic images. In the following section, we will propose improvements to enhance the computing time and the decision. Then we will apply the proposed fusion method to a real case of swimmer tracking.

4. Optimization

4.1. Input plane optimization

The input plane of the correlation methods known in the literature [4,24,23] has generally the standard configuration shown in Fig. 5a. The correlation plane obtained using these classical methods has two potential areas of correlation (top/left and bottom/right) and two areas having almost no information (bottom/left and top/right) (Fig. 5b). In order to improve the computation time, we propose reducing the size of the input plane to its absolute minimum by eliminating as much as possible the areas that are not information-bearing. For this, we propose the reduced input plane, shown in Fig. 10, generated as follows:

$$\begin{cases} c_x = padding \\ c_y = padding \end{cases} \quad (17)$$

$$\begin{cases} r_x = padding + w_c \\ r_y = padding + h_c/2 - h_r/2 \end{cases} \quad (18)$$

The input plane and the new reduced correlation plane have a different size than the plane from the literature presented in Fig. 5. Indeed, eliminating areas that are not information-bearing reduces the size of the plane, which becomes

$$\begin{cases} w_p = 2*w_c + 2*padding \\ h_p = h_c + 2*padding \end{cases} \quad (19)$$

The gain in the size of the planes obtained by this optimization method compared with those of the literature is given by

$$gain = \frac{h_c}{2*h_c + 2*padding} \quad (20)$$

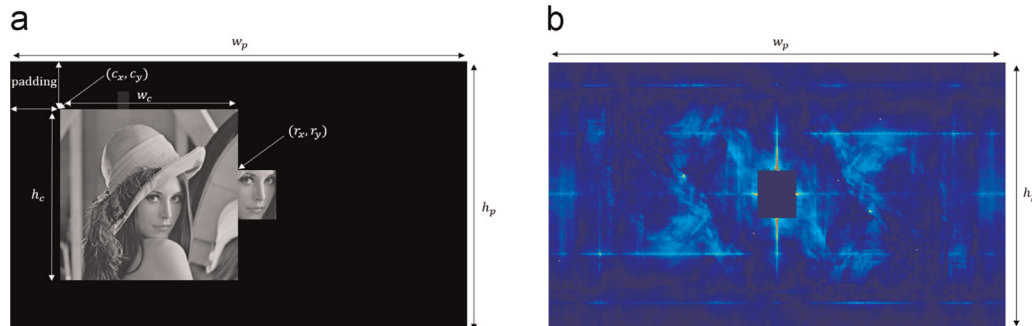


Fig. 10. Fusion of the correlation plane with the color scores map. (a) Correlation plane obtained using the NL-NZ-JTC technique ($k = 0.2$). (b) Color scores map. (c) Plane obtained using our fusion technique.

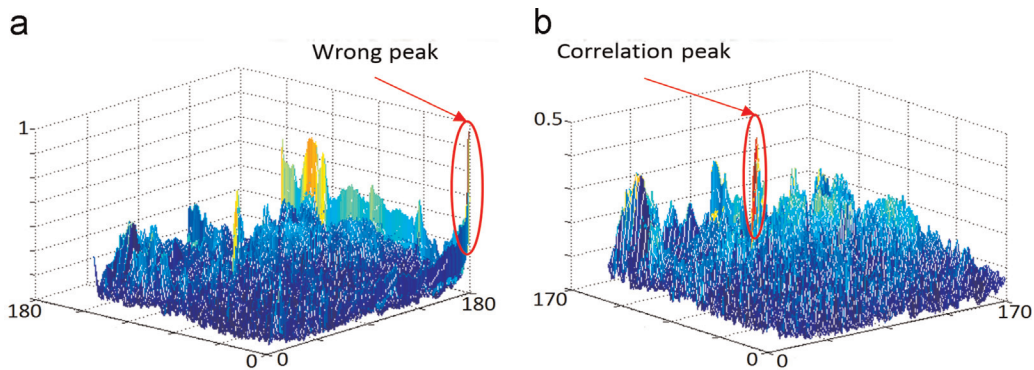


Fig. 11. Removing the region corresponding to edges correlation. (a) Potential zone where the highest peak corresponds to the correlation of edges (reference/target). (b) Restricted potential zone of correlation after removing the artifacts where the highest peak corresponds to the real peak of correlation.

Thus, our proposed technique of removing the areas that are not information-bearing both improves the computation time and limits the unusable areas of the correlation plane.

4.2. Removing the artifacts of edges correlation

In the case of low similarity between the reference and target images (noise, occlusion, rotation, etc.), the position of the expected correlation peak is often wrong (Fig. 11a). Indeed, in this case the maximum energy peak is frequently found on the edges limiting the potential zone of correlation. After an extensive study of this phenomenon, we noted that a zone of strong correlation appeared around the potential zone of correlation, particularly in its corners. This artifact is due to the correlation of the edges of the reference and target images. To avoid these errors, we propose to restrict the potential zone of correlation by retrieving a few pixels from the circumference of the zone (5 pixels in our implementation) in order to remove the area representing the correlation between the edges limiting the two images (Fig. 11b). However, this is only possible if it is assumed that the detected object is not

located on any of the edges of the target image. To take this into account, we generate an input plane where the target image has a larger size (in our case 5 pixels) to ensure that the object to be detected cannot be located on the edges of the target image.

4.3. Reducing the blur effect by the Sobel filter

Correlation techniques, especially the NL-NZ-JTC, are very sensitive to any deformation in the contour of the object to be detected. In our case, we had to deal with the blurred images extracted from the video, which greatly hindered detection and tracking. To reduce this effect, we applied contour enhancement to the reference and target images of the input plane as a pretreatment, using the Sobel filter [22].

As shown in Fig. 12a and b, where motion blur has affected the target image of the input plane, the original correlation plane is very noisy and does not allow the detection and localization of the reference in the target scene. In contrast, after applying the Sobel filtering method, the correlation plane is significantly enhanced and provides a correct localization of the searched object (Fig. 12c,

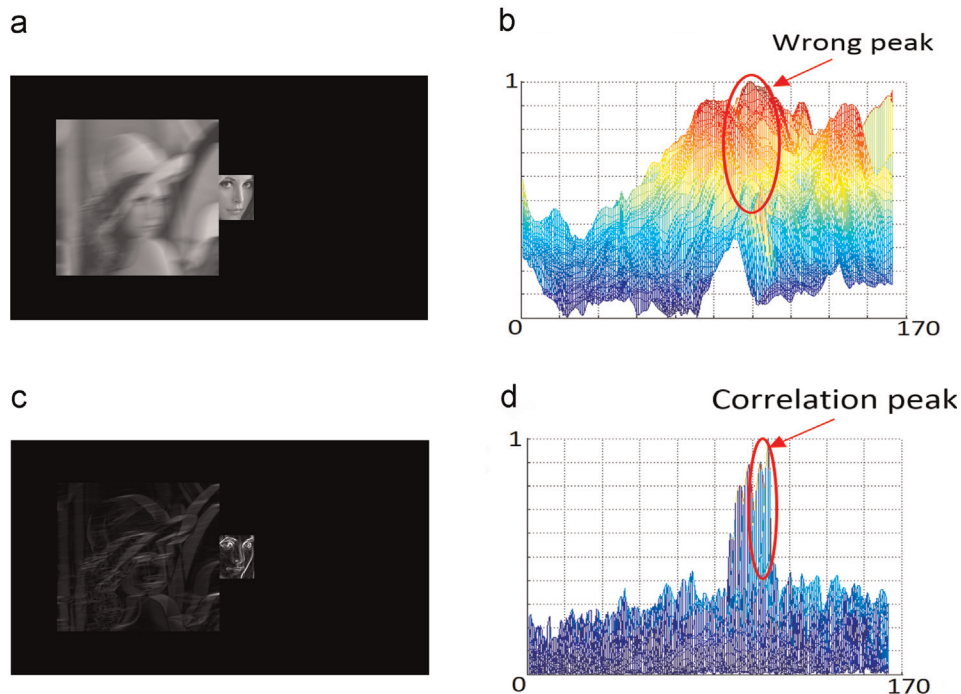


Fig. 12. Limitation of the blur effect by applying the Sobel filter. (a) Input plane with a target image blurred by a motion blur. (b) Very noisy correlation plane derived from the input plane (a) where the localization is false. (c) Filtered input plane using Sobel filter. (d) Denoised correlation plane derived from the input plane (c), where the localization is correct.

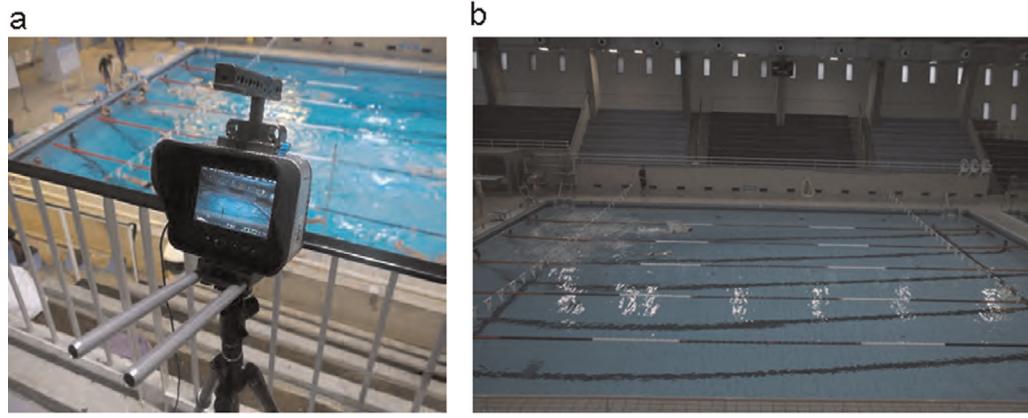


Fig. 13. Shooting environment. (a) The camera Blackmagic 4K used for shooting the videos of the test. (b) Example of a frame extracted from a video used for the tests.

d). Using the Sobel filter enhances the contours blurred by the movement of the object to be detected, and thereby provides more precise object localization in the noisy input planes.

5. Application on swimmer tracking

5.1. Swimmer tracking by dynamic fusion

As we have shown in the preceding sections, the use of only one technique – i.e. the NL-NZ-JTC approach or color histogram technique – is not sufficient to track swimmers. Although these methods are widely used to detect and locate an object in a target scene image, each of them has limitations in uncontrolled environments. In particular, occlusion and contour deformation problems (due to splashing and rotations) render swimmer tracking very difficult using the classical methods. We therefore propose our dynamic fusion technique to track swimmers robustly and accurately. The proposed system includes the following steps:

1. Calibration of the pool using the direct linear transformation (DLT) method [7], which matches the metric coordinates and each pixel in the image; this step is also necessary for the prediction step (step 3).
2. Manual selection of the reference to be tracked.
3. Prediction of the trajectory and the next location of the swimmer to improve the tracking. This will prevent the total loss of the tracked object and restrict the zone of interest around the predicted location.
4. Detection and localization of the tracked object in the zone of interest using the dynamic fusion method.
5. Updating the reference image with the new detected target.
6. Repetition of steps 3–5 until the last frame of the video.

5.2. Database and experimental protocol

To evaluate our dynamic fusion technique, we will compare it with the classical techniques (i.e. the NL-NZ-JTC and color histogram) on two video sequences of junior swimmers from the Brest nautical club (Brest city – France). The first video features the crawl and the second the backstroke. The videos are captured with a 4K resolution (3840×2160 pixels) in 25 frames per second for a period of 15 s each. The shooting environment and an example of a frame extracted from a video taken during training sessions are shown in Fig. 13.

The parameters used in the experiments are as follows:

- Size of the target:

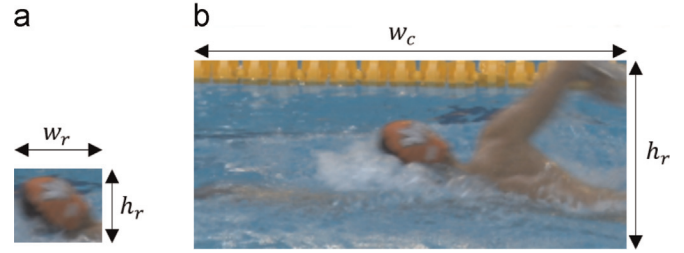


Fig. 14. Determination of the region of interest (target). (a) Reference image. (b) Target image

$$\begin{cases} w_c = 5 \cdot w_r \\ h_c = 3 \cdot h_r \end{cases} \quad (21)$$

This choice is explained by the fact that the swimmer moves horizontally in the lane and moves slightly in the vertical axis (Fig. 14).

- Value of the non-linearity degree of the NL-NZ-JTC method: $k=0.2$ This provides a good compromise between the noise and the sharpness of the peaks in the correlation plane.
- Number of color histogram bins: $Nb_{bin}=16$
- Weight factor for the weighted summation fusion: $\alpha=0.4$ (Eq. (16)) We made this choice because we have somewhat more confidence in the color histogram technique. It reduces the noise in the correlation plane and thus has a less noisy fusion plane.
- Number of pixels estimated for the correlation artifacts of the edges: 5 pixels. These pixels are removed all around the potential zone of correlation as well as on the color scores map in order to ensure the same size of the planes so that they can be merged.

5.3. Evaluation measures

To evaluate the performance of the swimmer tracking system based on our new dynamic fusion approach, we compare it with the two other systems in the literature. The first is based on correlations (NL-NZ-JTC), whereas the second uses the color histogram technique. To evaluate the different methods quantitatively, we propose the following three measures:

1. Tracking percentage.
2. Peak-To-Correlation energy (PCE).
3. Local standard deviation (Local-STD).

Tracking percentage is obtained by calculating the Euclidean

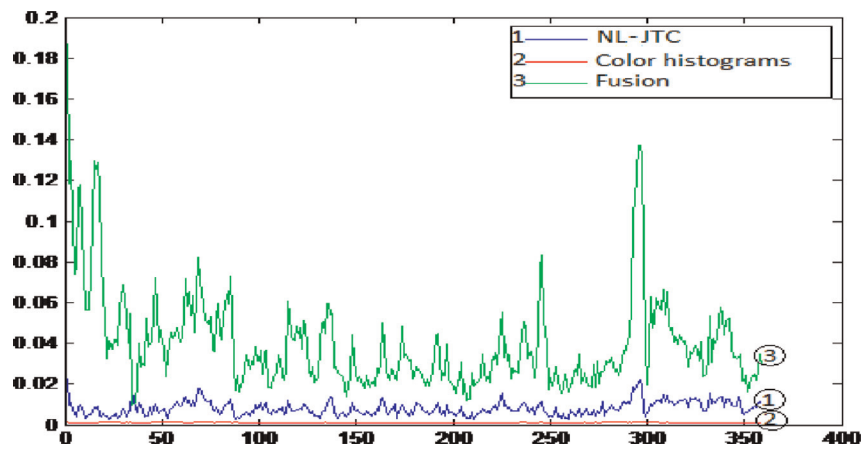


Fig. 15. Comparison between the NL-NZ-JTC correlator, color histograms and dynamic fusion in terms of PCE (tested on a video sequence (360 frames) of a backstroke competition). A high value of the PCE criterion implies a greater confidence towards the localization (sharp peak).

Table 2

Summary table of the comparisons between NL-NZ-JTC correlator, color histograms and the dynamic fusion technique in terms of tracking percentage, PCE and local standard deviation (Local-STD).

Comparison criteria		NL-NZ-JTC	Color histograms	Dynamic fusion
Backstroke video	Tracking percentage (%)	68.73	60	73.09
	PCE (mean) (*10e–2)	2	0.2	6
	Local-STD (mean) (*10e–2)	4.97	0.86	1.13
Crawl video	Tracking percentage (%)	54.17	57.22	71.74
	PCE (mean) (*10e–2)	2	0.2	4.6
	Local-STD (mean) (*10e–2)	5.13	1.44	1.88

distance between the coordinates given by each of the three tracking systems and the ground truth, which is established beforehand by manually annotating the swimmer location in each image of the videos.

PCE is used to measure the ratio of the peak energy to noise in the plane. This is given by the following formula:

$$PCE = \frac{\sum_{i,j}^{n,m} E(i,j)}{\sum_{i,j}^{w_p, h_p} E(i,j)} \quad (22)$$

where n and m represent the size of the area around the peak, while w_p and h_p represent the entire plane. Finally, E is the energy of a point in the plane.

The third criterion, which is the Local-STD, measures the noise level locally in the plane to estimate the accuracy and robustness

of the method in terms of localization (Section 3.6). The Local-STD is given by the following formula:

$$Local - STD = \frac{1}{L} \sum_{l=1}^L \sqrt{\frac{1}{n*m} \sum_{i=1}^{n,m} x_l(i,j) - \bar{x}} \quad (23)$$

where L represents the number of blocks and the size of each block is $(n*m)$, while $x_l(i,j)$ represents the height of the point with coordinates (i,j) in block l and \bar{x} is the mean of the values of the block.

Moreover, in the case of the dynamic fusion technique, the height of the fusion plane peak (Fig. 8c) can be used as a confidence factor. Because the plane has been normalized (Section 3.4), the values are included in the interval of $[0,1]$. Thus, if the fusion peak is near 1, then the two peaks of the two classical

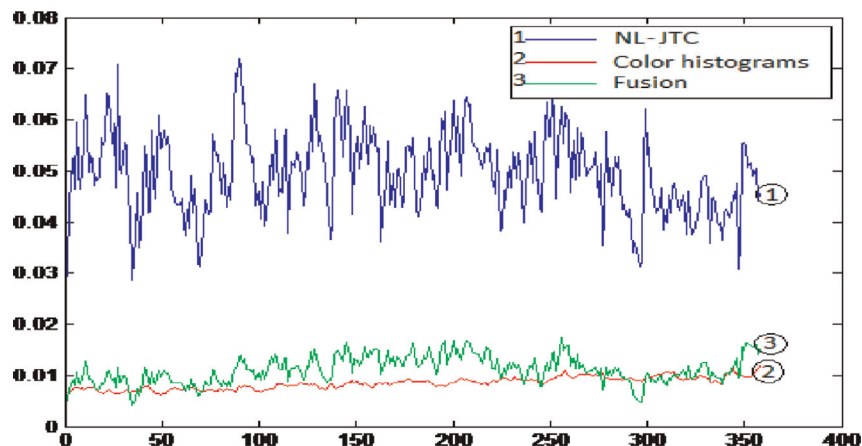


Fig. 16. Comparison between the NL-NZ-JTC correlator, color histograms and dynamic fusion in terms of Local-STD (tested on a video sequence (360 frames) of a backstroke competition). A low value of the Local-STD criterion implies a greater confidence towards the detection (less noisy plane).

techniques are located exactly at the same place, which means that they have detected the same object and therefore have very high confidence. Whereas, if the peaks are in different locations, the fusion peak is less than 1 and the confidence is lower. In our tests, we cannot use this criterion because it does not exist for the two classical methods and our aim is to compare the three methods. However, we use it as a validation criterion for the tracking in every iteration by comparing it with a threshold fixed empirically.

5.4. Experiments and results

Figs. 15 and 16 show the comparisons between the curves of the PCE and the Local-STD for the dynamic fusion method, the NL-NZ-JTC and the color histogram, respectively.

The curves in Fig. 15 show that the proposed technique significantly improves the PCE criterion. This is because the fusion plane often contains a sharper peak and less noise, so the decision is accurate and robust. However, the correlation approach often yields a noisy plane and is therefore less robust, as shown in the PCE curves. For the latter technique, based on color, the resulting PCE is very low because of the large width of the peaks, which yields good detection but an imprecise localization.

The curves in Fig. 16 show that the dynamic fusion and the color histogram technique generate less noisy planes compared with the correlation method. Their decision is more reliable and generally yields good detection.

Table 2 summarizes the results of the tests performed on the two videos presented above.

Table 2 clearly shows that the fusion significantly improves the tracking percentage, as well as the PCE and the Local-STD. For the first video (backstroke), the tracking percentage ranged from 68.73% for the NL-NZ-JTC and 60% for the color histogram to 73.09% for the dynamic fusion. The same was observed for the second video (crawl), with the tracking percentage ranging from 54.17% for the NL-NZ-JTC correlator and 57.22% for the color histogram to 71.74% for the dynamic fusion. The superior percentage for the fusion technique is due to the combined advantages of the two classical techniques to provide a more robust description.

In addition, the PCE values are improved, which reflects sharper peaks and more accurate tracking. The value of the Local-STD is enhanced, which implies noise attenuation for the dynamic fusion technique compared with the correlation, although the color scores map remains slightly less noisy than the fusion plane. It is important to note that cases judged as “false tracking” are often related to the invisibility of the tracked object in the target scene (swimmer's head completely under water). These experiments have shown that our dynamic fusion method has not only improved the tracking percentage but also the confidence toward the decision.

6. Conclusion

In this paper, a new tracking approach based on the fusion of an optical technique and a numerical technique (NL-NZ-JTC and color histogram) is proposed. The main advantage of this new method is the fusion of two descriptors: one based on contour information and the other based on color information. Moreover, the description of the tracked object is richer and can significantly improve the performance of the tracking algorithm.

We also present our contributions to optimize the tracking system. In particular, we propose reducing the input plane in comparison with the standard plane, extracting a sub-plane corresponding to a potential zone of correlation, eliminating the artifacts of the contour correlation, limiting the blur effect by

applying the Sobel filter, and using integral images to accelerate the computation time for the color histogram technique. We applied our dynamic fusion approach to real video sequences to validate our contributions to swimmer tracking. Finally, a comparative study is presented to show the interest of our approach compared with the two classical techniques.

References

- [1] A. Adam, E. Rivlin, I. Shimshoni, Robust fragments-based tracking using the integral histogram, in: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2006, pp. 798–805.
- [2] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **12** (2006) 2037–2041.
- [3] A. Alfalou, N. Ben Haj Yahia, M.S. Alam, Face recognition using a non-zero-order correlation plane and a nonlinear joint transform correlator, in: Proceedings of SPIE, vol. 8498, 2012, p. 84981A.
- [4] A. Alfalou, C. Brosseau, Understanding correlation techniques for face recognition: from basics to applications, in: Face Recognition book, INTECH, 2010, pp. 354–380.
- [5] A. Alfalou, C. Brosseau, W. Kaddah, Optimization of decision making for face recognition based on nonlinear correlation plane, *Opt. Commun.* (2015) 22–27.
- [6] E. Arnaud, E. Memin, B. Cernuschi-Frias, Conditional filters for image sequence based tracking—application to point tracking, *IEEE Trans. Image Process.* (1) (2005).
- [7] A.Y.I. Aziz, H.M. Karara, Direct linear transformation into object space coordinates in close-range photogrammetry, in: Proceedings of the Symposium on Close-Range Photogrammetry, Urbana, IL, 1971, pp. 1–18.
- [8] S. Birchfield, Elliptical head tracking using intensity gradients and color histograms, in: 1998 IEEE Computer Society Conference on Proceedings of Computer Vision and Pattern Recognition, 1998, pp. 232–237.
- [9] S. Birchfield, S. Rangarajan, Spatiograms versus histograms for region-based tracking, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, 2005, pp. 1158–1163.
- [10] W. Chuan-xu, L. Zuo-yong, A new face tracking algorithm based on local binary pattern and skin color information, in: Computer Science and Computational Technology, 2008. ISCCT '08. International Symposium on, 2008, pp. 657–660.
- [11] D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking, *IEEE Trans. Pattern Anal. Mach. Intell.* **5** (2003) 564–575.
- [12] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, 2005, pp. 886–893.
- [13] M. Elipot, Application of new techniques of camera calibration and the cinematique analysis of swimming (Ph.D. thesis), Université Paris Descartes, 2010.
- [14] FPN, (<http://www.fpnatation.fr/>), 2015.
- [15] N. Gagnanashvili, Chi-square tests for comparing weighted histograms, *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrom. Detect. Assoc. Equip.* **2** (2010) 287–296.
- [16] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M.K. Khan, K.O. Sentosa, Performance evaluation of score level fusion in multimodal biometric systems, *Pattern Recognit.* **5** (2010) 1789–1800.
- [17] A. Jacquot, P. Sturm, O. Ruch, Suivi adaptatif d'objets non rigides basé sur des histogrammes de couleur et une sélection automatique de paramètres, in: Journées ORASIS, Fournol, France, 2005.
- [18] G. Jaffre, A. Crouzil, Non-rigid object localization from color model using mean shift (regular paper), in: International Conference on Image Processing (ICIP), Barcelone, Espagne, 2003, pp. 317–320.
- [19] B. Javidi, Nonlinear joint power spectrum based optical correlation, *Appl. Opt.* **12** (1989) 2358–2367.
- [20] S. Jenicka, A. Suruliandi, Empirical evaluation of distance measures for supervised classification of remotely sensed image with modified multivariate local binary pattern, in: 2011 International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), 2011, pp. 762–767.
- [21] Z. Jiang, D.Q. Huynh, W. Moran, S. Challa, N. Spadaccini, Multiple pedestrian tracking using colour and motion models, in: DICTA, 2010, pp. 328–334.
- [22] N. Kanopoulos, N. Vasanthavada, R.L. Baker, Design of an image edge detection filter using the sobel operator, *IEEE J. Solid-State Circuits* **2** (1988) 358–367.
- [23] P. Katz, M. Aron, A. Alfalou, Joint transform correlation for face tracking: elderly fall detection application, in: Proceedings of SPIE, vol. 8748, 2013, p. 87480L.
- [24] I. Leonard, A. Alfalou, M.S. Alam, Nonlinear fringe-adjusted jtc-based face tracking using an adaptive decision criterion, in: Proceedings of SPIE, 2011, 81340M–81340M-6, (<http://spie.org/Publications/Proceedings/Paper/10.1117/12.905266>).
- [25] C.-T. Li, S. Yin, F.T.S. Yu, Nonzero-order joint transform correlator, *Opt. Eng.* **1** (1998) 58–65.
- [26] J. Liu, P. Carr, R.T. Collins, Y. Liu, Tracking sports players with context-conditioned motion models, in: CVPR, IEEE, Portland, Oregon, 2013, pp. 1830–183.
- [27] W. Lu, Y.-P. Tan, A color histogram based people tracking system, in: The 2001 IEEE International Symposium on Circuits and Systems, 2001, ISCAS 2001, vol.

- 2, IEEE, Sydney, NSW, 2001, pp. 137–140.
- [28] A. Miller, A. Basharat, B. White, J. Liu, M. Shah, Person and vehicle tracking in surveillance video, in: *Multimodal Technologies for Perception of Humans*, Springer Berlin Heidelberg, Baltimore, USA, 2008, pp. 174–178.
- [29] E. Pogalin, A.H.C. Thean, J. Baan, N. Schipper, A.W.M. Smeulders, *Video-based training registration for swimmers*, *Int. J. Comput. Sci. Sport* 6 (1) (2007).
- [30] F. Puel, J. Morlier, M. Avalos, M. Mesnard, M. Cid, P. Hellard, *3d kinematic and dynamic analysis of the front crawl tumble turn in elite male swimmers*, *J. Biomech.* 3 (2012) 510–515.
- [31] M. Sabet, R. Zoroofi, K. Niiat, M. Sabbaghian, Automated face tracking with self correction capability, in: *2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, 2011, pp. 280–284.
- [32] P. Viola, M. Jones, Robust real-time face detection, *Int. J. Comput. Vis.* (2004) 137–154.
- [33] Y. Yang, Z. Wang, D. Sun, M. Zhang, N. Cheng, Automatic object tracking using edge orientation histogram based camshift, in: *2010 Third International Conference on Information and Computing (ICIC)*, 2010, pp. 231–234.
- [34] Y. Yin, H. Man, S. Desai, H. He, Multiple human detection and tracking by using multiple-stage hog detector and pfgpdm, in: *Proceedings of SPIE 2009*, pp. 74800S–74800S-10, (<http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1341839>).

All-Automatic Swimmer Tracking System Based on an Optimized Scaled Composite JTC Technique

D. Benarab^{1,2}, T. Napoléon¹, A. Alfalou^{*1}, A. Verney², P. Hellard³

¹ISEN Brest, Département Vision, L@bISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France.

²ACTRIS Brest, 24 Rue Victor Grignard, 29490 Guipavas, France

³FFN (French swimming federation), 14 Rue Scandicci 93508 Pantin Cedex, France

ABSTRACT

In this paper, an all-automatic optimized JTC based swimmer tracking system is proposed and evaluated on real video database outcome from national and international swimming competitions (French National Championship, Limoges 2015, FINA World Championships, Barcelona 2013 and Kazan 2015). First, we proposed to calibrate the swimming pool using the DLT algorithm (Direct Linear Transformation). DLT calculates the homography matrix given a sufficient set of correspondence points between pixels and metric coordinates : i.e. DLT takes into account the dimensions of the swimming pool and the type of the swim. Once the swimming pool is calibrated, we extract the lane. Then we apply a motion detection approach to detect globally the swimmer in this lane. Next, we apply our optimized Scaled Composite JTC which consists of creating an adapted input plane that contains the predicted region and the head reference image. This latter is generated using a composite filter of n images chosen from the database. The dimension of this reference will be scaled according to the ratio between the head's dimension and the width of the swimming lane. Finally, applying the proposed approach improves the performances of our previous tracking method by adding a detection module in order to achieve an all-automatic swimmer tracking system.

Keywords: Scaled Composite JTC, background subtraction, frame differencing, motion detection, all-automatic, swimmer tracking, DLT

1 INTRODUCTION

The evaluation of swimming strategy is an important task to accompany amateur and high-level swimmers in order to enhance their performance. Various studies were conducted in this context¹⁻³ to analyze the swimmer's strategy in terms of diving period, swimming cycles, breathing cycles, speed, etc. In our study, we are interested in estimating the swimmer's instantaneous speed. Most of the actual systems are based on inertial sensors attached to the swimmer in order to measure the instantaneous speed. In fact, these systems are extremely accurate however they don't allow the athlete to perform his best performance. In addition, they are not allowed during competitions. Other range of systems consists of annotating the swimmers in video sequences.

Our aim is to develop an all-automatic swimmer tracking system which provides his positions in all the frames of the video sequence. These positions will be used later on to estimate the swimmer's instantaneous speed. Benarab et al. proposed a multi-tracking system⁴ where each track is based

*. Corresponding author: ayman.al-falou@isen.fr

independently on NL-JTC,⁵ HOG,⁶ LBP⁷ or color histogram.⁸ This proposed system showed better performance than the classical tracking systems that are based on only one tracking technique. However, the decision corresponds to the best detected target among the tracks according to a neutral criterion. Consequently, in each iteration only one technique is taken into account. In order to optimize this system they proposed dynamic fusion approach¹ which consists of enriching the description of the object to be tracked by fusing its contour and color information. These systems are initialized by a reference image of the swimmer's head. This latter is selected manually in the frame of his first appearance after diving, which is the main drawback of these systems.

Therefore, we propose an all-automatic JTC tracking system which is based on : (1) a special NL-JTC¹ and (2) the principle of composite filter. Using composite filters, one can reduce the number of correlation references required for swimmer tracking because a single composite reference is capable of generalizing several swimmer training images. Hester and Casasent presented the first composite filter, equal-correlation-peak (ECP) synthetic discrimination function (SDF) filters by imposing hard constraints on correlation peaks of training images.⁹ The proposed system is detailed in figure 1. First, we calibrate the swimming pool using the 2D DLT technique (Direct Linear Transformation). This method ensures the passage between pixel and metric coordinates and allows us to extract the swimmer's lane. Next, we localize the swimmer globally in the extracted lane based on the frame differencing technique.¹⁰ Then, we generate a composite reference from a pre-generated database (FINA World Championships, Barcelona 2013 and Kazan 2015). Afterwards, we scale it according to the lane dimensions and apply the NL-JTC to detect the swimmer's head. The detected and validated target will be used as a reference to initialize the swimmer tracking system proposed by Benarab et al.¹ In order to evaluate the proposed swimmer detection module, we tested it on video sequences recorded during crawl competitions of the French National Championship, Limoges 2015.

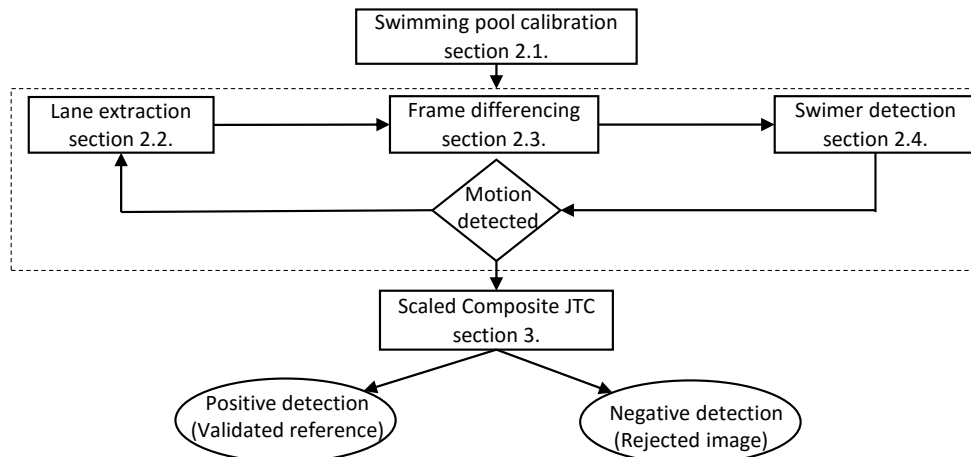


FIGURE 1: Motion detection process. First, the video is calibrated. Next, we extract the corresponding lane and apply the frame differencing technique on it in order to detect the motion. Once the motion of the swimmer is detected, the Scaled Composite JTC is applied to detect the swimmer's head. Finally, we take the detection decision by comparing the output PCE value to a threshold.

In this paper, we start, in section 2, by introducing the proposed motion detection approach and its different stages : 2D DLT calibration technique, lane extraction, frame differencing and global

swimmer detection. In section 3, we introduce the proposed Scaled Composite JTC approach. After, in section 4, we present the experimental protocol and the obtained results. Finally, we conclude with a summary of our contributions and perspectives in section 5.

2 MOTION DETECTION

2.1 Calibration

In the sake of detecting and localizing the swimmer in the pool, it is mandatory to calculate the homography matrix¹¹ that ensures the passage between pixel coordinates in the image and metric coordinates in the real scene. This operation is called calibration. It relies on the prior knowledge of the camera projection parameters which are related to the camera's settings and positioning. Therefore, the Direct Linear Transformation technique (DLT) was proposed.¹¹ In the case of swimming-pool calibration, which is a 2D space, it is sufficient to select four points whose their metric coordinates are known. This allows the calculation of the eight calibration parameters that solve the following standard 2D DLT equation :

$$\begin{cases} u = \frac{L_1x + L_2y + L_3}{L_7x + L_8y + 1} \\ v = \frac{L_4x + L_5y + L_6}{L_7x + L_8y + 1} \end{cases} \quad (1)$$

Where, u and v are the pixel coordinates of a point in the image, x and y are the real metric coordinates of the same point and $L_{(1..8)}$ represent the calibration parameters. The calculation of these parameters requires the knowledge of pixel and metric coordinates of four co-planar points.

2.2 Lane extraction

In order to track the swimmer, we analyze only the used lane (devoted for the chosen swimmer) to avoid the confusion with other swimmers and restrict the region of interest. Knowing that the metric coordinates of the swimming pool are conventional, we calculate the pixel coordinates of the four extremities of the swimmer's lane using the calibration matrix. Firstly, a rectangle includes the lane is cropped from the original image. Then, a mask is applied to extract only the concerned lane (figure 2a) which will be used in the rest of the process.

2.3 Frame differencing

Frame differencing is a motion detection technique used particularly in videos captured by static cameras. Two cases are distinguished, background subtraction and successive frames differencing. In the first technique, we consider the moving object as a foreground which will be segmented after subtracting the frame i from the background image. In our case, the first frame where the lane is empty is considered as a background. This technique shows very good results in case of static background. However, it is extremely sensitive to lighting variation.

Moreover, In our case, the background is not totally static especially when the swimmer generates splashing which creates noise all over the lane after background subtraction. Nevertheless, we noticed an insignificant difference between the successive frame. Therefore, we proposed to apply successive frames differencing in order to detect the swimmer. In the aim of decreasing the noise in the output

image we propose to use Median filter on the two images before the subtraction. The output image is calculated as follows :

$$Output_im(i) = |Median(im(i) - Median(im(i - 1)))| \quad (2)$$

The background is assumed to be the frame $i - 1$. This image would only show some intensity for the pixel locations relative to the swimmer movement as it is shown in figure 2b. A threshold is estimated empirically and applied on the output image in order to improve the subtraction, as shown in figure 2c.

2.4 Global swimmer detection

After the thresholding phase, we propose to apply image dilatation technique in order to get the connected components, where three types of them will be distinguished in the resulting image. The connected components corresponding to the swimming lane lines that will be eliminated by limiting the detection only in the inner lane zone. The small components will be eliminated as well because they are related to the noise. The dense connected components will be localized and considered as a swimmer (figure 2d).

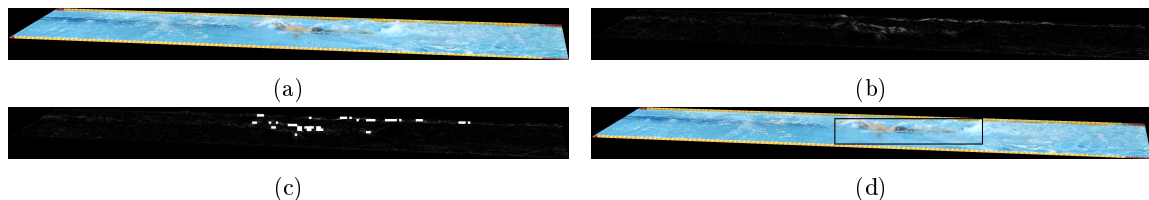


FIGURE 2: Motion detection Process. (a) Calibration and lane extraction (b) Frame differencing (c) Thresholding (d) Swimmer detection. (*French National Championship, Limoges 2015*).

3 SCALED JTC FOR SWIMMER DETECTION

3.1 Scaled composite reference

From a pre-generated database, we choose the n head reference images relevant to our case. Then, an operation of scaling is conducted to predict the dimensions of the swimmer's head in the image. For that, a scaling ratio is calculated between lane's and head's width. Then, the n chosen references are scaled regarding this ratio. Afterwards, we generate the Scaled Composite Reference taking advantage of the composite filter proposed for the JTC approaches.¹² It consists of calculating a weighted summation of the chosen reference images. The Scaled Composite Reference SCR is calculated as shown in the following equation :

$$SCR = \sum \alpha_i * S(ref_i) \quad (3)$$

Where α_i represents the weight that can be used to promote the reference ref_i and S refers to the scaling function.

3.2 NL-JTC (Non-Linear Joint Transform Correlator approach)

The relevant description of the swimmer is the standard form of his head, which looks like a deformed ellipse. Based on this, we chose the NL-JTC approach⁵ in order to localize accurately the position of the swimmer's head. NL-JTC is a correlation based technique which consists of

comparing two images in the aim of detecting, localizing or identify the reference in a target image. This method has proven its efficiency by the numerous publications in the literature.^{1, 4, 5, 12}

To detect the swimmer's head, the input plane of the NL-JTC approach is created using the *SCR* as a reference image coupled with the region of interest detected previously as a target image. Before generating the input plane, we apply Sobel filter on the two images in order to reduce the noise and the blur effect that are related to the quick movements of the swimmer in the water.¹ Once the input plane is created, we apply NL-JTC approach and generate the output correlation plane. The highest peak in this latter corresponds to the potential swimmer's head detection.

3.3 PCE and decision

Our goal is to detect accurately the swimmer's head in order to initialize the optimized swimmer tracking system published by Benarab et al.¹ For this, we apply the proposed swimmer detection approach on the m first frames of the videos. This period is supposed to contain the following events : empty lane, diving, resumption of swimming. During the first event, we apply only the motion detection process. Once the swimmer's movement is detected in the lane, we apply the whole process in order to get a potential swimmer's head after each iteration. At the end, for each potential detection, we associate a PCE value (Peak-to-Correlation Energy) as a confidence coefficient. PCE is used to measure the ratio of the peak energy to noise in the plane. It is given by the following formula :

$$PCE = \frac{\sum_{i,j}^{n,m} E(i,j)}{\sum_{i,j}^{w_p, h_p} E(i,j)} \quad (4)$$

Where n and m represent the size of the area considered as peak, while w_p and h_p represent the entire correlation plane and E is the energy. Finally, the potential detections are sorted by the order of their associated PCE values and the final decision corresponds to the highest PCE.

4 EXPERIMENTAL RESULTS

4.1 Experimental Protocol and databases

In order to evaluate the proposed swimmer detection system, We recorded videos of training (Foch swimming pool, Brest - France), national (French National Championship 2015, Limoges - France) and international championships (World Championship 2013, Barcelona - Spain and 2015, Kazan - Russia). These videos were recorded using a *Blackmagic 4K* camera (3840*2160 pixel, 25fps). Our database of videos contains :

1. The four types of swim.
2. Men and women competitions.
3. Different levels of swimming : Amateur, medium, high-level.

In the other hand, we generated an initial database of swimmers head references which will be used as a training set for the proposed Scaled Composite JTC approach. For this, we cropped swimmers head in various standard situations from videos in our train database. Some examples are shown in the figure 3. Indeed, our database is composed of trained references generated basically from videos of World Championship 2013, Barcelona and 2015, Kazan and test videos taken in the crawl competitions of the French National Championship 2015, Limoges. The train and test videos are chosen from different competitions in order to ensure the reliability of our proposed approach.



FIGURE 3: Examples of seven trained head references. (*World Championship, Barcelona 2013 and Kazan 2015*).

4.2 Experimental results

To evaluate our swimmer detection approach, we tested it on 5 swimmers during crawl competition sequences of 120 frame in the French National Championship, Limoges 2015. The proposed approach is compared to a ground truth that is annotated manually in the frames where the target is visible. In these frames, we measure the distance between detected and annotated targets. Then, the detection percentage is calculated and the results are shown in table 1.

Sequences	Crawl 1			Crawl 2	
Swimmers	Swimmer 1	Swimmer 2	Swimmer 3	Swimmer 1	Swimmer 2
Detection percentage (%)	83.33	86.44	85.71	78.31	82.14

TABLE 1: Detection percentage calculated for 5 swimmers during crawl competitions. (*French National Championship, Limoges 2015*).

Results in table 1 shows the efficiency of the proposed Scaled Composite JTC for swimmer detection. The detection percentage results vary from 78.31% and 86.44% for the tested videos. Our goal is to provide robust references detected automatically to initialize the optimized tracking system proposed by Benarab et al.¹ To achieve this, we need to reduce the false positives (*FP*) to avoid false initialization.

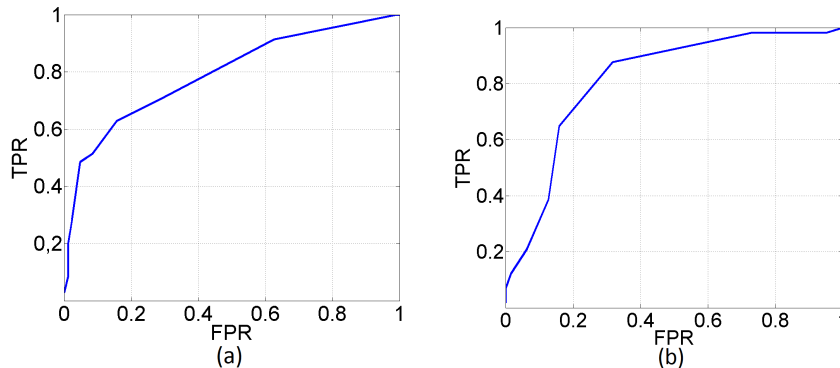


FIGURE 4: ROC curves for the evaluation of the Scaled Composite JTC detection approach applied on : (a) swimmer 3 - crawl 1, (b) swimmer 1 - crawl 2. (*French National Championship, Limoges 2015*).

The ROC curves in figure 4 calculated for the swimmer 3 - crawl 1 and swimmer 1 - crawl 2 show that we can reach 70% of *TP* by tolerating 20% of *FP*, which highlights the proposed approach efficiency, knowing that it is all-automatic. However, to increase the sensitivity ($FPR = 0$), we keep the target corresponds to the highest PCE value which is a true positive in all the sequences

that we tested. the detected and validated references for the five tested swimmers are shown in the figure 5.



FIGURE 5: Examples of two validated references for each tested swimmer. (*French National Championship, Limoges 2015*).

5 CONCLUSION

In this paper, we presented a new all-automatic swimmer detection approach based on Scaled Composite JTC. Coupling this Latter with the tracking system proposed by Benarab et al.¹ will produce a robust all-automatic swimmer tracking system. The proposed approach consists of detecting swimmer motion in the lane using an adapted frame differencing technique. Then, the head of the swimmer is detected accurately using the proposed Scaled Composite JTC technique. Finally, several tests are conducted on five swimmers during crawl competitions in the National French Championships, Limoges 2015. The proposed approach shows promising results in terms of detection percentage as well as its efficiency in providing robust references.

REFERENCES

- [1] Benarab, D., Napoléon, T., Alfalou, A., Verney, A., and Hellard, P., “Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques,” *Optics Communications* **356**, 256–268 (Dec. 2015).
- [2] Elipot, M., Hellard, P., Taar, R., Boissire, E., Rey, J. L., Lecat, S., and Houel, N., “Analysis of swimmers’ velocity during the underwater gliding motion following grab start,” *Journal of Biomechanics* **42**, 1367–1370 (June 2009).
- [3] Puel, F., Morlier, J., Avalos, M., Mesnard, M., Cid, M., and Hellard, P., “3d kinematic and dynamic analysis of the front crawl tumble turn in elite male swimmers,” *Journal of biomechanics* (3), 510–515 (2012).
- [4] Benarab, D., Napoléon, T., Alfalou, A., Verney, A., and Hellard, P., “A novel multitasking system for the evaluation of high-level swimmers performances,” *Proc. SPIE* **9094**, 90940A–90940A–7 (2014).
- [5] Javidi, B., “Nonlinear joint power spectrum based optical correlation,” *Appl. Opt.* (12), 2358–2367 (1989).
- [6] Dalal, N. and Triggs, B., “Histograms of oriented gradients for human detection,” in [*CVPR 2005. IEEE Computer Society Conference on*], 886–893 (2005).
- [7] Ahonen, T., Hadid, A., and Pietikainen, M., “Face description with local binary patterns : Application to face recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions* (12), 2037–2041 (2006).
- [8] Lu, W. and Tan, Y.-P., “A color histogram based people tracking system,” in [*Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*], **2**, 137–140 (2001).
- [9] Hester, C. F. and Casasent, D., “Multivariant technique for multiclass pattern recognition,” *Applied Optics* **19**(11), 1758–1761 (1980).
- [10] Long, Y., Xiao, X., Shu, X., and Chen, S., “Vehicle tracking method using background subtraction and meanshift algorithm,” in [*E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on*], 1–4 (2010).

- [11] Aziz, A. Y. I. and Karara, H. M., “Direct linear transformation into object space coordinates in close-range photogrammetry,” in [*Proc. of the Symposium on Close-Range Photogrammetry*], 1–18 (1971).
- [12] Alfalou, A. and Brosseau, C., [*Face Recognition Book*], ch. Understanding Correlation Techniques for Face Recognition : From Basics to Applications, 354–380, INTECH (2010).

A novel multitasking system for the evaluation of high-level swimmers performances

D. Benarab¹, T. Napoléon¹, A. Alfalou¹, A. Verney², and P. Hellard³

¹ISEN Brest, Département Vision, L@ISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France.

²ACTRIS Brest, 24 Rue Victor Grignard - Guipavas - BP 143 - 29803 Brest Cedex 9, France

³FFN (French swimming federation), 14 Rue Scandicci 93508 Pantin Cedex, France

Abstract

Swimmer tracking has specific difficulties compared to the other tracking systems due to some complex problems such as occlusion (by another person, a wall or splashing), variability of the target (in appearance, lighting or behavior). For the sake of conceiving a robust swimmer tracking system we started by developing mono-tracking systems based on some well-known pattern recognition techniques such as optical correlation (Non-linear JTC (Joint Transform Correlator)) and histogram based approaches (Color histogram, LBP (Local Binary Patterns) and HOG (Histogram of Oriented Gradients)). As an enhancement to these systems, we introduce the aspect of multi-tracking. Its basic idea is to track several potential targets which will give us several tracks then we choose the best track and relaunch the multi-tracking process. Forasmuch each technique has its own periods of good tracking and dropouts (loss of the object to be tracked), we propose a novel heterogeneous multi-tracking system by taking advantage of different tracking techniques. Each track represents an independent mono-tracking system. The use of the proposed heterogeneous multi-tracking system has improved significantly the tracking results from 81.34% to 94%.

Key words: swimmer tracking, multi-tracking, optical correlation, Joint Transform Correlator, histogram, LBP, HOG.

1 Introduction

The automatic tracking is an active field of research in computer vision. Among its applications, we can mention: video surveillance, video conference, recognition of movements... Therefore, it is used in several industries such as biomedical technologies, educational technologies, security or sport. The context of our work outlined in this paper is focused on the latter thematic and more specifically, the automatic tracking of a high-level swimmer in a video sequence. Our goal is to provide the location of the swimmer in each frame which will be used to evaluate his performances by calculating his instantaneous speed, the time of the passage by specific landmarks in the pool and other performances. As we progressed along this specific field of research we highlighted different difficulties such as variation in appearance, movement of the camera, fast movement of the swimmer, splashing and occlusions. In order to overcome these issues, we developed different tracking systems based on some well-known pattern recognition methods such as : non-linear JTC (Joint Transform Correlator),^{1,2} Histogram based approaches (Color histogram, LBP (Local Binary Approaches),³ HOG (Histogram of Oriented Gradient)⁴). Each one of these systems showed good results however it still has its own weakness regarding the difficulties of the swimmer tracking. Therefore, we propose a multi-tracking system which takes advantage of the strengths of the four approaches cited previously to build a robust swimmer tracking system.

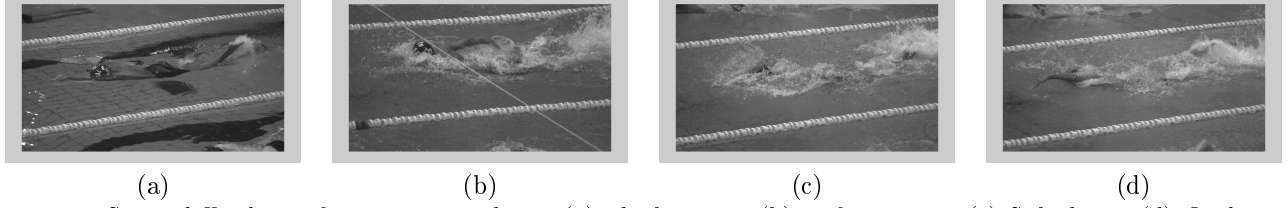


Figure 1: Some difficulties of swimmer tracking: (a) Ideal image. (b) Deformation. (c) Splashing. (d) Occlusion (passage of the hand).

Automatic tracking techniques are generally performed in two steps:

- Detection of the object to be tracked in the first frame.
- Comparison of each frame (i) with the previous ($i - 1$) to track the detected object (with $i \in [2..n]$ where n is the number of frames of the video sequence).

In our swimmer tracking system, the first step is done manually by selecting the region of interest (eventually the head) which will be considered as the reference image that will be tracked in the next frame. Once we get the right target, we will consider it, once again, as a reference and search for it in the next frame, and so on.

In this paper, (i) we introduce different pattern recognition methods that we adapted to fulfill the particularities of swimmer tracking, (ii) we propose a novel multi-tracking system taking advantage of several techniques to get a robust tracking, (iii) we conclude the paper with discussion on experimental results and comparisons between different tracking systems.

2 Swimmer recognition

2.1 Optical correlation based approaches

For the sake of conceiving a robust swimmer tracking system, We propose to adapt several well-known pattern recognition methods, such as the correlation which is an approach that consists in comparing a target image to a reference image (outcome from a pre-recorded data base). The optical correlation has proven its efficiency as evidenced by the numerous publications in the literature.^{1,2,5,6} Indeed, it allows to achieve simultaneously the three necessary operations for the recognition of the targeted object: detection, identification and localization. The optical correlation uses the light and the properties of lenses to realize the recognition. The basic idea is to use two lenses, one for the Fourier transform of the input plane that contains the images to be compared and the other to the inverse Fourier transform to get the correlation plane. Between the two lenses, a spatial light modulator (SLM) is set to retrieve and display the Fourier plane.

For the reason of simplicity we have chosen to implement and adapt this method (originally all optical) on a programmable and reconfigurable numerical method. Thus, we expect to have the flexibility of numerical methods and the approximation of the optical speed.

For the Classical architecture Joint Transform Correlator (Cl-JTC),^{1,6} The correlation plane is obtained by calculating the inverse Fourier transform of the intensity of the joint spectrum (Fourier Transform of the input plane).

$$G_{CL_JTC}(u, v) = |F(u, v)|^2 = |C(u, v)|^2 + |R(u, v)|^2 + |C(u, v)|\exp[\phi_c(u, v)]|R(u, v)|\exp[-\phi_r(u, v) + j(ud + vd)] + |C(u, v)|\exp[-\phi_c(u, v)]|R(u, v)|\exp[\phi_r(u, v) - j(ud + vd)] \quad (1)$$

Where G_{CL_JTC} denotes the joint spectrum intensity of the input plane, F , C and R are the Fourier transform of the input plane, the target and the reference, respectively. In the case of a similarity between the target and the reference, the correlation plane will contain three peaks: one auto-correlation peak that represents the correlation between each image and itself, and two symmetrical peaks of the correlation between the reference and the target.

In the correlation plane, we are interested in the information extracted from the two cross-correlation peaks,

however, the information recovered from the auto-correlation peak is useless (Figure 2). That is where the idea of the architecture Non-Zero JTC (NZ-JTC) came from, which is to remove the auto-correlation peak.^{1,6} The removal of the auto-correlation peak can be performed by subtracting the TF^{-1} of the term $|C(u, v)|^2 + |R(u, v)|^2$ from the equation (1).

$$G_{NZ_JTC}(u, v) = |F(u, v)|^2 - (|C(u, v)|^2 + |R(u, v)|^2) \quad (2)$$

A non-linear thresholding is applied to the the joint spectrum intensity in order to control the width of the peaks according to the need. This architecture is called Non-Linear JTC (NL-JTC).^{1,6}

$$G_{NL_JTC}(u, v) = \alpha |G_{NZ_JTC}(u, v)|^k \quad (3)$$

Where α is a weighting factor and k is the non-linearity degree. If ($k \simeq 0$) then we will get sharp and intense peaks which means a discriminative and less robust system. If ($k \simeq 1$) then we will get large peaks which means a robust and less discriminative system. Between the two, we can vary k according to the need.

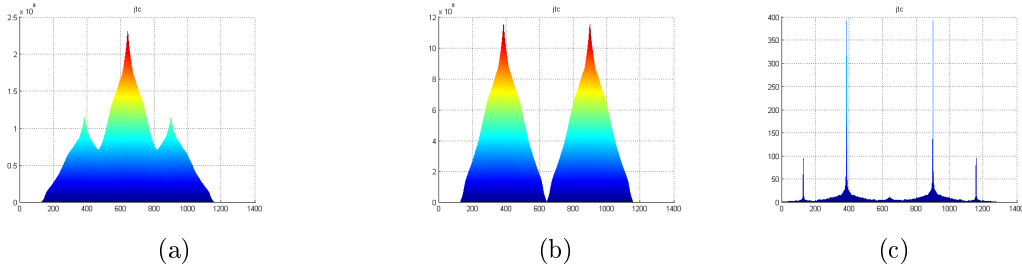


Figure 2: Correlation planes: (a): Classical-JTC. (b): Non-Zero-JTC. (c): Non-Linear-JTC

The choice of the NL-JTC approach for our case comes from its capacity of discrimination and accuracy in localizing the objects (Figure 2). However, this approach is slightly sensitive to occlusions and splashing which may engender some dropouts. To avoid this situation, we propose several histogram based approaches that offer different definitions of the object to be tracked : color, gradient and texture.

2.2 Histogram based approaches

Color histogram based approach is one of the well-known techniques used for tracking.⁷ It uses the color as the main tracking tool. The basic idea is to represent each of the two images, the reference and the target with a histogram of color levels then compare them. The selected target is the one that gives us the closest distance (highest similarity). The color-histogram-based methods are robust with respect to the orientation of the object, but they are sensitive to the variation in lighting conditions.

In the other hand, the appearance of the feature to be tracked can be characterized by the distribution of intensity gradients or contours without the need to know their locations. This approach proposed by⁴ is based on building histograms to store the orientation of the gradients of an image.⁴ The gradient of each pixel is performed by applying a 1-D derivative filter centered in the horizontal and vertical directions. Then, for each pixel, the angle and the magnitude of the gradient are stored in a histogram known as HOG (Histogram of Oriented Gradient). This robust method is mainly based on contours to perform the tracking. However, in the case of swimmer tracking, the distortion due to the splashing will strongly affect the contours and therefore the gradients.

The LBP descriptor describes the texture of the image by calculating the gray level relative neighboring pixels. The particular strength of this descriptor comes from its robustness with respect to the variation of lighting which comes from the local relationship between the pixel and its neighbors. For each pixel, the LBP descriptor is obtained by applying the algorithm described in Figure 3. The final descriptor is an array of LBP

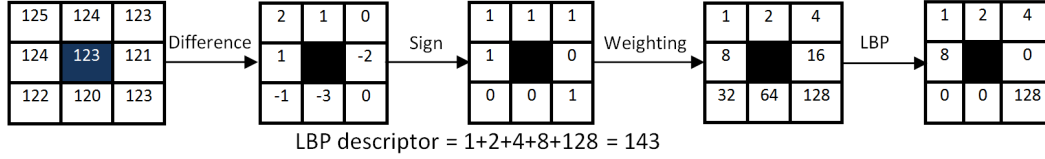


Figure 3: LBP algorithm : Calculate the LBP descriptor of a pixel from its neighbors.

codes which will be represented by a histogram.

Several variants of the LBP have been proposed in the literature.^{3,8} Among the well-known variants we can mention the LBP uniform that overcomes the problem of the sensitivity to the noise and the rotation-invariant LBP to obtain the rotation invariance.

The histogram comparison is made between the two histograms with a specific distance measures such as χ^2 or *Bhattacharyia*.⁹ The tracking process is based on the Mean-Shift algorithm⁷ which allows to define the region of interest in which we compare the histograms of the reference image and the target image.

Each histogram based approach has its own strengths and weakness which leads us to propose a system that benefits from the advantages and avoids the drawbacks by combining several approaches in one multi-tracking system.

3 Multi-tracking

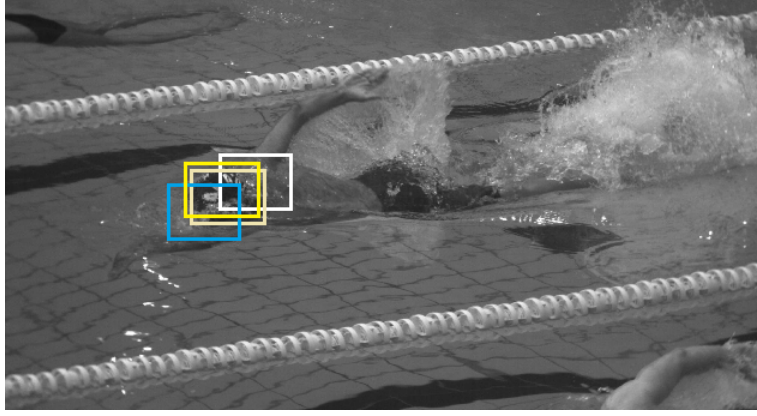
3.1 Swimmer tracking

Our aim is to follow a part of the swimmer body (eventually the head) which must be defined and manually extracted in the first frame of the video sequence or in an appropriately chosen frame. The extracted region is considered as a reference in the frame (i) then we intend to detect it in the frame ($i+1$). To achieve this, we define a restricted region of interest where we will look for the target in the frame ($i+1$) (it should be centered like the reference image and have three times the size of the reference). The reason of defining this region is to avoid the heaviness of computations by searching on the target in the entire frame. Using the NL-JTC approach, we put the two images (the reference and the region of interest of the frame ($i+1$)) into an input plane. In the case where there is a correlation between the two images, two correlation peaks have to be noticed out of the correlation plane. It is theoretically proven that the distance between the two correlation peaks is equal to twice the distance between the correlated points in the images of the input plane. From this equality, it is possible to determine the location of the target in the frame ($i+1$). Using the histogram-based approaches, a distance comparison is performed between the histogram of the reference and the histograms of the possible targets extracted by scanning the restricted region of interest. The chosen target corresponds to the highest similarity. The object detected in the frame ($i+1$) will be considered as a reference for the tracking in the frame ($i+2$).

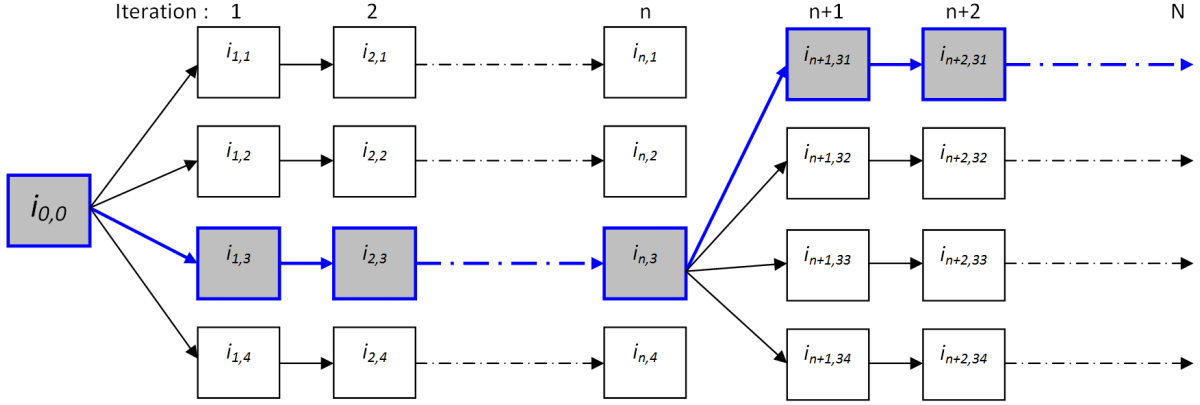
3.2 Multi-tracking systems

In the case of a good similarity between the object in the reference image and in the target image, using the NL-JTC approach will give rise to two high and sharp correlation peaks. Consequently, we can localize accurately the object in the target image. However, in the case of a bad similarity, we will have several peaks around the right peaks and they may have approximately the same height which can be confusing for taking the decision. Based on this observation, we proposed to choose several highest peaks and localize the object in the target image according to them, which will give us several potential detections (Figure 4). The same concept can be applied on the histogram-based approaches by tracking several detections that give the highest similarity in terms of histogram comparison with the reference image.

In order to enhance our multi-tracking system, we propose to take advantage of different tracking techniques (correlation and histogram-based techniques). Each track represents an independent mono-tracking system. After several iterations, we choose the best track to reinitialize the multi-tracking system which enhances the reliability



(a) Multi-tracking (4 tracks).



(b) Architecture of a multi-tracking system (4 tracks). The colored track is the chosen one by performing a color histogram comparison between the targets $i_{n,m}$ and the references $i_{n-1,m}$. Where n denotes the iteration number and m denotes the track number.

Figure 4: Multi-tracking process.

and avoids the situation of a total dropout (loss of the object to be tracked).

3.3 Decision

In a multi-tracking system, the decision of choosing the best track is a crucial step to conceive a robust tracking system. We propose to use the color-histogram comparison as a decision criterion in a homogeneous multi-tracking system based on several NL-JTC tracks or in a heterogeneous multi-tracking system based on NL-JTC, LBP, and HOG.

4 Experiments and results

4.1 Data sets

For the sake of evaluating our swimmer tracking systems, we used the video sequences disposed by the FFN which were taken in different competitions but they were not sufficient in terms of quality. So, we had to move in order to collect the convenient data sets by attending to different regional, national and international competitions (regional championship in Paris, French championship in Rennes, international swimming championships in Barcelona). The videos used in our tests are taken by a high definition camera with the following characteristics : HD (1280*720 pixels), 30 frame/s, Progressive (Not interlaced).

4.2 Comparison between different swimmer tracking systems

To evaluate our swimmer tracking systems, we performed several tests on two video sequences of 150 frames (5s), the first video is used to set the parameters of the tracking systems and the second one is used for the evaluation using the same parameters. In the Figure 5, we compared the distance in pixels between the automatic tracking systems and the ground truth (manual tracking) along a video sequence). The Table 1 shows the performances of the different tracking systems applied on the two video sequences. The comparisons are performed in terms of the percentage of good tracking, the mean and the standard deviation (sdv) of the distances between the automatic tracking systems and the ground truth.

	NL-JTC		LBP		HOG		Color hist		Multi-JTC-NL		Hetero-Multi	
	vid1	vid2	vid1	vid2	vid1	vid2	vid1	vid2	vid1	vid2	vid1	vid2
Mean	25.7	32.3	25.5	24.62	28.63	43.21	24.21	26.06	21.7	23	19.03	20.37
sdv	17.02	23.16	27.63	25.5	17.76	39.02	18.6	19	11.18	12.3	12.98	14.65
Tracking%	88	76.67	81.34	83	84.67	60.67	88.34	86.67	94	93.33	94.67	93.67

Table 1: Comparison between tracking systems and the ground truth. vid1: Video used to set the optimal parameters. vid2: Video used to evaluate the tracking systems

We consider that the distance between 0 and 40 pixels (half of the width of the reference image) are insignificant errors because of the inaccuracy of the ground truth. Beyond that, we are in a state of dropout. The mean and the standard deviation describes generally the distribution of the distances between automatic tracking systems and the ground truth.

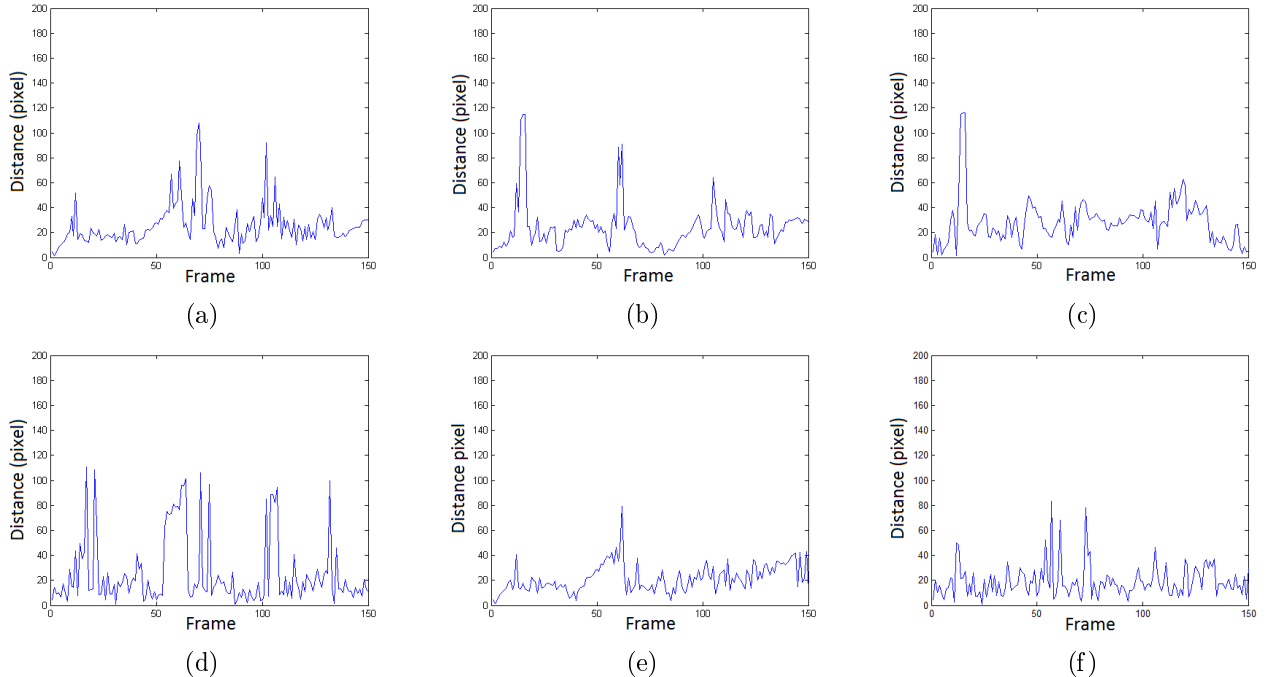


Figure 5: Distance between automatic tracking systems and the ground truth in function of frame number. (a): NL-JTC with $k = 0.2$ (k is the non-linearity coefficient). (b): Color histogram (32 bins). (c): HOG (32 bins). (d): LBP uniform rotation-invariant (32 bins). (e): Multi-tracking system based on NL-JTC (4 tracks), the curve shown in this subfigure corresponds to the best track chosen every five frames by performing a color histogram comparison. (f): Heterogeneous Multi-tracking based on NL-JTC, LBP, HOG and the color histogram is used as a decision criterion.

Like we mentioned in the Section 2 each mono-tracking system has advantages and drawbacks which influences on their performances as shown in the Table 1 and the Figure 5, where each tracking system has its own periods of good tracking and dropouts. This fact leads as to imagine a system that combines the four systems by taking only the best tracking in each frame. For instance, if we take manually the min distance of the four mono-tracking systems then we will get a percentage of 98.67% of good tracking, mean distance: 11.19 pixel and sdv : 7.25 pixel. The homogeneous multitasking system based on NL-JTC that we proposed shows good performances on the two videos (94% and 93.33) compared to the NL-JTC mono-tracking system (88% and 76.67%) which can be interpreted by the fact of checking out four possible detections of NL-JTC and choosing the best one by performing a color-histogram comparison. The enhancement brought by the heterogeneous multi-tracking system (NL-JTC, LBP, and HOG. Decision criterion: Color-histogram comparison) is more interesting (94.67% and 93.33%) because we benefited from the advantages of the three mono-tracking systems, in addition to the use of the Color-histogram comparison as decision criterion. This latter is still a subject of research to find the best criterion for choosing the right track, which will enhance the tracking system significantly.

The tests are performed on the videos described in the subsection 4.1 using Matlab R2012b. The mean duration of the iterations of our proposed system is : 0.93s , which is a subject of an ulterior research study.

5 Conclusion

In this paper, we have adapted several pattern recognition techniques in order to develop a new swimmer tracking systems. Firstly, we proposed several mono-tracking systems based on NL-JTC, color histogram, LBP or HOG. Secondly, a multi-tracking system is proposed to take advantage of the strengths of different tracking techniques. Finally, the results of the tests that we performed show clearly the enhancement brought by the use of the multi-tracking system. The performances of this latter system lays on a crucial step which is finding the best criterion to choose which track we trust the most.,In our work, we used the color histogram comparison as a decision criterion which showed good results nevertheless this step can be enhanced by using a hybrid decision criterion or taking in account the historical detections of each track which may give us an idea of the robustness of this track.

REFERENCES

- [1] Alfalou, A., Ben Haj Yahia, N., Alam, M. S., "face recognition using a non-zero-order correlation plane and a nonlinear joint transform correlator," *Proc. Of SPIE 8498*,, 84981A (2012).
- [2] Katz, P., Aron, M., Alfalou, A., "joint transform correlation for face tracking: elderly fall detection application," *Proc. SPIE 8748*,, 87480I (2013).
- [3] Ahonen, T., Hadid, A., Pietikainen, M., "face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(12), 2037–2041 (2006).
- [4] Dalal, N., Triggs, B., "histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, **1**, 886–893 vol. 1 (2005).
- [5] Alam, M., Wai, C. N., "color pattern recognition using fringe-adjusted joint transform correlation," *National Aerospace and Electronics Conference, Proceedings of the IEEE 2000*,, 427–433 (2000).
- [6] Alfalou, A., Brosseau, C., [*Face Recognition Book*], "Understanding Correlation Techniques for Face Recognition: From Basics to Applications," 354–380, INTECH (2010).
- [7] Ju, M.-Y., Ouyang, C.-S., Chang, H.-S., "mean shift tracking using fuzzy color histogram," *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, **6**, 2904–2908 (2010).
- [8] Fang, Y., Luo, J., Lou, C., "fusion of multi-directional rotation invariant uniform lbp features for face recognition," *Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on*, **2**, 332–335 (2009).
- [9] Jenicka, S., Suruliandi, A., "empirical evaluation of distance measures for supervised classification of remotely sensed image with modified multivariate local binary pattern," *Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on*,, 762–767 (2011).

