Estimation du regard à partir de la vidéo

Afifa Dahmane

▶ **To cite this version:**

Afifa Dahmane. Estimation du regard à partir de la vidéo. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Lille 1, Sciences et Technologies; CRIStAL UMR 9189; Université des Sciences et Technologies Houari Boumediene (USTHB), Alger, 2015. Français. NNT : . tel-01703903

HAL Id: tel-01703903
https://hal.science/tel-01703903

Submitted on 8 Feb 2018

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 des Sciences et Technologies de Lille
Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL)
École Doctorale Sciences pour l'Ingénieur
En cotutelle avec :
Université des Sciences et Technologies Houari Boumediene (USTHB), Alger
Laboratoire de Recherche en Intelligence Artificielle (LRIA)

Thèse

présentée par

Afifa DAHMANE

en vue d'obtenir le grade de

Docteur de l'Université des Sciences et Technologies de Lille
(Discipline : Informatique)

Estimation du regard à partir de la vidéo

Soutenue publiquement le 01/02/2015 devant le jury :

Président :	M. Ahmed GUESSOUM (Professeur)	Université d'Alger USTHB
Rapporteurs :	M. Mohamed Chawki BATOUCHE (Professeur)	Université Constantine 2
	M. François POULET(HDR)	Université Rennes 1
Examinateur :	M. El Mustapha MOUADDIB (Professeur)	Université de Picardie
Directeurs de thèse :	M. Chaabane DJERABA (Professeur)	Université Lille 1
	M. Slimane LARABI (Professeur)	Université d'Alger USTHB

Remerciements

J'exprime mes profonds remerciements à mes directeurs de thèse, le professeur Slimane Larabi et le professeur Chaabane Djeraba pour l'aide compétente qu'ils m'ont apportée. L'œil critique de M. Slimane Larabi et ses conseils éclairés m'ont été très précieux pour structurer le travail. Il a toujours été disponible et à l'écoute de mes nombreuses questions. C'est également grâce à sa collaboration avec le laboratoire LIFL que j'ai eu la chance de travailler avec l'équipe de recherche Fox, ce qui c'est avéré une expérience très enrichissante. Je tiens à remercier M. Chaabane Djeraba pour l'accueil et les conditions de travail privilégiées qu'il m'a offertes. Les nombreuses discussions que nous avons eues ainsi que ses conseils m'ont encouragé à finir ce travail.

Je remercie les professeurs Mohamed Chawki Batouche et François Poulet d'avoir accepté d'être rapporteurs de ma thèse et le professeur El Mustapha Mouaddib d'en être examinateur. Je tiens à remercier le professeur Ahmed Guessoum d'avoir accepté d'être président du jury. J'exprime ma gratitude à tous les membres de mon jury de thèse, qui me font l'honneur de juger ce mémoire.

Mes sincères remerciements s'adressent au Dr. Marius Bilasco. Sa relecture méticuleuse de chacun des chapitres m'a permis de mieux préciser mon propos et d'améliorer la qualité des différentes sections. Ses remarques sont pour beaucoup dans le résultat final de ce travail.

Je remercie tous les thésards et les autres membres du laboratoire LIFL de Lille, notamment ceux avec qui j'ai eu l'occasion de travailler. Entre autres Amel, Rémi, Pierre, Jean, José, Tarek ...

Je remercie également tous les membres du laboratoire LRIA d'Alger pour la bonne ambiance de travail.

Enfin, je voudrais dédier ce mémoire aux membres de ma famille et à mes amis qui m'ont soutenue dans la réalisation de cette thèse. Je tiens à remercier tout particulièrement mes parents.

Résumé

L'objectif de cette thèse est l'estimation de la pose de la tête humaine à partir d'images monoculaires. Le but est d'avoir un retour d'information de l'utilisateur sur l'orientation de son regard, et ce, de manière non-intrusive. Ce domaine de recherche est très actif compte tenu de l'évolution des interfaces de communication entre l'homme et la machine, d'autant plus que plusieurs défis sont toujours ouverts. Notamment, la robustesse du système, son invariabilité à l'identité des personnes et à l'illumination ainsi que la qualité du matériel de capture requis.

L'approche que nous avons proposée est basée sur la symétrie bilatérale du visage. Nous utilisons des caractéristiques extraites de la symétrie pour estimer la pose de la tête par le biais de l'apprentissage. Les caractéristiques utilisées sont géométriques mais extraites de manière globale à partir de toute la texture du visage, sans que des points ou des contours spécifiques ne soient requis.

Ces caractéristiques ont été validées expérimentalement à l'aide de bases d'images et de vidéos publiques dédiées à l'estimation de la pose de la tête. L'apprentissage supervisé ainsi que la régression sont utilisés pour construire des modèles de poses. Ces modèles ont été testés sur des séquences vidéo indépendantes des bases utilisées pour l'apprentissage. L'erreur d'estimation a été calculée et les résultats sont supérieurs ou équivalents à l'état de l'art.

Mots clés : Estimation de la pose de la tête, Détection de la symétrie, Reconnaissance de formes.

Title: Gaze estimation from video

abstract: The aim of this thesis is to estimate the pose of a human head from monocular images. The goal is to have a feedback from the user on the direction of his gaze, and this is done in a non-intrusive manner. This area of research is very active given the evolution of the communication interfaces between a human and a machine, especially as many challenges are still there. In particular, the robustness of the system, its invariance to the identity of individuals and enlightenment.

The approach we propose is based on the bilateral symmetry of the face. We use features extracted from the symmetry to estimate head pose through learning. The features used are geometric but extracted holistically from the whole texture of the face, without the need for specific points or contours.

These features have been experimentally validated by means of publicly available databases of images and videos dedicated to head pose estimation. Supervised learning and regression have been used to build models of poses. These models have been tested on video sequences different from the bases used for learning. The estimation error has been calculated and the results are superior or equivalent to the state of the art.

Keywords: Head pose estimation, Symmetry detection, Pattern recognition.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction

Sommaire

La vision par ordinateur permet l'analyse du comportement humain sans interférences entre les utilisateurs, leur environnement et à faible coût. L'analyse du regard est un des piliers de l'analyse du comportement, elle permet la compréhension approfondie du processus visuel humain. L'analyse du regard tend à redéfinir la relation entre l'utilisateur et l'ordinateur en permettant à ce dernier d'attribuer un sens aux mouvements de l'utilisateur.

L'estimation de la direction du regard s'appuie de manière significative sur l'analyse des mouvements de la tête. Le mouvement de la tête joue un rôle dans la communication entre individus, tels que les hochements de tête, les mouvements brusques ou le détournement du visage. Certains mouvements peuvent traduire les réactions et les intérêts des utilisateurs. Par exemple, si une personne suit un cours de e-learning et qu'elle détourne la tête de son écran, cela souligne son manque de concentration. D'autre part, l'orientation de la tête peut être utilisée afin de rendre des applications de reconnaissance de formes invariables à la pose de la tête.

L'estimation de la pose de la tête est un domaine de recherche très actif, d'autant plus qu'il constitue un pilier de l'analyse du comportement humain et spécialement, de l'analyse du regard. Nous présentons une méthode géométrique d'estimation de la pose de la tête à partir d'images monoculaires. Nous allons démontrer que la taille des régions symétriques du visage est un bon indicateur du mouvement pan (*cf.* figure 1.1) de la tête [DLD10]. L'approche proposée ne nécessite pas la détection de points spécifiques sur le visage ce qui la rend robuste aux occultations partielles ainsi qu'à la résolution des images. Les dimensions de la région du visage qui contient une symétrie bilatérale sont utilisées pour décider de l'orientation de la tête. Nous utilisons un modèle de classification afin de valider l'utilisation des caractéristiques symétriques pour l'estimation de la pose [DLDB12].

1.1 Contexte

L'estimation de l'attention du regard consiste à déterminer le sens du regard d'une personne. L'orientation de la tête contribue, conjointement avec la position des yeux, à

la direction du regard. L'orientation de la tête consiste à déterminer la valeur des angles de la rotation de la tête selon les 3 degrés de liberté : pan, tilt et roll. Le pan et le tilt sont représentés par les mouvements hors plan de la tête autour des axes Y et X respectivement. Le roll est l'angle de rotation de la tête dans le plan image selon l'axe Z (*cf.* figure 1.1).

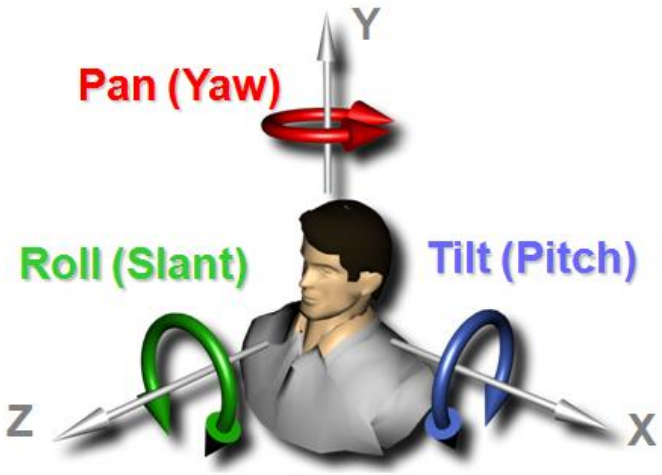

FIGURE 1.1: Les 3 degrés de liberté de l'orientation de la tête.

L'estimation de l'orientation de la tête constitue la première étape d'un système d'analyse du regard. Mais aussi, une étape qui rendrait un système de reconnaissance de forme basé sur le visage, telle que la reconnaissance du genre, de l'expression faciale ou même la reconnaissance du visage lui même, plus performant tout en traitant plus de cas. En effet, la reconnaissance du visage en tenant compte de l'orientation de la tête permet un traitement sous différentes poses, ce qui est plus proche des conditions réelles que d'avoir des visages frontaux seulement.

Afin de pouvoir travailler sur des projets divers de reconnaissance de forme (ce qui se fait actuellement au sein de notre équipe), nous avons besoin de développer un système qui nous renvoie une information sur l'attention visuelle à partir d'une seule image sans poser de contraintes particulières sur la qualité du matériel de capture et sans gêner

l'utilisateur, dont on va estimer la pose, dans sa tâche. L'estimation doit pouvoir se faire à partir de vidéos aussi bien que des images statiques. Ce travail constitue l'objectif de cette thèse.
L'information obtenue sur la pose de la tête nous permet de :

- Avoir un retour sur l'attention visuelle qui nous indique de quelle manière l'utilisateur effectue une tâche précise, comme suivre un cours de e-learning ou regarder une vitrine de magasin. C'est l'analyse du comportement.
- Appliquer d'autres algorithmes de reconnaissance de forme, comme la reconnaissance faciale ou la reconnaissance du genre en tenant compte de la pose de la tête.

1.2 Problématique

Plusieurs paramètres liés aux conditions extérieures de prise d'images et aux personnes, peuvent influencer la robustesse du système d'estimation de l'orientation de la tête. Un système robuste doit prendre en considération la précision des résultats ainsi que la manière d'interagir avec l'utilisateur. Le système doit fonctionner indépendamment de l'identité de la personne ainsi que de la résolution de l'image. En effet, un nouvel utilisateur ne doit pas être prié d'initialiser son visage ou sa pose. Aussi, une seule caméra doit être utilisée et sa distance de l'utilisateur peut être grande comme elle peut être petite. Avec ces pré-requis satisfaits, le système doit être capable d'estimer l'orientation de la tête de plusieurs personnes et en temps réel.

Ces critères qui définissent un système robuste sont cités dans l'état de l'art établi par Murphy et Trivedi [MCT09]. Nous avons défini une approche, selon notre contexte de travail. Notre système se doit d'être non intrusif et ne nécessite pas la collaboration des utilisateurs afin de ne pas les gêner dans leur tâches. Les résultats doivent être indépendants de l'identité de l'utilisateur. La pose d'un nouvel utilisateur inconnu doit pouvoir être estimée sans qu'il fasse partie des données d'apprentissage. Ceci permet au modèle créé avec un ensemble de données d'être déployé sur de nouvelles images ou vidéos.

1.3 Approche proposée

Nous présentons un modèle symétrique dans le but de l'estimation de la pose de la tête. Nous exploitons la symétrie bilatérale du visage qui est définie en utilisant des régions globales de la peau et non des points caractéristiques locales.

L'approche proposée ne nécessite pas la localisation de points spécifiques sur le visage et peut être déployée avec du matériel à faible coût et de grande disponibilité. De plus, ni initialisation du système ni calibration ne sont requises. La pose estimée est grossière mais suffisante pour déduire la direction globale de l'attention visuelle.

Notre travail présente trois contributions majeures (*cf.* figure 1.1) :

- Une méthode pour détecter la position de l'axe de symétrie et son orientation sur une image.
- Estimation de l'angle roll de l'inclinaison de l'axe de symétrie.
- Les angles pan et tilt sont calculés en utilisant la région définie par les pixels symétriques.

Nous proposons une approche qui allie l'efficacité des méthodes locales et globales. Nous sélectionnons la région symétrique du visage relativement à l'intensité des pixels de la peau. Nous utilisons la taille de cette région et son orientation pour estimer la pose.

La méthode proposée (Figure 1.2) commence par détecter le visage en utilisant un algorithme basé sur celui de Viola Jones [VJ01]. Un pré-traitement (égalisation d'histogramme) est appliqué afin de réduire l'influence de l'illumination. Ensuite, l'axe de symétrie est recherché dans la région du visage. Cette tâche est réalisée en utilisant un algorithme de détection de la symétrie. Une fois le visage et son axe de symétrie détectés, nous extrayons les caractéristiques de la symétrie. Nous déduisons l'angle roll à partir de l'orientation de l'axe de symétrie et nous estimons le pan et le tilt en effectuant un apprentissage sur la région symétrique du visage.

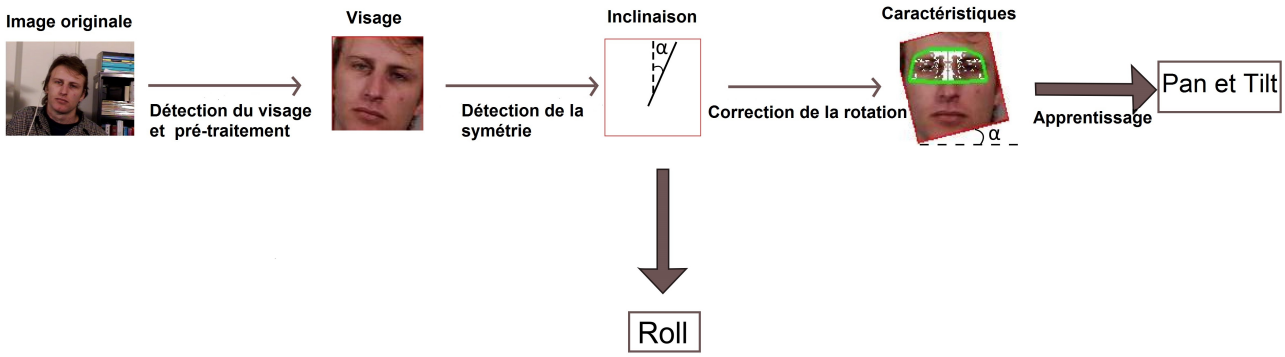

FIGURE 1.2: Approche proposée

1.4 Organisation du document

Dans la suite du document, après avoir introduit le schéma général de l'approche proposée, le Chapitre 2 dresse un état de l'art des méthodes d'estimation de la pose de la tête. Nous présentons ces techniques au travers de deux axes : descripteurs et outils. Nous analysons l'usage des différents descripteurs utilisés pour caractériser la pose sur des images de visages. Ensuite, nous détaillons les différents outils utilisés tels que l'apprentissage et la comparaison de prototypes.

Le Chapitre 3 traite le problème de détection de la symétrie sur des images de visages. Nous présentons un résumé des techniques de détection de la symétrie au travers des deux catégories : symétrie 2D et symétrie 3D. Nous détaillons ensuite notre méthode de détection de la symétrie bilatérale du visage. Nous expliquons le rôle des paramètres utilisés et nous montrons comment appliquer le processus sur un flux vidéo.

Le Chapitre 4 met l'accent sur la corrélation entre la pose de la tête et la symétrie détectée sur l'image de cette dernière. Nous effectuons une étude théorique sur la déformation de la projection du visage sur le plan image relativement à la pose de la tête. Nous définissons par la suite, les caractéristiques qui nous permettent de discriminer les poses de la tête ainsi que la méthodologie suivie pour l'apprentissage de ces caractéristiques.

Le Chapitre 5 évalue notre méthode d'estimation de la pose de la tête. Nous commen-

çons par valider la détection de la symétrie bilatérale du visage sur des images dont nous connaissons l'inclinaison et la position de l'axe de symétrie. Nous évaluons ensuite nos descripteurs de symétrie sur des bases publiques annotées d'images et de vidéos. Nous terminons par comparer nos résultats avec l'état de l'art.

Enfin, nous concluons ce document au Chapitre 6 et nous présentons les perspectives ouvertes par ces travaux.

Chapitre 2

Estimation de l'orientation de la tête : État de l'art

Sommaire

L'estimation de l'orientation de la tête fait partie de l'analyse du comportement humain. Nous mettons l'accent, dans ce chapitre, sur l'estimation de l'attention visuelle. Par la suite, nous présentons un état de l'art des méthodes d'estimation de la pose de la tête existantes regroupées en deux parties. Premièrement, par la manière de caractériser la pose, en utilisant l'apparence ou la forme. Deuxièmement, par la méthode utilisée pour obtenir la pose à partir de ces caractéristiques où plusieurs outils sont utilisés. Parmi eux, nous trouvons l'apprentissage, la comparaison de prototypes et le suivi de mouvement.

2.1 Analyse du comportement humain

Les techniques actuelles permettent d'extrapoler beaucoup de détails à partir de l'information sur le mouvement humain. Ces détails procurent des indices concernant l'action que l'individu effectue. Les caractéristiques des actions accomplies par un utilisateur ou un groupe d'utilisateurs peuvent être utilisées pour analyser le comportement humain.

L'analyse du comportement humain offre la possibilité de doter les ordinateurs d'une capacité d'attribuer un sens à l'attitude des utilisateurs, leurs préférences, leurs relations sociales, etc. L'ordinateur n'a plus le rôle d'observateur mais il serait un participant actif dans plusieurs domaines y compris l'éducation et le divertissement. L'estimation de l'attention visuelle constitue un élément de l'analyse du comportement humain lié aux micro activités effectuées par l'utilisateur en mouvement (*cf.* figure 2.1). Notons qu'une macro activité de l'utilisateur concerne par exemple les déplacements dans l'environnement et l'interaction avec d'autres objets.

L'estimation de l'attention visuelle consiste à déterminer où est-ce qu'une personne regarde sur un plan prédéfini. Il s'agit d'une combinaison de l'orientation de la tête et de l'orientation des yeux. La localisation et le suivi des yeux n'est pas une tâche facile et requiert souvent un matériel dédié tels que les systèmes infrarouge ainsi qu'une étape de calibration.

Les méthodes utilisées pour localiser l'orientation des yeux se divisent en deux catégories [HJ10] :

– Les méthodes basées sur l'apparence de l'œil où des caractéristiques extraites de

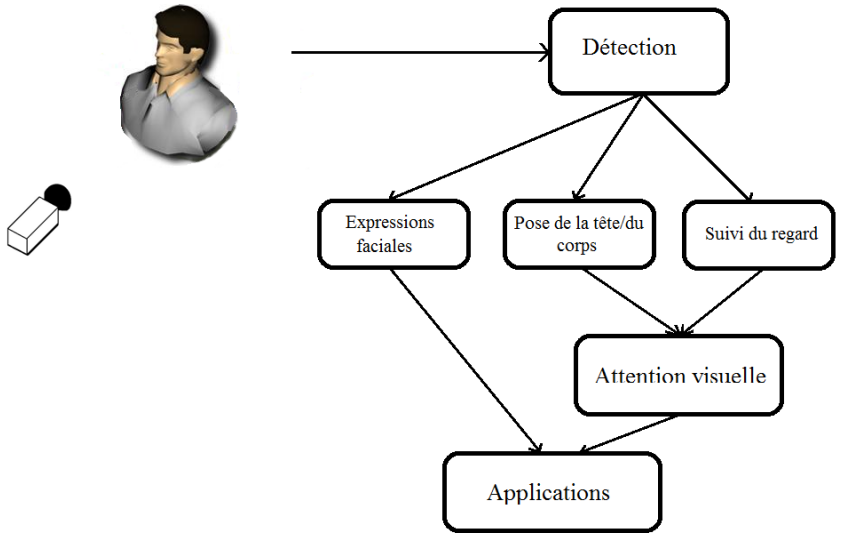

FIGURE 2.1: Analyse du comportement lorsque l'utilisateur effectue de petits mouvements (micro activité).

l'image de l'œil sont reliées aux points d'attentions.

– Les méthodes basées sur un modèle 3D de l'œil où les points d'attention sont déterminés par l'intersection de l'axe visuel et le plan d'attention.

La perception humaine de la direction du regard est fortement influencée par l'orientation de la tête. Des expérimentations ont été conduites dans ce sens par Langton et al. [LHT04]. Les participants ont été sommés de décider si des personnes posaient le regard directement sur eux ou pas. Les auteurs ont montré qu'il existe un mécanisme, basé sur les contours du visage et l'angle de déviation du nez, qui influence la perception du regard.

La pose de la tête est plus facile à estimer que l'orientation des yeux. Notamment lorsque l'utilisateur est loin de la caméra. La détection des visages sur des images de

faible résolution est plus précise que celle des yeux. D'autant plus que les yeux sont détectés seulement lorsque le visage est frontal ou semi frontal. La détection et le suivi des yeux sont donc, le plus souvent, normalisées par rapport à la pose de la tête. La pose de la tête peut être associée à l'orientation des yeux ou bien utilisée à part afin de déduire la direction du regard. Dans ce qui suit, nous citons les méthodes existantes d'estimation de la pose de la tête. Nous mettons l'accent sur les méthodes basées sur la symétrie bilatérale du visage et nous résumons les avantages et inconvénients de chaque méthode.

2.2 Méthodes d'estimation de la pose de la tête

Nous avons choisi de présenter l'état de l'art de l'estimation de la pose de la tête au travers de deux familles d'approches. Nous distinguons entre la caractérisation de la pose et les différents outils utilisés pour l'estimation de cette dernière. Nous commençons par analyser l'usage des différents descripteurs utilisés pour caractériser la pose sur des images de visages. Ensuite, nous citons les différents outils utilisés tels que les classifieurs ou les régresseurs. Descripteurs et outils peuvent être combinés.

2.2.1 Méthodes de caractérisation du visage

Le problème crucial des approches d'estimation de la pose proposées, est de trouver les caractéristiques les plus discriminantes qui permettent de trouver la meilleure estimation de la pose de la tête. Il existe deux familles de descripteurs : les descripteurs de forme et les descripteurs d'apparence.

2.2.1.1 Descripteurs de forme

La forme du visage est définie par un ensemble de points ou de régions caractéristiques. Ils apportent une information sur la géométrie du visage et par conséquent sur la pose de la tête. Les points caractéristiques peuvent être anatomiques tels que les yeux, le nez et la bouche et/ou géométriques tels que les coins ou une intersection de contours. Ces

points ou régions peuvent être détectés manuellement ou à l'aide de détecteurs génériques tels que le détecteur de Harris [HS88]. Il existe aussi des méthodes pour détecter des régions spécifiques comme celle de la bouche ou des yeux. Nous analysons l'usage de ces descripteurs au travers des trois catégories d'estimation de la pose suivantes : les méthodes géométriques, les méthodes basées sur un modèle et les méthodes basées sur des descripteurs locaux.

2.2.1.1.a Méthodes géométriques

Elles utilisent les positions des points caractéristiques du visage afin de déterminer la pose à partir de leur configuration relative (*cf.* figure 2.2). Wang et Sung [WS07] déterminent l'orientation de 3 lignes sur le visage, considérées parallèles sur l'image plane. Celles reliant les coins intérieurs des yeux, les coins extérieurs des yeux et les coins de la bouche. Le point de fuite de ces lignes est calculé en supposant le ratio de leurs longueurs connu. Pan et al. [PZJ05] utilisent la relation entre la projection de 7 points caractéristiques et leur pose 3D. Parmi les points utilisés, 4 sont les coins des deux yeux et trois se trouvent sur l'axe de symétrie du visage.

FIGURE 2.2: Points caractéristiques faciaux utilisés pour l'estimation de la pose.

Ces méthodes sont simples à implémenter. Cependant, elles ont un inconvénient ma-

jeur, celui de la nécessité de détecter tous les points utilisés dans le calcul et avec précision. Si un seul point venait à manquer, l'orientation ne peut plus être calculée. Ces méthodes requièrent souvent une initialisation avec une vue frontale afin d'essayer de trouver tous les points caractéristiques.

2.2.1.1.b Méthodes basées sur un modèle

Nous présentons dans cette section les différentes approches utilisant des modèles construits à partir de points caractéristiques spécifiques du visage. Ces points sont utilisés comme nœuds d'un modèle non rigide. Les modèles sont utilisés pour l'estimation de la pose de la tête en les superposant sur l'image du visage de telle sorte qu'ils correspondent à la structure faciale. Nous pouvons distinguer trois familles de modèles : Active Shape Model (ASM), Elastic Graph Matching (EGM) et Active Appearance Model (AAM).

Un ASM [CTCG95] modélise la distribution des niveaux de gris dans le voisinage de chaque point par une gaussienne multidimensionnelle afin de représenter la forme. L'initialisation de l'ASM se fait avec un ensemble d'images en entrée où les positions moyennes des points pertinents se trouvent dans des zones prédites. Ainsi, grâce à l'analyse d'un nombre d'images important, il est possible d'estimer des corrélations entre les points, et de calculer les valeurs moyennes de leurs positions afin d'établir une position moyenne du modèle par rapport à l'image.

Kruger et al. [KPM97] représentent les caractéristiques faciales avec un modèle EGM qui est comparé à l'image afin de minimiser la distance entre les points caractéristiques détectés et les nœuds. L'estimation de l'orientation de la tête est effectuée par la création d'un graphe pour chaque pose. Le graphe est déformé itérativement jusqu'à trouver la distance minimale entre les nœuds et les points caractéristiques détectés. La pose associée au graphe qui maximise la similarité avec l'image est choisie.

Le modèle actif d'apparence (AAM) a été proposé par Cootes et al. [CET01]. Les AMMs permettent de tirer profit de toute l'information de texture, ce qui constitue un plus par rapport aux ASMs. La construction d'un AAM commence par la génération d'un ASM à partir d'un ensemble de données d'apprentissage. Le modèle est construit en déformant les images de telle sorte que les points caractéristiques correspondent à

ceux de la forme moyenne. Un modèle de texture est construit en transférant la texture des exemples d'apprentissage. Le modèle de forme est combiné au modèle de texture en faisant un apprentissage sur la corrélation entre la forme et la texture. La figure 2.3 montre l'application du modèle AAM sur un visage.

Plus tard, Baker et al. [BMX+04] ont amélioré l'algorithme précédent pour qu'il puisse traiter 200 trames par seconde. Les auteurs ont, aussi, converti le 2D AMM en un modèle 3D en utilisant une approche d'optimisation trilinéaire afin d'essayer de traiter les cas d'occultations dues aux larges rotations de la tête.

Cependant, les modèles flexibles restent peu robustes aux occultations et difficiles d'utilisation avec des images à basse résolution. En effet, la mise en correspondance avec précision, entre un modèle de visage déformable et une séquence d'images avec une grande amplitude de mouvement, est le principal inconvénient de ces méthodes. Pour pallier ce problème, Caunce et al. [CTC10] recherchent explicitement les contours occultés et déterminent l'importance de chaque correspondance avec le modèle en utilisant une méthode de pondération basée sur la correspondance obtenue.

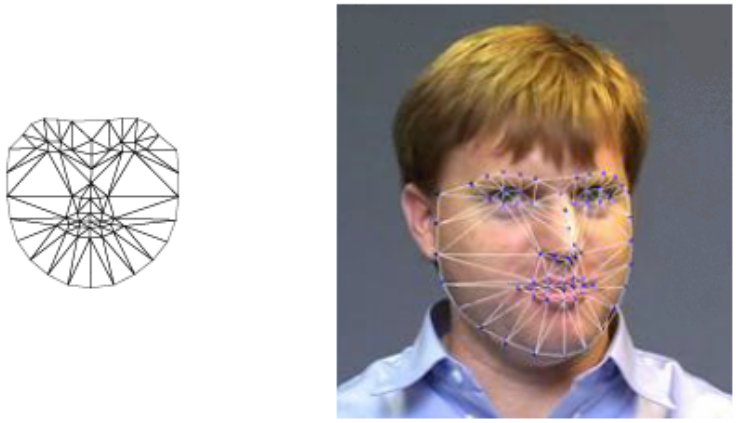

FIGURE 2.3: (a) Un modèle AAM de base. (b) Modèle AAM appliqué sur un visage.

2.2.1.1.c Méthodes basées sur des descripteurs locaux

Les descripteurs locaux ne caractérisent qu'une zone restreinte de l'image. Ils sont utilisés pour qualifier le voisinage des points caractéristiques identifiés. Le descripteur le plus couramment utilisé est SIFT (Scale invariant feature transform) [Low04]. Il présente l'avantage d'être invariant à l'orientation et à la résolution de l'image. Le descripteur SIFT peut être utilisé pour estimer la pose comme étant un vecteur de caractéristiques à classifier [HC12],[GACZE12] ou comme technique pour aligner les images (SIFT Flow) [LYT11]. Dans [WYS12] SIFT Flow est utilisé pour aligner le visage à un prototype prédit représenté par la moyenne des images dans la pose prédite. Cependant, les mêmes points doivent être détectés au même endroit à chaque prise de vue.

2.2.1.2 Descripteurs d'apparence

Les descripteurs d'apparence considèrent l'estimation de la pose de la tête comme étant un signal où l'apparence de la tête entière est utilisée au lieu de se concentrer sur des caractéristiques spécifiques du visage.

Nous présentons ces descripteurs au travers des deux catégories suivantes : le filtrage de l'image et les histogrammes.

2.2.1.2.a Les filtres

Le but du filtrage est d'extraire l'information pertinente dans l'image. Le filtre de Gabor par exemple, est utilisé dans l'analyse de l'image et l'estimation de la pose de la tête. C'est un filtre orienté qui permet de mettre en évidence des textures ainsi que des zones homogènes d'une image. Sa réponse est une sinusoïde modulée par une fonction gaussienne également appelée ondelette de Gabor. Sherrah et Gong [SG01] ont fait varier l'orientation des filtres de Gabor à chaque pose afin d'évaluer le ratio de similarité entre les poses. Lablack et al. [LZD08] utilisent les ondelettes de Gabor pour construire un vecteur de caractéristiques composé de coefficients qui sont échantillonnés en utilisant des échelles et des orientations différentes pour chaque image de pose. Ce vecteur est ensuite utilisé comme entrée pour apprendre un classifieur de poses. Dans [KS02] le

Gabor wavelet network (GWN) est appliqué pour l'estimation de la pose qui est calculée à partir des coefficients des ondelettes. Wu et Trivedi [WT08] utilisent l'analyse des sous espaces dans l'espace des ondelettes de Gabor pendant la première étape de leur système qui affine la pose estimée en deux étapes. Les auteurs appliquent le filtre de Gabor à six échelles et huit orientations puis réduisent la taille des dimensions trouvées pour représenter les poses. La figure 2.4 montre la réponse réelle des ondelettes de Gabor sur huit orientations allant de 0 à $7\pi/8$.

FIGURE 2.4: Réponse réelle des ondelettes de Gabor sur 8 orientations

Des opérateurs non linéaires tels que les Local Binary Patterns (LBP) [OPH96] sont également utilisés. Ils consistent à prendre le voisinage local de chaque pixel et de le seuiller par la valeur du pixel central. Le résultat du seuillage donne une combinaison de valeurs binaires. Dans le cas d'un voisinage 3×3, on obtient un code à 8 bits. Dans [MZS+06] les auteurs combinent les ondelettes de Gabor et les LBP pour estimer la pose de la tête. Le filtre de Gabor pour extraire les caractéristiques d'orientation de la tête et le LBP pour extraire les caractéristiques d'orientations locales du visage. Ce qui a donné le LGBP (Local Gabor Binary Patterns).

Les filtres sont, cependant, sensibles à l'illumination ainsi qu'à l'identité de la personne. En effet, le changement d'illumination ou de personne peut donner des réponses différentes pour la même pose.

2.2.1.2.b Les histogrammes

La représentation par histogramme est un outil fondamental du traitement d'images, simple et puissant avec de très nombreuses applications.

Les histogrammes de couleur sont résistants à un certain nombre de transformations sur l'image. Ils sont invariants en translation et en rotation dans le plan image et varient lentement lors des rotations hors-plan, des changements d'échelles et des occultations. Robertson et Reid [RR06] proposent un modèle basé sur la couleur de la peau et des cheveux en utilisant l'histogramme de couleur afin d'estimer la pose de la tête. Les histogrammes sont en revanche sensibles aux changements d'illumination et aux conditions d'éclairage.

Pour parer à ce problème, Funt et Finlayson [FF95] proposent d'utiliser les dérivées du logarithme des trois canaux de couleur afin d'obtenir des caractéristiques invariantes à l'illumination. Schiele et Crowley [SC00] proposent d'intégrer aux histogrammes de couleur, l'information sur la structure en construisant des histogrammes à champs récepteurs gaussiens multidimensionnels. Les histogrammes de l'orientation des gradients (HOG) sont aussi utilisés pour caractériser des objets [DT05]. L'idée principale des descripteurs HOG est basée sur l'information du contour. Des régions de l'image sont décrites par la distribution locale de l'orientation du contour et la magnitude du Gradient correspondante. Dahmane et Meunier [DM11] appliquent le gradient, comme pré-traitement de l'image, pour définir les régions pertinentes où la magnitude des ondelettes de Gabor sera calculée.

2.2.1.3 Réduction de dimension

Appelées aussi "Apprentissage de variétés" (manifold learning), ces techniques cherchent une relation entre l'espace d'origine des données de dimension D et un espace de plus petite dimension d. Dans notre cas, le sous-espace doit représenter au mieux les variations de pose du visage (*cf.* figure 2.5). La réduction de dimension du point de vue reconnaissance de forme (pattern recognition) est équivalente à l'extraction de caractéristiques où le vecteur de caractéristiques est représenté par les caractéristiques réduites.

Nous pouvons distinguer deux catégories majeures de méthodes de réduction de dimension : linéaires ou non linéaires. Supposons que le vecteur de données $y = (y_1, y_2, ..., y_D)$ de l'espace de grande dimension $\mathbb{R}^D$ est projeté dans l'espace $\mathbb{R}^d$ en $x = (x_1, x_2, ..., x_d)$. Les techniques linéaires de réduction de dimension nous donnent $x = Wy$ où $W_{d\times D}$

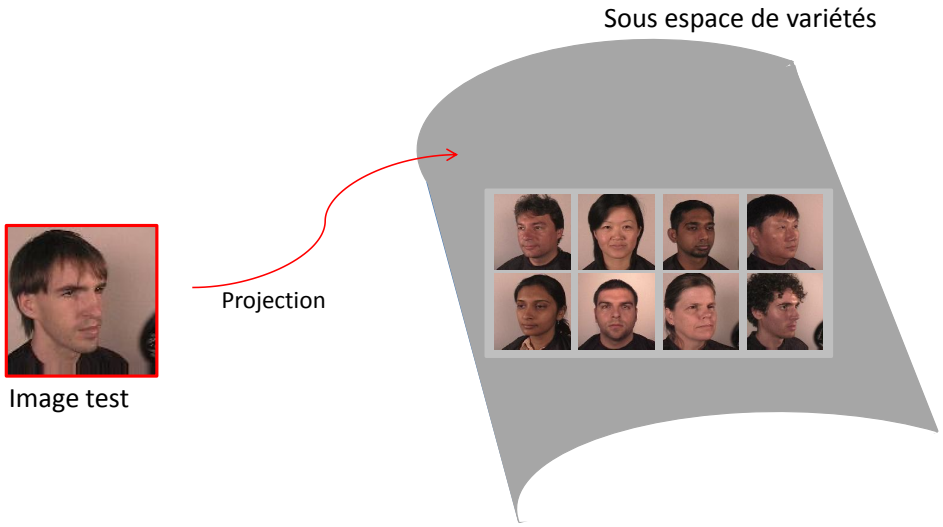

FIGURE 2.5: La réduction de dimension. L'image test est projetée sur un espace de faible dimension qui représente au mieux les variations de pose du visage.

est la matrice de pondération de transformation linéaire. Chaque composant x_i est une combinaison linéaire des variables originales.

$$x_i = w_{i,1}y_1 + ... + w_{i,D}y_D \quad i \in \{1, ..., d\}$$

Les données originales peuvent être représentées par les pixels de l'image brute ou des caractéristiques extraites de l'image. Gong et al. [GMC96] ont appliqué la transformée en ondelettes de Gabor avant de réduire la dimension des données pour distinguer les changements de pose en utilisant l'Analyse en Composantes Principales (ACP). Ptucha et Savakis [PS10] ont testé un ensemble de techniques de réduction de dimension telles que l'ACP [Pea01], Linear Discriminant Analysis (LDA) [Fis36], Locally Linear Embedding (LLE) [RS00], Isomap [TSL00], unsupervised Locality Preserving Projections (LPP) et supervised LPP (SLPP) [HN03] et ceci en utilisant des points caractéristiques du visage

(les points ASM) au lieu de prendre tous les pixels de l'image.

2.2.1.3.a Méthodes linéaires

Les méthodes linéaires sont plus simples et plus faciles à implémenter que les techniques récentes basées sur des transformations non linéaires. L'Analyse en Composantes Principales (ACP) est la technique de réduction de dimension linéaire la plus largement utilisée. Elle est basée sur les statistiques du deuxième ordre. L'ACP est également connue sous le nom SVD (singular value decomposition). Elle cherche des combinaisons linéaires orthogonales (les composantes principales) des variables avec la plus large variance. La base du nouveau sous-espace est alors formée par les d premiers vecteurs propres de la matrice de covariance. Gong et al. [GMC96] ont montré qu'il suffisait de trois axes principaux pour représenter la distribution des visages en fonction de leur pose.

Dong et al. [HSlTB11] ont proposé le Supervised Local Subspace Learning (SL2) pour apprendre un modèle linéaire local où la correspondance entre les données en entrée et l'espace réduit est le résultat de l'apprentissage via un Generalized Regression Neural Network (GRNN). Liu et al. [LLL09] utilisent des images pré-traitées pour construire le sous espace propre (l'espace qui réunit les vecteurs propres) avec la méthode de factorisation de matrice NMF (Non-negative Matrix Factorization [LS99]). Les résultats sont légèrement plus performants que ceux de l'ACP en terme d'erreur moyenne.

2.2.1.3.b Méthodes non-linéaires

Parmi les approches les plus populaires, nous retrouvons ISOMAP [TSL00] qui préserve la distance géodésique entre les points, Locally Linear Embedding (LLE) [RS00] qui préserve la distance basée sur une combinaison linéaire locale du voisinage et Laplacian Eigenmaps (LE) [BN03] qui préserve la distance décrite par un graphe pondéré construit par le voisinage.

Xianwang et al. [WHGY08] ont combiné ISOMAP et LFDA qui est une technique d'apprentissage supervisé de distance métrique afin d'avoir un meilleur regroupement intra-classes.

Le système BME (Biased Manifold Embedding) a été proposé par Balasubramanian et al. [BYP07] pour l'estimation de la pose de la tête. L'information de pose issue des images d'apprentissage est utilisée pour construire l'espace de petite dimension, contrairement aux techniques de réduction de dimension qui déterminent seulement la relation géométrique entre les données de l'espace de grande dimension. Les labels sont utilisés pour calculer un voisinage partial (biased neighborhood) pour chaque point de l'espace de données. Ainsi, le système BME pondère chaque distance inter-exemples, par un facteur qui dépend de la différence des labels.

La figure 2.6 illustre la réduction de dimension via le Laplacian Eigenmaps (LE) [BYP07]. Les images utilisées représentent trois poses, une frontale et deux de profil. Nous pouvons remarquer que le sous-espace représente bien les variations de pose du visage. La vue frontale se trouve au centre d'une trajectoire elliptique, les poses négatives d'un coté et les poses positives de l'autre.

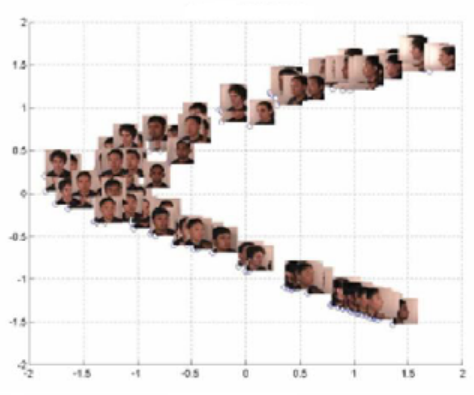

FIGURE 2.6: Exemple d'intégration d'images dans un sous-espace LE ([BYP07]).

Cependant, les techniques de réduction de dimension ont tendance à créer un sous espace pour l'identité aussi bien que pour la pose. Les méthodes de caractérisation du visage ont été regroupées selon leur usage dans l'estimation de la pose. Ces différents descripteurs peuvent être utilisés par des outils génériques que nous présentons dans le paragraphe suivant.

2.2.2 Méthodes génériques d'estimation de la pose de la tête

2.2.2.1 Méthodes par classification

Ces méthodes considèrent l'estimation de la pose comme un problème de classification de forme. Les visages qui ont la même orientation sont regroupés dans la même classe. Une méthode d'apprentissage est alors appliquée pour déterminer si un visage inconnu appartient à une classe d'orientation donnée.

La classification peut se faire avec un ensemble de classifieurs en parallèle (*cf.* figure 2.7 (a)), chacun vise à détecter une classe d'orientation. Dans ce genre d'architectures, l'ensemble d'apprentissage est partitionné entre les classifieurs. Ce qui n'est pas évident car les frontières entre les classes sont continues. Cependant, une autre manière de faire consiste à utiliser toutes les données pour apprendre simultanément toutes les classes d'orientation. Cette architecture est appelée multi-classes (*cf.* figure 2.7 (b)).

Une large gamme de classifieurs a été utilisée dans l'estimation de la pose, tels que les Séparateurs à Vaste Marge (SVMs), les réseaux de neurones et la classification bayesienne.

Plusieurs méthodes d'estimation de la pose ont utilisé les SVMs [HSW98a] et plus récemment [MZS+06] et [DM11]. Les classes sont regroupées à l'aide des marges maximales de l'hyperplan qui sépare une classe de l'autre [BGV92]. Isarun et al. [CSS+11] ont utilisé les forêts d'arbres décisionnels aléatoires [Bre01] en plus des SVMs. Les forêts d'arbres décisionnels sont une combinaison d'arbres de décision en faisant la moyenne. Un autre type d'arbre de décision appelé Randomised ferns, qui est une combinaison d'arbres en utilisant les règles de Bayes, sont définis par Özuysal et al. [OFL07]. Les réponses sont intégrées d'une manière bayésienne naïve ce qui permet d'obtenir une meilleur extensibilité en terme de nombre de classes. Ce type d'arbre de décision est utilisé dans l'estimation de la pose, par Benfold et Reid [BR08]. Les images sont normalisées à 10×10 pixels, segmentées en six segments de couleur en utilisant les k-means. Chaque segment est manuellement annoté (cheveux, peau ou background). Ces labels seront utilisés pour construire l'arbre afin de classifier la pose.

Le classifieur naïf de Bayes est, donc, utilisé dans l'estimation de la pose [WT00].

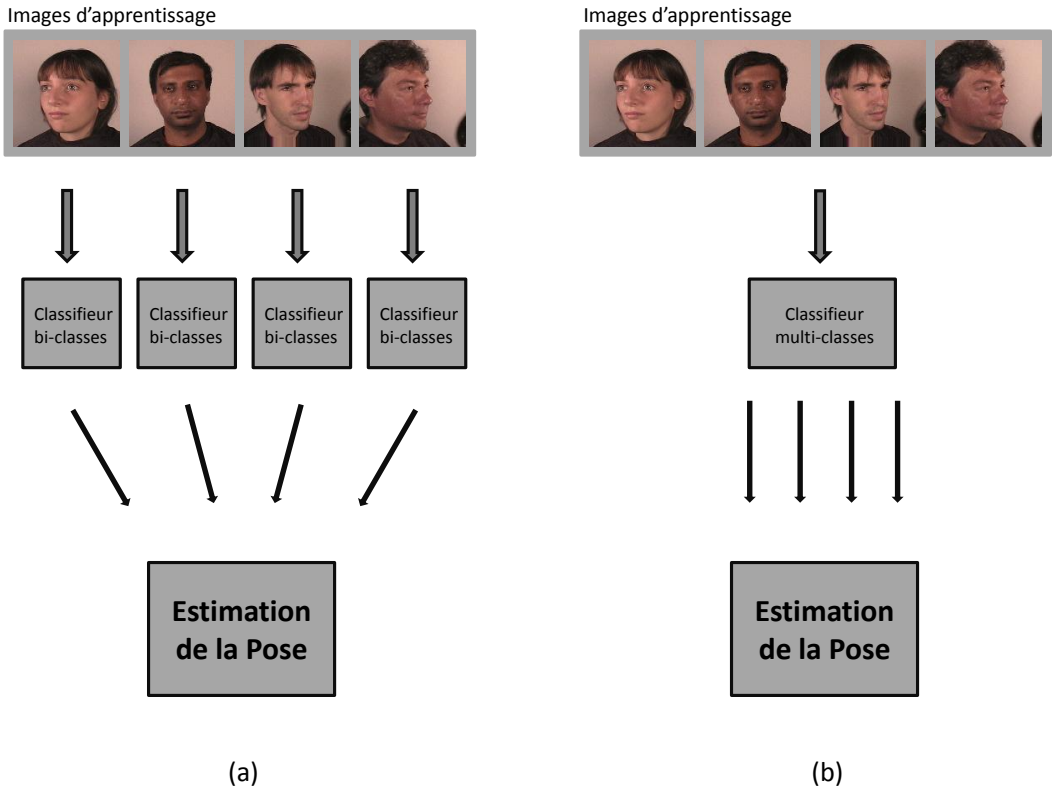

FIGURE 2.7: Classification de la pose. (a) Architecture parallèle. (b) Architecture multi-classes.

C'est un type de classification bayésienne probabiliste qui suppose que les caractéristiques d'une classe sont indépendantes les unes des autres. Dans [ZHLH07], le classifieur naïf de Bayes est appliqué afin de fusionner l'information de pose à partir de vues provenant de plusieurs caméras.

Le perceptron multicouche MLP (Multi Layer Perceptron) est un modèle de réseau de neurones organisé en plusieurs couches où chaque couche est constituée d'un nombre variable de neurones. Dans [VNS06] chaque perceptron sur la couche de sortie correspond à une classe d'orientation. Gourier et al. entraînent un réseau de neurones pour chaque

classe afin de synthétiser une image ressemblant à celle en entrée. La pose estimée sera celle du réseau qui a obtenu l'erreur de reconstruction la plus faible. Rowley et al. [RBK98] ont étendu la détection de visages frontaux aux cas où une rotation dans le plan leur est associée. La méthode consiste à utiliser un réseau de neurone appelé "routeur" qui va analyser les images en entrée avant d'être traitées par le détecteur de visages. Si l'entrée contient un visage, le routeur retourne l'angle de rotation du visage. Ainsi, l'architecture proposée inverse les étapes de détection de visage et d'estimation de la pose.

Les méthodes par classification sont robustes lorsqu'elles s'appuient sur des méthodes d'apprentissage robustes (les SVMs, les arbres de décision, ...). L'avantage des méthodes de classification est que le problème de détection du visage peut être traité conjointement avec l'estimation de la pose. Cependant, cet avantage peut s'avérer un problème car la détection du visage cherche à trouver les caractéristiques partagées par tous les visages quelle que soit leur orientation, alors que l'estimation de la pose cherche à différencier les orientations.

2.2.2.2 Méthodes par régression

Les outils de régression sont utilisés pour apprendre une relation fonctionnelle entre l'apparence d'un visage et une mesure de l'orientation de la tête. La relation fournit une estimation continue de la pose pour un visage inconnu. Un ensemble d'apprentissage est nécessaire pour modéliser cette relation fonctionnelle. Le principe de la régression est illustré par la figure 2.8.

Les outils les plus souvent utilisés pour apprendre la régression sont les réseau de neurones. Tia et al. [lTBC$^+$03], utilisent un système multi caméras et estiment la pose de la tête par des réseaux de neurones. Parmi les réseaux existants, le plus utilisé est le MLP. Dans [SYW02] deux MLP à une seule sortie sont utilisés, chacun pour un degré de liberté. Contrairement au MLP qui utilise la rétro-propagation, LLM (Locally Linear Map) est un autre réseau de neurones composé de cartes linéaires qui combine l'apprentissage supervisé et non-supervisé. Rae et Ritter [RR98] ont utilisé trois LLMs pour identifier la pose de la tête. La segmentation des couleurs est réalisée afin de différencier la région du visage de l'arrière plan. Elle est basée sur un réseau LLM entrainé à produire une valeur

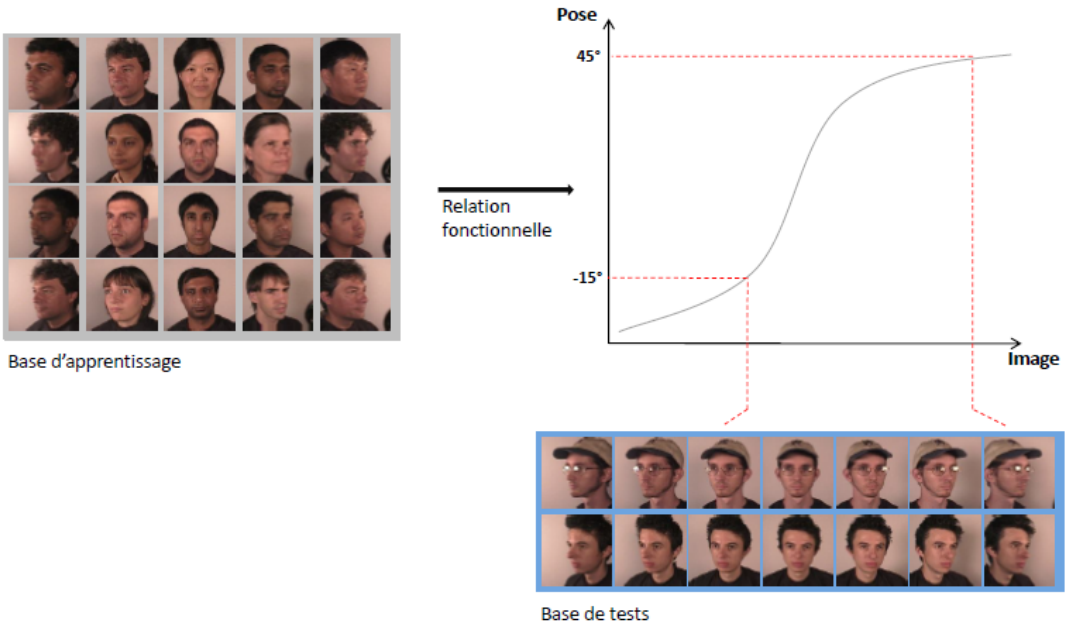

FIGURE 2.8: Les méthodes par régression. La pose est estimée pour une image test moyennant une relation fonctionnelle apprise entre l'apparence d'un visage et sa pose.

de probabilité pour chaque pixel appartenant à la zone du visage. Un autre réseau LLM est utilisé pour donner une localisation plus précise du visage et le troisième réseau LLM est entrainé pour l'estimation de la pose.

Les machines à vecteurs de support régressif SVR (Support Vector Regressors) ont été utilisées avec succès pour l'estimation de la pose. Murphy-Chutorian et al. [MCDT07] utilisent les SVRs sur une concaténation d'histogrammes d'orientations de contours calculés pour différentes régions de l'image. Dans [LGSL04] les SVRs ont été appliqués après une réduction de dimension par l'ACP sur des images filtrées par un opérateur de Sobel et dans [MKK+06] sur 20 points faciaux labellisés manuellement.

D'autres méthodes de régression sont possibles, telle que la CRSR (Convex Regularized Sparse Regression) [JLS+11] et la GPR (Gaussian Process Regression) [RY08].

Murad et al. [HGD12] proposent une méthode basée sur la régression des moindres carrés partiels PLS (Partial Least Squares).

Les résultats des méthodes par régression sont dépendantes du résultat de la détection du visage, car ces méthodes estiment la pose indépendamment de l'étape de détection du visage. Elles sont rapides et précises mais elles restent complexes.

2.2.2.3 Méthodes par comparaison de prototypes

De manière intuitive, ces méthodes consistent à comparer l'apparence d'un visage inconnu à un ensemble d'images étiquetées correspondant à plusieurs poses. La pose estimée sera celle du visage le plus ressemblant dans la base de données (*cf.* figure 2.9). La comparaison se fait en utilisant des images brutes ou des images filtrée afin d'accentuer les caractéristiques pertinentes liées à la pose. Les critères de correspondance peuvent se baser sur plusieurs métriques telles que l'erreur quadratique moyenne, ou la corrélation croisée normalisée.

FIGURE 2.9: Les méthodes par comparaison de prototypes. La pose de l'image test est celle de l'image prototype la plus ressemblante.

Dans [Bey94], les auteurs représentent les visages par des modèles qui couvrent différentes poses et cherchent la meilleure ressemblance avec une donnée en entrée en utilisant la corrélation avec le modèle. Le même principe de similarité avec des prototypes est adopté par Jamie et Sherrah [SGO99] afin de calculer le ratio de similarité de la pose.

Cependant, une grande base d'images entraine un nombre de calculs important. Aussi, elles sont sensibles aux erreurs de localisation du visage. Néanmoins, l'inconvénient majeur de ces méthodes reste l'influence de l'identité du visage. L'effet de l'identité peut causer plus de dissemblance que la pose elle même.

2.2.2.4 Méthodes par suivi de mouvement

Le suivi de mouvement ou le tracking pour l'estimation de la pose de la tête utilise l'information temporelle en suivant cette dernière sur plusieurs images. Parmi les méthodes les plus utilisées pour représenter le mouvement dans les vidéos, on trouve *le flux optique* (ou flot optique). Il permet d'estimer le déplacement des pixels entre deux images successives. L'hypothèse initiale pour le calcul du flux optique est que l'intensité des points de l'image à travers le temps est approximativement constante pour des petites durées [HS81].

Benfold et Reid [BR09] ont proposé une méthode de suivi basée sur le filtre de Kalman où la vélocité du mouvement de la tête est calculée en combinant les vélocités du suivi de plusieurs points d'intérêt. Sugimura et al. [CSS+11] ont utilisé la méthode proposée par Benfold et Reid pour le suivi de piétons afin d'étiqueter automatiquement la pose de la tête à partir des déplacement des piétons. Ba et al. [BO11] déterminent l'attention visuelle des personnes en utilisant un système de suivi basé sur les filtres particulaires qui constituent souvent une alternative aux filtres de Kalman. Valenti et al. [VSG12] utilisent le modèle cylindrique CHM (Cylindrical Head Model) pour le suivi de la tête. Les pixels de l'image sont projetés en points 3D. Le flot optique est alors utilisé pour estimer l'orientation de la tête. Les auteurs cherchent à pallier le problème de perte de la trajectoire du mouvement en combinant le modèle avec la position des yeux. Effectivement, le suivi de mouvement a besoin d'être réinitialisé lorsqu'il devient instable afin de récupérer la trajectoire.

2.2.3 Estimation de la pose de la tête en utilisant la symétrie

Une famille spécifique des approches exploitant les caractéristiques globales du visage tout en réduisant la dépendance à l'identité, est représentée par les approches qui exploitent la symétrie bilatérale du visage. En effet, malgré le fait que le visage humain ne soit pas parfaitement symétrique, la symétrie faciale des personnes est significative et peut être exploitée dans l'estimation de la pose. Quelques travaux, se basant sur les caractéristiques faciales, utilisent la propriété de symétrie du visage pour l'estimation de la pose [LMH+09b], [PDG08], [GH04], [HCZZ04]. La symétrie faciale est aussi utilisée comme indicateur des régions utilisées pour apprendre à détecter un visage dans une image [RBK98]. Pour cela, les auteurs estiment la symétrie de chaque région.

Nous trouvons dans l'état de l'art quelques travaux d'estimation de la pose basées sur la symétrie. L'hypothèse est que la quantité de symétrie peut donner des indices sur l'orientation de la tête. Luhandjula et al [LMH+09b] utilisent la symétrie faciale comme indicateur de l'attention du regard pour les personnes handicapées. L'hypothèse est que les différentes positions de la tête, droite, gauche, haut et bas, donnent une courbe de symétrie différente. La courbe de symétrie [LvWKW06] est définie par une valeur de symétrie S calculée pour chaque colonne x de l'image I $(X \times Y)$, en utilisant à chaque fois toutes les colonnes distantes d'un maximum de k pixels de la colonne dont on calcule la valeur de symétrie. Le calcul se fait à l'aide de l'équation suivante :

$S(x) = \sum_{w=1}^{k} \sum_{y=1}^{Y} [abs(I(x-w, y) - I(x+w, y))]$

Luhandjula et al. estiment la pose de la tête en classifiant les courbes de symétrie. Pathangay et al. [PDG08] estiment la pose à partir d'une seule vue en exploitant la structure symétrique du visage. Les auteurs utilisent l'image miroir de l'image test comme une deuxième vue virtuelle. La pose est évaluée en mettant en correspondance les points caractéristiques AAM de l'image test et de son miroir (*cf.* figure2.10).

Un modèle d'illumination basé sur la symétrie est proposé par Gruendig et Hellwich [GH04]. Un opérateur de symétrie radiale [LZ03] est appliqué afin de détecter trois points caractéristiques faciaux qui sont les yeux et le nez. Pour chaque combinaison des deux yeux et du nez, la pose est calculée en utilisant une projection géométrique. Par la suite,

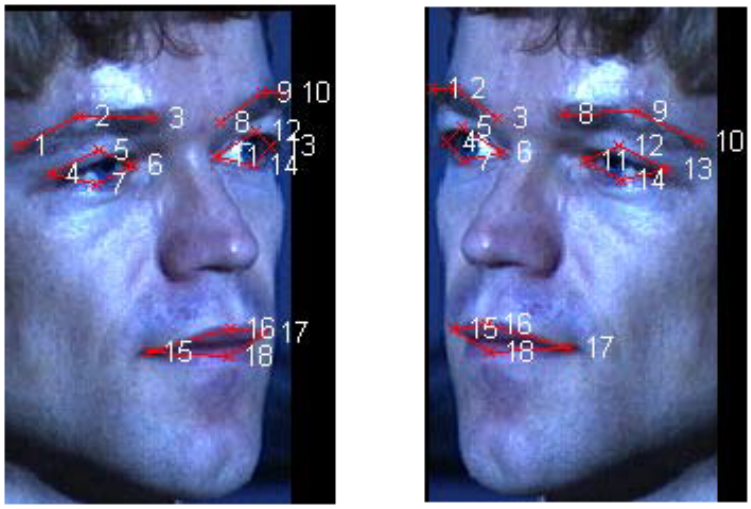

FIGURE 2.10: Détection des points caractéristiques AAM sur l'image test, à gauche, et son miroir, à droite, afin de les mettre en correspondance (tiré de [PDG08]).

l'estimation de la pose est raffinée moyennant un modèle d'illumination construit à partir d'images de visages chacune illuminée d'un endroit différent. Ce raffinement est formulé comme étant un problème de moindre carrés cherchant à trouver la pose qui satisfait le modèle d'illumination avec les positions des trois points caractéristiques détectés. Ma et al. [HCZZ04] estiment de manière grossière la pose. Pour cela, ils établissent une relation entre le visage et la pose en utilisant la distribution de cinq points caractéristiques qui sont les deux coins de la bouche, les deux centres des yeux et le bout du nez. L'asymétrie de l'apparence est alors détectée pour les poses non frontales.

Toutes ces méthodes manquent de robustesse du moment qu'elles se basent sur des caractéristiques faciales difficiles à extraire et en même temps indispensable pour l'estimation de la pose. L'utilisation de certaines caractéristiques réduit considérablement l'ampleur des mouvements (par exemple, un coin d'œil ou un coin de la bouche caché sur une pose de profil). De plus, d'autres inconvénients apparaissent : dans [LMH+09b], la

courbe de symétrie n'est pas représentative de la symétrie présente sur l'image pour les poses non frontales. Dans [GH04], la détection des points caractéristiques via l'opérateur radial est loin d'être efficace. 15 possibilités de positions des yeux sont retenues dans la partie du haut du visage afin de s'assurer que les bonnes positions soient incluses. Aussi les régions des yeux et du nez doivent être connues car ces caractéristiques sont recherchées chacune dans sa région. Un autre inconvénient apparait dans [HCZZ04], est le fait que le traitement doit commencer avec une pose frontale. Donc, l'utilisateur doit initialiser sa pose face à la caméra et lorsque les paramètres d'initialisation sont perdus en cours de traitement (par exemple un point caractéristique disparait), la pose ne peut plus être estimée à moins qu'une autre initialisation ne soit à nouveau effectuée.

2.2.4 Synthèse

Les familles d'approches citées plus haut peuvent être combinées afin de se renforcer les unes les autres. En effet, la robustesse peut être augmentée dans un système hybride où plusieurs modules d'estimation de la pose sont combinés. Les différents modules peuvent être combinés de deux manières :

- séquentielle, comme par exemple l'incorporation d'un algorithme d'estimation de la pose dans un système de filtre particulaire afin de suivre la tête et d'estimer sa pose en même temps [BmO04].
- parallèle, la fusion des résultats d'estimation provenant de modules indépendants afin d'obtenir un seul résultat [WPP+04].

Cependant, chaque famille d'approches a ses limitations. Les approches basées sur l'apparence souffrent de l'information sur l'identité et la luminosité qui sont contenues dans l'apparence globale. Pour les méthodes par comparaison de prototype, l'effet de l'identité peut causer plus de dissemblance que la pose elle même. D'autre part, les approches basées sur des modèles sont indépendants de l'identité lorsque les points caractéristiques utilisés sont liés à la morphologie humaine. En effet, les variations anthropométriques entre les individus sont petites. Par exemple, d'une personne à une autre,

l'écart entre les yeux, n'est pas très différent et de même pour les autres mesures. Cependant, si les points caractéristiques sont liés à la texture (par exemple, intersection de contours), l'effet de l'identité est influent.

Aussi, les approches basées sur l'apparence nécessitent un temps de calcul considérable ce qui rend le système temps réel difficile à implémenter. Les approches basées sur la forme sont rapides mais sensibles aux occultations et nécessitent des images de haute résolution. La difficulté réside dans la détection des points d'intérêt avec précision. Les points caractéristiques nécessaires dans un système d'estimation de la pose doivent être détectés tous, à défaut, le système s'arrête de fonctionner. Il est important de mentionner que les dispositifs à haute résolution ne sont pas toujours disponibles. Un autre problème se pose pour les approches basées sur la forme, est que le système requiert souvent d'être initialisé avec une vue frontale. Un résumé des inconvénients de chaque type de caractéristiques utilisées dans l'estimation de la pose, est présenté dans le tableau 2.1. Nous prenons en compte les critères de robustesse à l'illumination et l'indépendance à l'identité ainsi que la contrainte de l'obligation d'initialisation de la pose. Les caractéristiques très affectées par une contrainte se voient octroyer le sumbole (-) et ceux qui surmontent bien la contrainte, le symbole (+).

Tableau 2.1: Résumé des caractéristiques utilisées dans l'estimation de la pose.

Type de caractéristiques	Initialisation requise	Robustesse à l'illumination	Indépendance de l'identité
Géométriques	-	-	+
AAM	-	-	+
EGM	-	-	+
Filtres	+	-	-
Histogramme	+	-	-

Nous cherchons à exploiter une approche basée sur des caractéristiques globales du visage afin de ne pas dépendre de la présence et de la bonne détection de points caractéristiques spécifiques. Nous souhaitons aussi éliminer la dépendance à l'identité et réduire l'effet de l'illumination. Les caractéristiques géométriques de la symétrie bilatérale du

visage, extraites de manière globale semblent répondre à nos exigences. Nous utilisons, donc, la symétrie bilatérale du visage pour l'estimation de la pose de la tête à condition que la détection de cette symétrie se fasse de manière globale. Nous présentons dans le chapitre suivant l'état de l'art des méthodes de détection de la symétrie, ainsi que la méthode proposée.

Chapitre 3

La symétrie dans l'image

Sommaire

La symétrie est omniprésente dans notre environnement, qu'elle soit naturelle ou artificielle. Elle joue un rôle important dans la perception humaine. Des tests psychophysiques soulignent que la détection de la symétrie permet de signaler la présence d'objets et d'attirer l'attention visuelle [DH98]. En effet, deux régions symétriques se trouvent rarement, dans une image, par hasard. Il y a de fortes chances qu'il existe un lien entre ces régions. Ainsi, les applications de reconnaissance de forme et de compréhension de la scène peuvent exploiter les symétries présentes sur une image. Il est possible, par exemple, de lancer la segmentation de l'image sans avoir de connaissances préalables sur son contenu, en se basant sur les régions contenant des symétries locales. La symétrie peut être utilisée aussi comme caractéristique afin de guider la mise en correspondance lors de la reconnaissance d'objets.

Il existe plusieurs types de symétrie : la symétrie bilatérale (ou symétrie axiale ou encore réflexion), la symétrie rotationnelle et la symétrie radiale qui est la combinaison des deux premières (*cf.* figure 3.1). La symétrie bilatérale est la plus couramment observée et analysée. Le visage humain est caractérisé par une symétrie bilatérale. De ce fait, nous nous basons sur ce type de symétrie.

3.1 Méthodes génériques pour la détection de la symétrie bilatérale

Différents champs applicatifs peuvent bénéficier de l'apport de l'information fournie par l'analyse des symétries présentes dans l'image. Nous distinguons l'analyse d'imagerie médicale [MGM05], [LCR01], la détection de véhicules [Kue91] et notre domaine d'intérêt, l'analyse faciale [ML04].

L'axe de symétrie peut être droit ou courbé (*cf.* figure 3.2). Même si on peut observer souvent des symétries bilatérale avec un axe courbé dans la nature, pour le visage humain, l'axe reste droit pour une grande partie du visage.

Les méthodes de détection de la symétrie peuvent se baser sur des caractéristiques locales ou globales. Elles peuvent aussi utiliser des images 3D ou 2D. Nous présentons

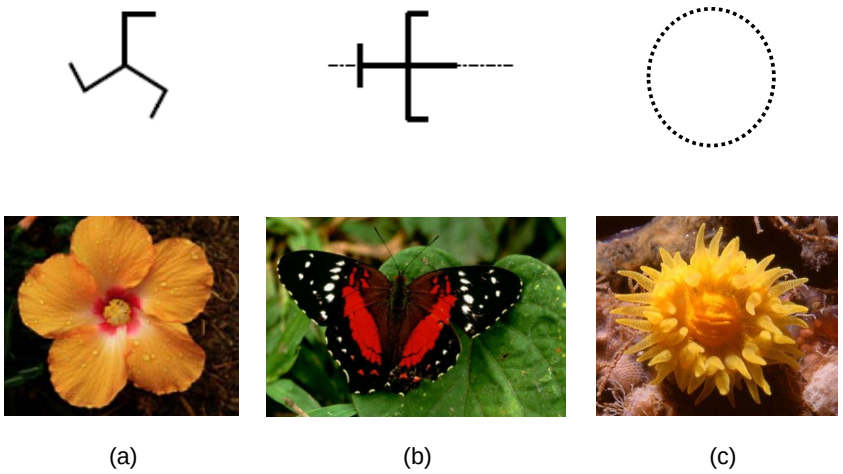

FIGURE 3.1: Les trois types de symétrie (schémas et images réelles). (a) rotationnelle, (b) bilatérale, (c) radiale.

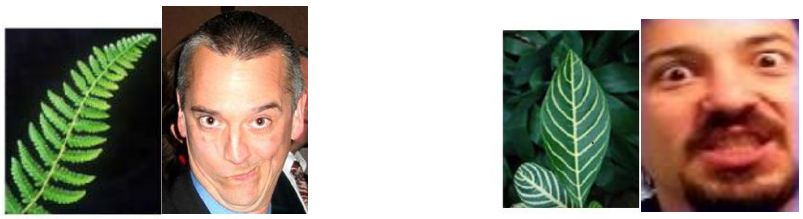

FIGURE 3.2: L'axe de symétrie dans la nature. (a) courbé, (b) droit. Nous pouvons remarquer que pour le visage humain, même avec des grimasses, une grande partie du visage reste symétrique.

par la suite quelques méthodes de l'état de l'art au travers des deux catégories 3D et 2D. La différence entre ces deux catégories réside essentiellement dans la nature du matériel de capture utilisé. Les dispositifs 3D nous donnent les coordonnées de tous les points de

la scène. Alors que les images 2D sont une projection des points de la scène qui sont face à la caméra. Les symétries extraites dans les deux cas, sur le même objet, peuvent être différentes.

3.1.1 Symétrie 3D

Plusieurs dispositifs de capture 3D existent tels que les scanners laser et les systèmes stéréoscopiques. Toutefois, les images 2D + profondeur qu'on appelle aussi 2D et demi peuvent êtres utilisées. La profondeur est calculée en illuminant la scène avec du proche infra rouge. La caméra calcule ensuite le temps que la lumière prend pour effectuer le trajet jusqu'à l'objet, c'est les caméras temps de vol ou TOF (Time Of Flight). Un capteur photosensible peut aussi être utilisé pour convertir le rayonnement Infra Rouge réfléchi par les objets, en un signal numérique. Les objets de la scène se verront attribuer une profondeur qui correspond à la quantité de la lumière infra rouge réfléchie. C'est le mécanisme utilisé avec la Kinect par exemple (*cf.* figure 3.3). Cependant, ces méthodes ne peuvent pas être utilisées dans n'importe quel environnement car la lumière ambiante, quelle soit naturelle ou issue d'autres sources lumineuses, peut interférer avec la mesure de la distance.

Sato et Tamura [ST96] cherchent à détecter les symétries présentes sur des silhouettes à partir d'images 3D acquises avec un scanner. La méthode étant basée sur le concept de la transformée de Hough, les auteurs définissent les conditions qui doivent être satisfaites par les paires de points symétriques afin d'appartenir à l'espace des paramètres. Les paramètres estimés des plans de symétrie sont obtenus en détectant les pics dans l'espace des paramètres. Les conditions de symétrie entre deux points $x1$ et $x2$ sont en relation avec leurs normales $n1$ et $n2$, de sorte que,

$(n_1 + n_2) \times (x_1 - x_2) = 0 \quad$ et $\quad (n_1 - n_2) = \lambda(x_1 - x_2)$,

avec λ étant un coefficient scalaire. Le résultat est ensuite utilisé comme estimation initiale pour la méthode des moindres carrés itérativement repondérés, afin d'être affiné.

La symétrie est extraite des modèles 3D de manière partielle ou approximative par Mitra et al. [MGP06]. La méthode est basée sur la mise en correspondance des signatures

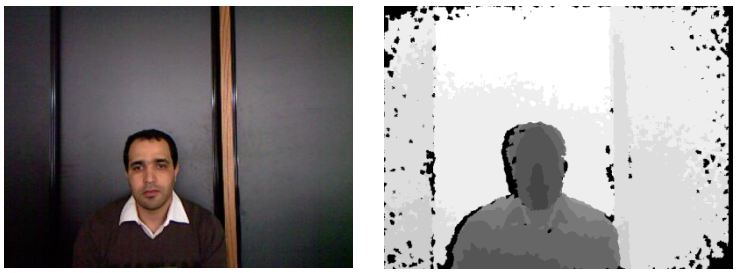

FIGURE 3.3: Exemple d'image 2D et demi. (a)Image de texture. (b) Image de profondeur correspondante, prise avec une Kinect. Le niveau de gris de chaque pixel représente la profondeur du point sur la scène.

locales de la silhouette. Les mêmes auteurs [MGP07] proposent une méthode de symétrisation de modèles 3D, qui consiste en le rehaussement des symétries approximatives déjà présentes. Deux poses 3D du modèle sont nécessaires. La symétrisation est faite en calculant la correspondance entre les deux poses. Une symétrie bilatérale courbée est définie à partir de symétries localement droites.

Le calcul de la symétrie est relativement précis selon la qualité des images 3D. Cependant, le matériel de capture 3D est le plus souvent très onéreux et non disponible au grand public, spécialement celui qui possède une grande précision et effectue les mesures en temps réel. La symétrie 3D est le plus souvent utilisée dans les domaines du graphisme et de la reconstitution de scène.

Pour que la détection de la symétrie soit associée au plus grand nombre possible d'applications, la majorité des travaux réalisés dans ce domaine requièrent l'utilisation d'une seule caméra. C'est même exigé par Murphy et Trivedi [MCT09] comme critère de robustesse d'un système d'estimation de la pose de la tête.

3.1.2 Symétrie 2D

La détection automatique de la symétrie dans les images est un intérêt de recherche durable en vision par ordinateur. Plusieurs approches ont été proposées [LE06], [CFB05], [Ste05], [DG04], [SS99].

Pour trouver l'orientation des axes de symétrie dans une image, Sun et Si [SS99] utilisent le vecteur gradient de la surface. Les orientations des vecteurs gradients sont représentées avec un histogramme. S'il existe des symétries dans l'objet, alors on les retrouve dans l'histogramme d'orientation du gradient à un ou plusieurs angles d'orientation. La convolution de l'histogramme avec lui même, qui est obtenue par la transformée de Fourrier, produit des pics à ces angles de symétrie. Après avoir trouvé la direction de ou des axes de symétrie, leur position est déterminée en utilisant le centre de gravité de l'image qui doit être connu. L'axe de symétrie est supposé passer par le centre de gravité.

Dans [LE06], la méthode proposée permet de détecter les symétries locales ou globales. Elle est basée sur la détection de points caractéristiques invariants à l'échelle (SIFT par exemple). Ces points ainsi que leurs miroirs sont décrits en fonction de leurs positions relatives, leur rotation et de leur échelle. Les deux ensembles de descripteurs sont ensuite comparés afin de définir la "quantité" de symétrie de chaque paire de points caractéristiques. Un vote est alors effectué pour trouver l'axe de symétrie dominant. Cette méthode reste sensible à la détection des points caractéristiques pertinents de l'image.

Dans [DG04], la symétrie est détectée, en utilisant la représentation de la silhouette dans le domaine de Fourrier Melin. Cependant, l'objet doit être préalablement extrait de la scène et son centre de gravité connu.

Chen et al. [CFB05] utilisent le GLDH (Grey Level Difference Histogram) pour détecter l'axe de symétrie du visage. Seulement, cette méthode s'applique sur des images de visages qui regardent directement la caméra avec un arrière plan net et une lumière contrôlée.

Ces méthodes nécessitent des hypothèses particulières ou des connaissances préalables. La méthode que nous avons choisie comme base de travail, est celle de F. Stentiford [Ste05]. Cette méthode est basée sur un modèle d'attention visuelle. Des réflexions

et des translations aléatoires sont appliquées à des régions aléatoires de l'image avant de tester leur correspondance. La correspondance de deux régions indique un axe de symétrie potentiel. L'inclinaison qui comptabilise le plus grand nombre de régions symétriques est celle qui est choisie. La méthode ne nécessite pas de connaissances supplémentaires ou des conditions particulières sur l'image ni d'intervention manuelle. Dans la section suivante, nous allons détailler le fonctionnement de cette méthode et la tester. Nous allons par la suite faire ressortir ses lacunes pour pouvoir l'adapter à notre contexte qui est la détection de la symétrie bilatérale sur des visages à différentes poses.

3.2 Extraction statique de la symétrie dans les visages

Notre but est de trouver la symétrie bilatérale du visage, définie par l'axe qui passe entre les deux yeux. Dans ce qui suit, nous allons présenter l'algorithme de base de notre méthode de détection de la symétrie. Nous allons le tester dans le contexte d'estimation de la pose, mettre l'accent sur ses lacunes, et l'adapter à notre cas applicatif.

3.2.1 Détection de la symétrie basée sur un modèle d'attention visuelle

Le principe utilisé pour détecter les régions saillantes dans une image est exploité par F. Stentiford [Ste05] dans la détection de la symétrie de réflexion. La méthode proposée dans [Ste05] détecte les axes de symétrie sans requérir d'hypothèse préalable sur l'image à analyser. Ce qui nous a incité à nous baser sur le même principe pour la détection de la symétrie bilatérale du visage.

F. Stentiford propose une méthode basée sur un modèle d'attention visuelle qui détermine ce qui est important dans une scène. L'idée est qu'une région qui ne correspond pas à la plupart des autres régions de l'image est très susceptible d'être anormale et se démarque au premier plan. Pour ce faire, un ensemble de pixels dans une petite région

de l'image est comparé à sa translation dans une autre région. Une région de l'image qui ne ressemble pas à beaucoup d'autres se voit attribuer un score d'attention élevé. La symétrie est détectée en utilisant le même mécanisme que pour mesurer l'attention. Un ensemble de pixels est comparé à un autre ensemble dont la position est calculée en appliquant une translation et une réflexion aléatoires. L'angle de réflexion qui fait correspondre le plus grand nombre d'ensembles déterminera l'angle de l'axe de symétrie. Le principe de base de la méthode de F. Stentiford repose sur les définitions suivantes :

Définition 1 :
Deux pixels se correspondent si la différence entre leurs niveaux de gris est inférieure à un seuil ε. Soit $F(P)$ la valeur du niveau de gris du pixel $P\begin{bmatrix} x \\ y \end{bmatrix}$, alors

- Si $|F(P_1) - F(P_2)| > \varepsilon$, pas de correspondance entre P_1 et P_2
- Si $|F(P_1) - F(P_2)| < \varepsilon$, correspondance entre P_1 et P_2

Définition 2 :
Une région de l'image est symétrique à une autre si chaque pixel de la première correspond à son symétrique dans la seconde.

L'ensemble $S_1 = \{P_1^1, P_1^2, \ldots, P_1^m\}$ représente m pixels appartenant à une région de l'image. Le pixel P_1^1 est choisi de manière aléatoire sur l'image et les pixels de P_1^2 jusqu'à P_1^m sont pris aléatoirement dans le voisinage de P_1^1. L'ensemble S_1 doit être constitué de manière à avoir au moins h pixels qui ne correspondent à aucun des autres pixels de l'ensemble afin de s'assurer que la surface à traiter n'est pas homogène et éviter les symétries insignifiantes. Soit r le rayon de la région, comme illustré dans la figure 3.4. Le pixel rouge représente P_1^1 et les pixels noirs P_1^2 à P_1^m sont pris aléatoirement dans le voisinage de P_1^1.

Une translation δ et une réflexion selon un angle α, aléatoires, sont appliquées à S_1

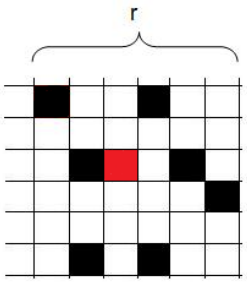

FIGURE 3.4: L'ensemble des pixels qui représentent la région à traiter.

qui nous donnent l'ensemble $S_2 = \{P_2^1, P_2^2, \ldots, P_2^m\}$. La translation se fait comme suit :

$$P_2^1 = P_1^1 + \delta \tag{3.1}$$

La réflexion se fait comme suit :

$$\begin{bmatrix} x_2^i - x_2^1 \\ y_2^i - y_2^1 \end{bmatrix} = \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix} \begin{bmatrix} x_1^i - x_1^1 \\ y_1^i - y_1^1 \end{bmatrix} \quad i \in [2, m] \tag{3.2}$$

La figure 3.5 (a) représente un ensemble S_1 (en noir) et l'ensemble S_2 (en rouge) après une translation $\delta = \binom{0}{6}$ et une réflexion de $\alpha = \frac{\pi}{2}$. Les pixels pleins représentent P_1^1 et P_2^1.

On vérifie si S_1 est symétrique à S_2 en faisant varier l'angle de réflexion α ainsi que la distance de translation δ dans des intervalles choisis. L'opération est répétée un maximum de p fois jusqu'à trouver une correspondance. Dès qu'une symétrie est détectée, on incrémente la classe correspondante de l'histogramme qui représente la distribution des angles de réflexions présentes dans l'image. L'axe de symétrie, dans ce cas, est la ligne passant par les milieux des segments $(P_1^i P_2^i) \quad i = 1, \ldots, m$. La figure 3.5 (b) représente deux ensembles symétriques et les points qui définissent l'axe de symétrie.

On répète le processus pour M ensembles S_1 aléatoires. Les maximums dans la dis-

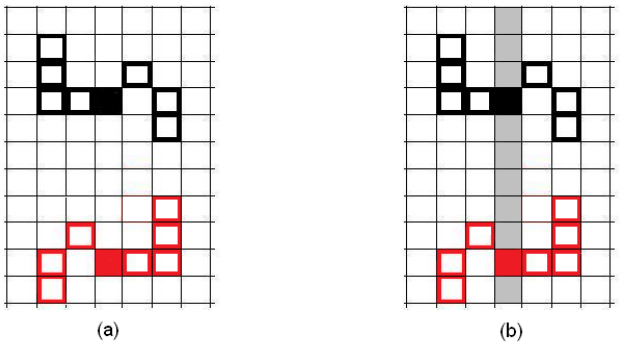

FIGURE 3.5: Détection d'un axe de symétrie potentiel. (a) Application d'une translation et d'une réflexion à une région de pixels. (b) En gris, l'axe de symétrie qui relie les points de milieu entre chaque deux pixels correspondants.

tribution des angles d'axes de réflexion vont représenter les angles des axes de symétrie. L'auteur [Ste05] a utilisé quelques images de la base Yale [GBK01] avec les paramètres $m = 12, h = 3, M = 10000, p = 100, 45° < \alpha < 135°, \varepsilon = 80$. Les images choisies représentent des personnages en face de la caméra. Les résultats montrent un seul axe de symétrie mais n'indiquent pas de quelle manière sa position a été choisie parmi tous les axes ayant la même inclinaison.

Nous allons tester cette méthode afin de vérifier si elle peut être appliquée pour l'estimation de la pose de la tête. Les images utilisées doivent être en adéquation avec le contexte d'estimation de la pose en présentant des variations de cette dernière.

3.2.2 Adéquation de la symétrie détectée au contexte de l'estimation de la pose

Nous avons implémenté la méthode de F. Stentiford [Ste05] pour la tester sur des données couvrant un spectre plus large de poses. Nous avons utilisé des images provenant de bases destinées à l'estimation de la pose de la tête. Les paramètres utilisés sont les mêmes utilisés par l'auteur sauf le seuil de la mise en correspondance $\varepsilon = 25$ au lieu de 80. Ceci est dû au fait que notre base de données est moins contrastée que les images

utilisées par l'auteur. Chose que nous pouvons constater en visualisant l'histogramme des niveaux de gris. La taille r de la région représentée par l'ensemble S_1 est $r = 20$ pixels.

Nous avons fait varier l'intervalle de l'angle α (*cf* figure 3.6). Nous pouvons remarquer que plus l'intervalle est petit, la concentration des points appartenant aux axes de symétrie détectés, devient plus visible. Nous allons donc utiliser l'intervalle $[60°, 120°]$ pour la suite des expérimentations afin d'augmenter la concentration des symétries détectées entre les régions bilatérales du visage en diminuant les symétries entre le haut et le bas du visage. La figure 3.7 illustre les résultats pour trois images avec quatre exécutions pour chaque image. Les résultats montrent plusieurs axes de symétrie correspondant à la même inclinaison. Ceci est du au fait que la méthode se base sur le maximum de la distribution des angles pour lesquels des symétries sont détectées. La localisation de l'axe de symétrie optimum n'est pas efficace car il existe, souvent, plusieurs axes de symétrie ayant la même inclinaison. Effectivement, l'histogramme des angles d'axes de réflexion se calcule pour tous les axes colinéaires. Le fait d'avoir un grand nombre d'axes colinéaires, influence le choix de l'angle car le nombre d'ensembles symétriques pour une inclinaison est le cumul des symétries correspondant à tous les axes ayant cette inclinaison.

Le nombre de symétries existantes est considérable sur les visages frontaux et le fait de prendre quelques régions aléatoires s'avère suffisant pour détecter la symétrie.

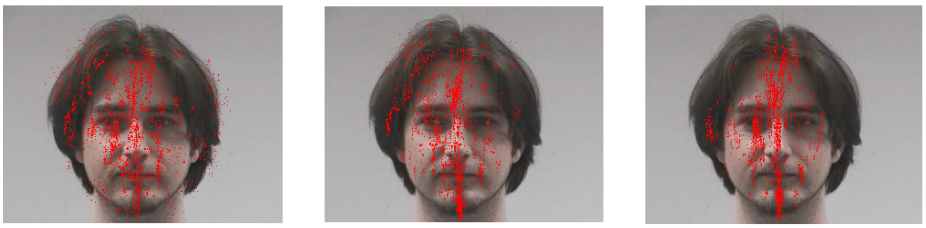

FIGURE 3.6: Points appartenant aux axes de symétrie détectés. (a) $\alpha \in [45°, 135°]$. (b)$\alpha \in [60°, 120°]$. (c)$\alpha \in [80°, 110°]$.

Vue la façon aléatoire de la constitution des régions, les résultats diffèrent d'une

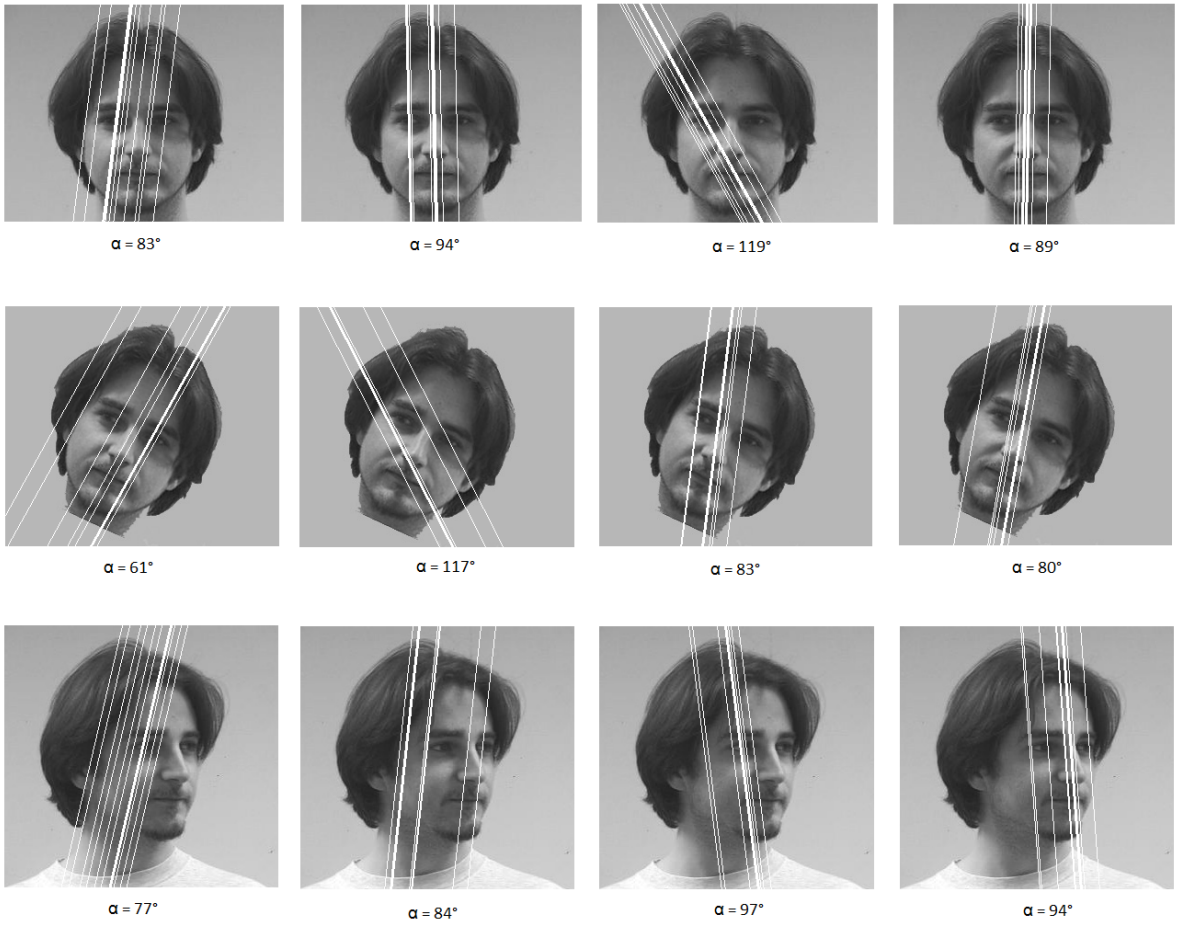

FIGURE 3.7: Symétries détectées. La première ligne représente un visage de face, dans la deuxième, une rotation dans le plan de 25° est appliquée et la troisième ligne représente un visage de profil. L'inclinaison des axes de symétrie détectés est notée sous chaque image.

exécution à une autre avec la même image. D'autre part, la constitution d'ensembles de pixels aléatoires et non-homogènes, nécessite plusieurs essais. Pour générer M ensembles, il a fallu pas moins de $5 * M$ tentatives.

Pour cela, nous avons opté pour une affectation plus structurée des régions et nous

avons adapté la méthode à la symétrie bilatérale et non la réflexion, en modifiant l'équation 3.1 qui génère une réflexion (*cf.* figure 3.5). Les positions des pixels des deux ensembles doivent êtres symétriques par rapport à un axe donné. Nous avons aussi, pallié au plus gros inconvénient de la méthode qui est le choix de l'axe optimum parmi ceux ayant la même inclinaison choisie. Notre approche est détaillée dans la section suivante.

3.2.3 Nouvel algorithme de détection de la symétrie bilatérale du visage

Notre but est de trouver la symétrie bilatérale du visage, définie par l'axe qui passe entre les deux yeux. Notre algorithme est basé sur celui de F. Stentiford [Ste05]. Il est nécessaire d'adapter l'algorithme original qui met en évidence les symétries de réflexion présentes sur une image de manière automatique mais qui manque de robustesse de part son aspect aléatoire et la difficulté de trouver l'axe optimum parmi plusieurs correspondants à l'inclinaison choisie. Nous avons, donc, cherché explicitement les symétries bilatérales du visage humain sous différentes poses.

Nous commençons par délimiter la région d'intérêt sur l'image en nous basant sur l'hypothèse qu'il existe un visage avec une symétrie bilatérale. Pour cette raison, la symétrie ne doit pas disparaitre complètement de l'image. Par exemple lorsque l'angle du mouvement pan (gauche/droite) dépasse 45°. Par conséquent, les images ne contenant pas de visage ne sont pas traitées. La détection du visage se fait via un classifieur de Haar [VJ01], nous considérons une ellipse à l'intérieur du cadre englobant le visage qui va représenter la région d'intérêt.

Nous utilisons l'intensité des pixels afin de détecter la symétrie sur l'image. La séparation entre l'avant et l'arrière plan est nécessaire pour ne pas détecter une symétrie insignifiante entre les régions du visage et celles de l'arrière plan. Notre but n'est pas de déterminer si un pixel appartient au visage ou non mais de ne pas faire correspondre un pixel de peau à un pixel de l'arrière plan. Puisqu'un pixel de peau est plus similaire à un autre pixel de peau qu'à un pixel de l'arrière plan, la séparation peut se faire pendant le processus de mise en correspondance pour la détection des symétries. Après avoir testé

plusieurs espaces de couleurs, nous utilisons l'espace RGB car il conserve les valeurs qui permettent de mettre en correspondance les pixels pour déduire les symétries locales. Les trois canaux sont utilisés de la manière suivante :

$$Si|F^R(x)–F^R(y)| < \varepsilon \quad ET \quad |F^V(x)–F^V(y)| < \varepsilon \quad ET \quad |F^B(x)–F^B(y)| < \varepsilon,$$
alors il y a correspondance entre le pixel x et le pixel y relativement au seuil ε. (3.3)

Les deux pixels sont, donc, considérés symétriques. La figure 3.8 montre des exemples de détection de pixels symétriques dans la partie supérieure du visage, avec différents espaces de couleur. Les pixels symétriques mis en évidence sont superposés sur les images et entourés. Sur l'ensemble des images que nous avons utilisées, les régions qui englobent les pixels symétriques sont mieux détectées avec l'espace RGB et les espaces de couleur de type teinte/saturation. Contrairement aux images en niveau de gris ou bien avec des espaces luminance/chrominance.

D'autre part, l'illumination influence la détection et dans certains cas, cause des erreurs. Pour cela, un pré-traitement peut être appliqué sur les images avant de commencer la détection de la symétrie. Nous appliquons l'égalisation d'histogramme sur chaque canal RGB de l'image.

Une fois la région du visage définie, le pré-traitement est appliqué suivi par l'algorithme de détection de la symétrie bilatérale du visage. L'axe de symétrie est défini par son inclinaison α par rapport à l'horizontale et par les coordonnées d'un point P lui appartenant. On fait varier α de $AngleInf$ à $AngleSup$, puis, pour chaque inclinaison, on cherche la position de l'axe de symétrie dominant en utilisant les symétries locales.

Soit α_ε le pas de variation de α de $AngleInf$ jusqu'à $AngleSup$. Nous obtenons j axes $A_{i\{\alpha_i\}}$, chacun correspond à une inclinaison avec $j = (AngleSup - AngleInf)/\alpha_\varepsilon + 1$. Ces axes sont pondérés par le nombre de symétries locales qu'ils satisfont. La figure 3.9 présente deux images exemples avec la distribution des symétries pour chaque angle d'inclinaison. Ensuite, un vote est effectué sur les m maximum de la distribution des j axes pour définir l'axe de symétrie du visage A :

$$d(C, A_{\{\alpha\}}) = min\{d(C, A_{i\{\alpha_i\}})\} i \in [1, m] \tag{3.4}$$

Dans ce qui suit, nous allons détailler comment définir l'axe de symétrie relatif à chaque inclinaison en utilisant les symétries locales présentes sur la région d'intérêt. Ainsi que la méthode que nous avons définie afin de détecter les symétries locales.

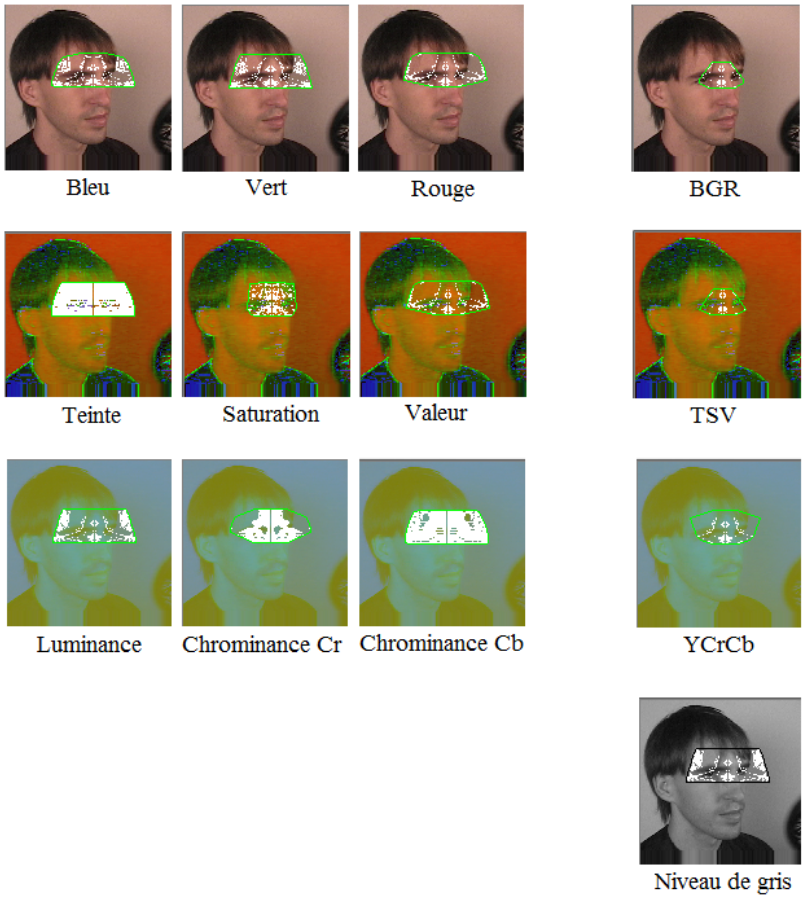

FIGURE 3.8: Détection des pixels symétriques sur la même image avec différents espaces de couleur, en utilisant tous les canaux ou bien chaque canal à part.

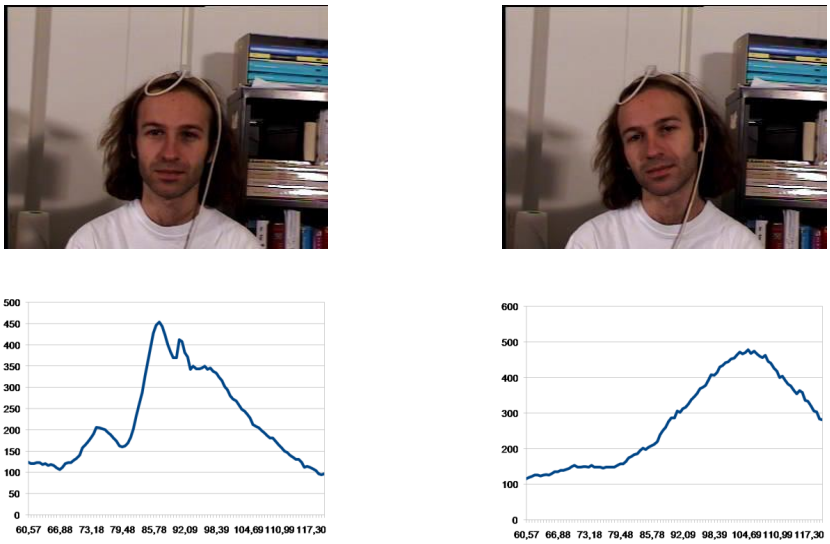

FIGURE 3.9: Distribution des symétries pour chaque angle d'inclinaison. Pour l'image de gauche, le plus grand nombre de symétries correspond à 86° et pour celle de droite le pic est à 104°. Les images appartiennent à une séquence vidéo de Boston University dataset [VG09].

3.2.3.1 Détection de l'axe de symétrie relatif à une inclinaison

La région d'intérêt est divisée en un ensemble de petits block rectangulaires que nous appelons "cellules" (*cf.* figure 3.10). Chaque axe de symétrie sur l'image est basé sur les symétries locales relatives aux cellules.

On cherche la *cellule* symétrique de chaque *cellule* de la région d'intérêt. Lorsque deux *cellules* symétriques sont détectées, nous pouvons déterminer la position de leur axe de symétrie. Cet axe local passe perpendiculairement par le milieu du segment qui relie les deux *cellules* (*cf.* figure 3.11).

Après la détection de tous les axes de symétrie $A_{i\{\alpha\}}$ avec i allant de 1 au nombre d'axes pour une inclinaison α, on choisit le meilleur axe en utilisant le même mécanisme

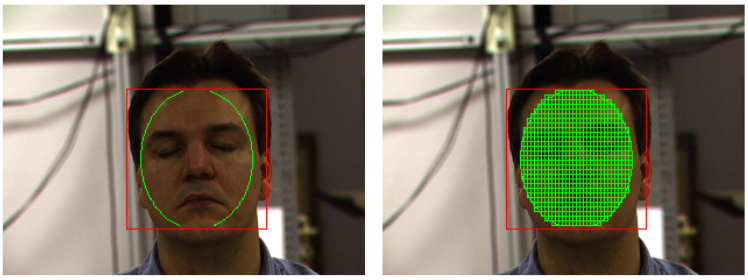

FIGURE 3.10: Détection du visage et la division de la région d'intérêt en *cellules*.

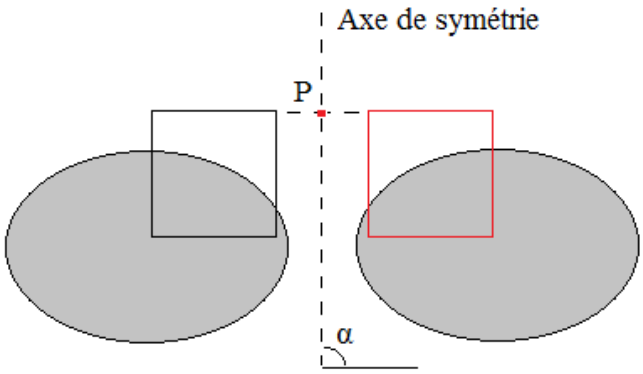

FIGURE 3.11: Une symétrie locale avec un couple de cellules symétriques.

de vote que précédemment (équation 3.4). Les symétries locales entre deux cellules très proches ou très éloignées ne sont pas comptabilisées. En effet, les constatations psychophysiques [THM95] nous indiquent que les caractéristiques trop proches ou trop éloignées de l'axe de symétrie ne contribuent pas beaucoup à la perception de la symétrie. La distance maximale dans notre cas applicatif est limitée par la région d'intérêt. La cellule miroir ne doit pas sortir du périmètre du visage et ne doit pas non plus chevaucher avec la cellule originale. L'axe de symétrie choisi pour l'inclinaison α est alors, celui qui maximisent le nombre de symétries locales tout en étant le plus proche du centre du

visage.

On peut observer la distribution des symétries locales détectées sur les images de la figure 3.12.

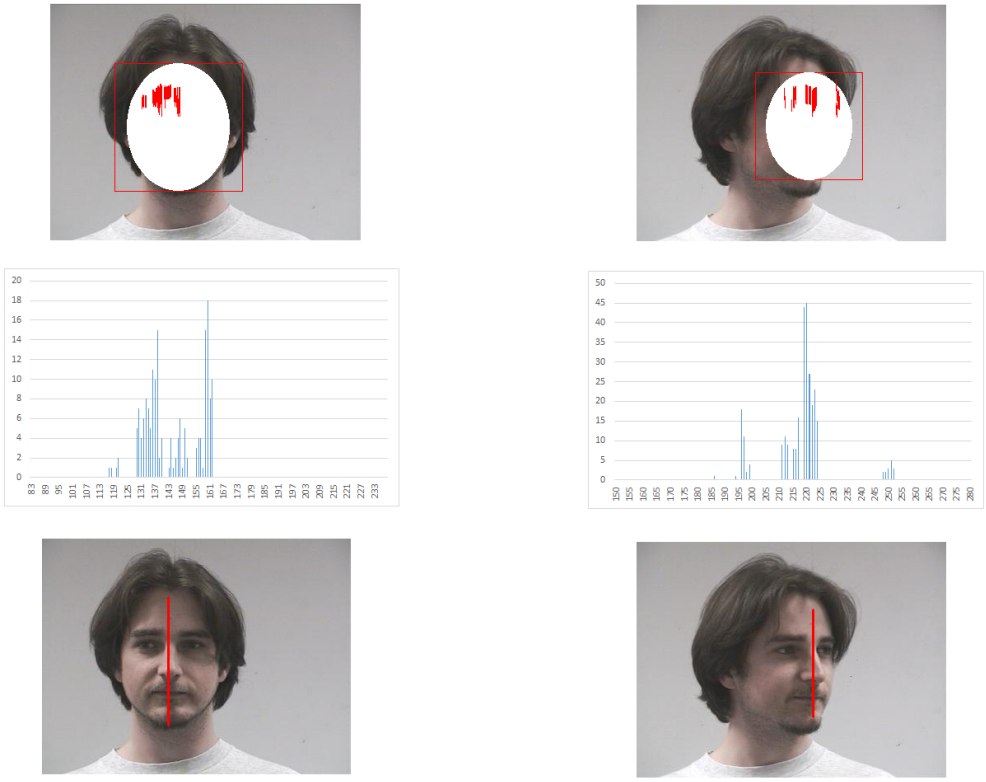

FIGURE 3.12: Axe de symétrie et distribution des symétries détectées avec un angle $\alpha = 90°$. Les deux images proviennent de Pointing dataset [GHC04]. Le premier visage est de face et le deuxième avec une rotation de 45° vers la gauche. La première ligne représente tous les axes de symétrie locaux détectés, la deuxième, la distribution de ces axes par rapport au nombre de symétries et la troisième ligne montre l'axe choisi.

3.2.3.2 Détection des symétries locales

Une symétrie locale est celle identifiée entre deux cellules. Afin de trouver les symétries locales on prend en considération un ensemble de pixels qui forment la diagonale de la cellule. On s'assure que l'ensemble n'est pas homogène, c'est à dire qu'il existe une différence dans l'intensité entre au moins deux pixels de l'ensemble.

Si l'ensemble des pixels n'est pas homogène, on teste s'il y a correspondance entre la *cellule* originale et ses *cellules* miroir relativement à α jusqu'à trouver une correspondance. La position de la *cellule* miroir est calculée via l'équation 3.5. Les coordonnées d'un pixel (x,y) après la réflexion seront (x^r,y^r). On fait varier x^r tout au long de la largeur de l'image et obtenons y^r (*cf.* figure 3.13).

$$y^r = y + ((\tan \alpha) \times (x^r - x)) \tag{3.5}$$

La correspondance de deux *cellules* est testée par rapport à tous les pixels de la diagonale. Un pixel correspond à un autre si la différence en intensité des trois canaux de couleur entre les deux pixels ne dépasse pas un certain seuil. Chaque pixel de la *cellule* originale doit correspondre à son homologue de la *cellule* miroir testée.

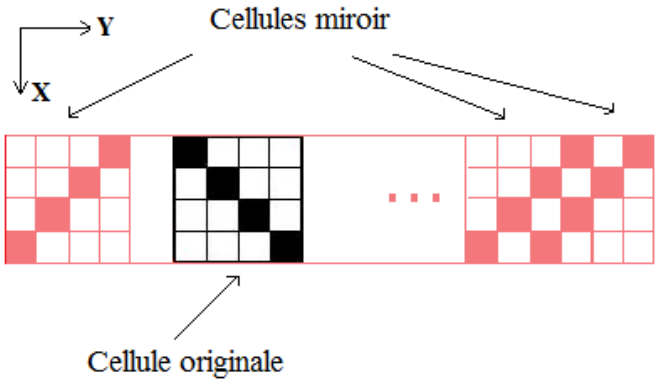

FIGURE 3.13: Une cellule et ses cellules miroirs relativement à 90°.

La méthode proposée est résumée dans l'algorithme suivant (algorithme 1) :
Les variables utilisées sont :
ROI : La région d'intérêt (l'intérieur de l'ellipse)
C : Cellule appartenant à la ROI
r : Diagonale de C
C^r : Cellule miroir de C (après réflexion)
x^r : Premier pixel appartenant à la diagonale de C^r
$Correspond$: Booléen pour indiquer une correspondance entre deux cellules.

Nous pouvons remarquer que la complexité du processus de détection de la symétrie est linéaire. Le temps d'exécution dépend de la taille de l'image ainsi que de l'intervalle $[AngleInf, AngleSup]$.

Algorithm 1 Détection de l'axe de symétrie A

$\alpha \leftarrow AngleInf$
tant que $\alpha < AngleSup$ **faire**
  tant que *non fin ROI* **faire**
    Prendre une cellule C
    si C *non-homogène* **alors**
      $x^r \leftarrow largeurROI$
      $Correspond \leftarrow faux$
      tant que $Correspond = faux\ ET\ x^r > r$ **faire**
        Définir C^r (équation 3.5)
        si $C^r \in ROI$ **alors**
          Tester correspondance entre C et C^r
          si *correspondance* **alors**
            $Correspond \leftarrow vrai$
            Sauvegarder l'axe
          finsi
        finsi
        $x^r \leftarrow x^r - 1$
      fin
      Prendre n maximum dans la distribution des A_i (i allant de 1 au nombre d'axes pour une inclinaison α)
      Voter pour un axe selon équation 3.4
    finsi
  fin
  $\alpha \leftarrow \alpha + \alpha_\varepsilon$
  fin
  Prendre n maximum dans la distribution des A_i (i allant de 1 au nombre d'inclinaisons)
  Voter pour A selon équation 3.4

3.3 Extraction dynamique des zones de symétries dans les visages

Afin d'avoir un retour d'information sur l'orientation de la tête, nous nous sommes concentrés sur la symétrie bilatérale du visage. Dans la section précédente, nous avons analysé des images pour en extraire l'axe de symétrie faciale de manière statique, sans aucune corrélation entre deux images. Cependant, nous pouvons exploiter l'information temporelle contenue dans le flux vidéo afin de réduire le temps de calcul et d'avoir plus de précision dans la détection de la symétrie. La position et l'inclinaison de l'axe de symétrie dans une trame donnée ne sont pas très différents de ceux de la trame suivante. Nous exploitons la continuité du mouvement et réduisons ainsi l'intervalle de recherche. Les informations qui peuvent être transmises d'une trame à une autre sont : la position de la région d'intérêt sur l'image, la position de l'axe de symétrie et son inclinaison. Une vérification se fait à des intervalles réguliers pour rattraper les éventuelles erreurs.

A la première trame où un visage est détecté, la position de l'axe de symétrie et son inclinaison sont calculées. Puis, pour chaque trame, si le visage n'est pas détecté ou si sa position est trop éloignée de la position à la trame précédente (i.e. le décalage est supérieur à une distance minimum d), nous considérons comme position du visage celle de la trame précédente. Sauf dans le cas où aucun visage n'est détecté depuis plus de trois trames successives car la continuité du mouvement n'est plus assurée. Dans ce cas, nous considérons une absence de visage avec symétrie bilatérale et nous passons à la trame suivante. Au fur et à mesure de l'avancement de la vidéo, l'intervalle de calcul de la position et de l'inclinaison de l'axe de symétrie est réduit au voisinage de l'axe de symétrie à la trame précédente. Une détection sur l'intervalle complet est effectuée à des intervalles réguliers afin de corriger d'éventuelles erreurs et empêcher leur propagation à travers les trames de la vidéo. La détection de l'axe de symétrie se fait via l'algorithme 1.

Le pseudocode suivant (algorithme 2) résume la manière d'exploiter le flux vidéo afin de détecter la symétrie du visage.
Les variables utilisées sont :
V : Les coordonnées du point haut gauche du cadre du visage.
P : Un point appartenant à l'axe de symétrie.
α : L'inclinaison de l'axe de symétrie.
$Cycle^{\alpha}$ et $Cycle^{p}$: Les intervalles respectifs pour calculer α et P.

Algorithm 2 Traitement d'une vidéo

répéter
  Lire une trame
  Détecter le visage
jusqu'à *Trouver un visage*;
Détecter la position et l'inclinaison de l'axe de symétrie (algorithme 1)
tant que *non fin vidéo* **faire**
  Lire une trame
  Détecter le visage
  si *Visage non détecté depuis plus de trois trames* **alors**
    $V \leftarrow null$
  sinon
    si *Visage non détecté OU décalage(*V*,*$V_{prcdent}$*)* $> d$ **alors**
      $V \leftarrow V_{prcdent}$
    sinon
      $V \leftarrow V_{courant}$
    finsi
  finsi
  si $V \neq null$ **alors**
    si $Cycle^{\alpha}$ *atteint* **alors**
      intervalle de recherche de α est $[45°, 135°]$
    sinon
      intervalle de recherche de α est *le voisinage de* $\alpha_{prcdent}$
    finsi
    si $Cycle^{p}$ *atteint* **alors**
      intervalle de recherche de P est $[1, LargeurROI]$
    sinon
      intervalle de recherche de P est *le voisinage de* $P_{prcdent}$
    finsi
    Détecter la position et l'inclinaison de l'axe de symétrie (algorithme 1)
  finsi
fin

Selon la qualité des vidéos, nous pouvons fixer les cycles de calcul des données à $Cycle^{\alpha} = 5\,trames$, $Cycle^{p} = 7\,trames$ pour une fréquence d'images de $30\,trames/$s. Ce qui était le cas pour les vidéos de la base de données Boston University [VG09] que nous avons utilisées. Les valeurs des cycles sont des nombres premiers pour que le calcul des

deux valeurs position de l'axe P et l'inclinaison α, ne se fasse jamais en même temps. Pas plus d'une seule donnée n'est calculée par trame.

Dans ce chapitre, nous avons présenté une méthode de détection de la symétrie bilatérale du visage. La détection se fait de manière globale, en utilisant la texture du visage. La mise en correspondance des différentes cellules constituant la région d'intérêt permet d'éviter l'utilisation de points caractéristiques spécifiques du visage afin de rester en adéquation avec le principe choisi pour l'estimation de la pose. Dans le chapitre suivant, nous étudierons le lien entre la symétrie et la pose de la tête. La validation de la détection de la symétrie ainsi que de l'estimation de la pose sont détaillés dans le chapitre 5.

Chapitre 4

La pose à partir de la symétrie

Sommaire

La perception humaine de la pose de la tête est basée sur deux repères : la déviation de la silhouette de la tête par rapport à la symétrie bilatérale et la déviation de l'orientation du nez par rapport à la verticale [WFLC00]. De plus, il est établi que la pose de la tête est liée à sa géométrie. Pour ces deux raisons nous supposons que la symétrie du visage est un bon indicateur sur sa géométrie et plus spécialement, sur la pose de la tête.

Puisque la symétrie procure des connaissances de haut niveau sur la géométrie du visage, nous l'utilisons afin de contribuer à la solution du problème d'estimation de la pose. Nous exploitons la texture de l'image, en plus de la géométrie du visage. Ceci dans le but de bénéficier des avantages des méthodes géométriques d'estimation de la pose (simplicité) et de celles des méthodes globales (robustesse).

Nous proposons une approche pour estimer les mouvements dans le plan (roll) et hors plan (pan et tilt) de la tête. Le roll est déduit de l'inclinaison de l'axe de symétrie. Le pan et le tilt sont estimés à partir de l'apprentissage des caractéristiques de la symétrie bilatérale du visage. Notre approche est basée sur la détection de l'axe de symétrie sur le visage, à partir duquel la région symétrique est définie. La région symétrique est délimitée à partir de la texture globale du visage et non pas en suivant des points spécifiques. Des caractéristiques sont extraites de cette région afin de discriminer les différentes poses de la tête.

Dans la suite du chapitre, nous allons étudier la relation entre la symétrie détectée sur une image 2D du visage et la pose de ce dernier sur la scène par rapport au point de prise de vue. Nous allons décrire les caractéristiques utilisées pour discriminer les différentes poses et comment ces caractéristiques sont elles utilisées.

4.1 Corrélation entre la symétrie et la pose de la tête

La pose de la tête que nous cherchons à estimer est relative au repère de la caméra utilisée pour capturer l'image. Nous allons étudier la corrélation entre cette pose et la symétrie détectée dans l'image projetée sur le plan (*cf.* figure 4.1). Cette relation réside dans la transformation induite par la projection du visage sur le plan de l'image.

La symétrie du visage est liée à la déformation de la projection du visage sur le

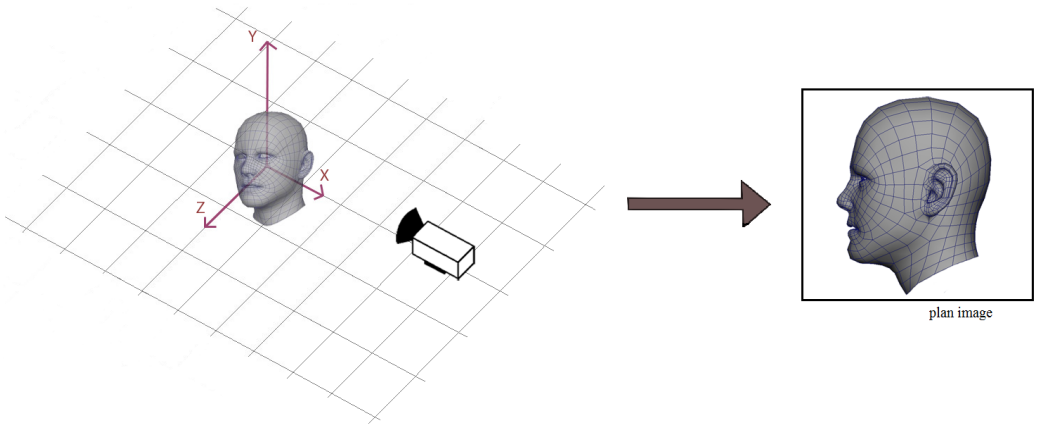

FIGURE 4.1: L'estimation de la pose de la tête est relative à la projection de cette dernière sur le plan image.

plan image. Cette déformation est le résultat d'un mouvement hors plan de la tête. La symétrie présente dans l'image va nous permettre de caractériser et par la suite, de discriminer les rotations hors plan (pan et tilt). Nous exploitons, toutefois, l'inclinaison de la symétrie détectée sur le visage pour estimer le mouvement dans le plan (roll). La figure 4.2 montre la variation de la région symétrique pour différentes poses de la tête en présence de rotations hors plan et dans le plan. L'axe de symétrie ainsi que les pixels symétriques sont superposés en vert sur les images. Nous pouvons constater que lorsque le visage est en face de la caméra, la symétrie entre ses deux parties apparait clairement et la ligne qui passe entre les deux yeux en passant par le bout du nez définit l'axe de symétrie. Cependant, lorsque la tête effectue une rotation selon les axes de son repère, cette symétrie change.

4.1.1 Le mouvement roll

Le mouvement roll est une rotation dans le plan qui n'est pas représentative de la direction du regard. Sa détection sert à redresser le visage par rapport à la verticale afin de passer à l'estimation du mouvement hors plan.

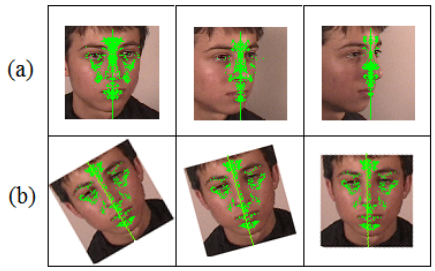

FIGURE 4.2: (a) Variation de la taille et de la forme de la région symétrique pendant un mouvement hors plan. (b) Variation de l'inclinaison de l'axe de symétrie pendant un mouvement dans le plan.

L'orientation du visage dans le plan peut être estimée pendant ou après la détection de la tête. En effet, il existe des méthodes de détection du visage invariantes à l'orientation de ce dernier [JV03],[DBD14]. Dans notre approche l'orientation dans le plan est estimée après la détection de la tête avec un algorithme basé sur celui de Viola-Jones [VJ01] (*cf.* annexe B).

L'inclinaison de l'axe de symétrie correspond à l'orientation du visage dans le plan. Le redressement du visage par rapport à cette inclinaison permet d'obtenir un visage droit (*cf.* figure 4.3).

4.1.2 Le mouvement pan

Le mouvement pan représente la rotation de la tête autour de l'axe des Y (droite/gauche). Nous analysons la taille des régions symétriques sous différentes poses en changeant l'angle pan.

Nous supposons deux points symétriques a et b sur le visage. m est le milieu du segment $[ab]$ (*cf.* figure 4.4 (a)). La projection de ces trois points sur l'image est (a_i, b_i, m_i). Lorsque (ab) est parallèle au plan image, le point m_i est le milieu de $a_i b_i$, et les segments $[a_i m_i]$ et $[m_i b_i]$ sont symétriques par rapport au point m_i comme indiqué par la figure 4.4 (a).

FIGURE 4.3: Détection de l'angle roll (a) et redressement du visage (b) afin de poursuivre l'estimation de la pose hors-plan.

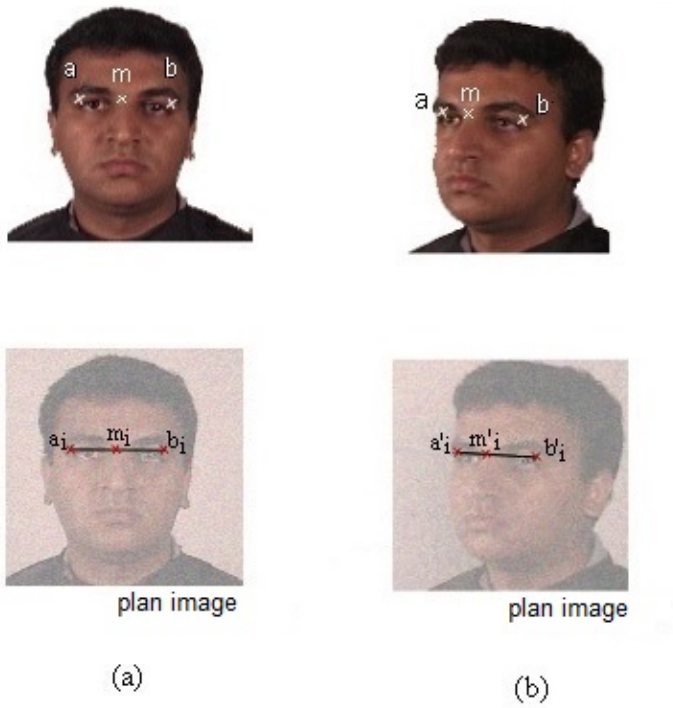

FIGURE 4.4: Projection de deux points symétriques du visage avant (a) et après (b) un mouvement pan.

Lorsque la tête effectue un mouvement pan (*cf.* figure 4.4 (b)), les points caractéristiques (a, b, m) sont projetés en (a'_i, b'_i, m'_i). La figure 4.5 montre le changement de la projection des trois points (a, b, m) après un mouvement pan.

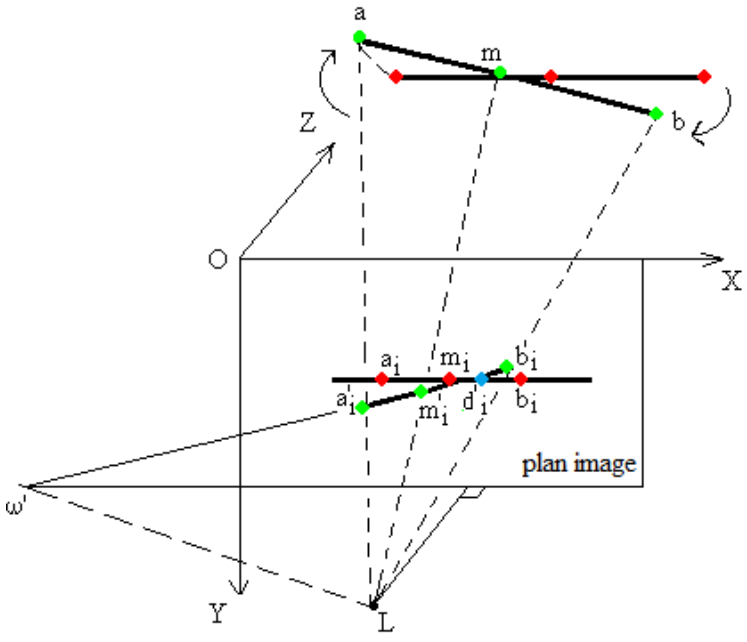

FIGURE 4.5: Changement de la projection des points caractéristiques (a, b, m) de (a_i, b_i, m_i) en (a'_i, b'_i, m'_i).

Soit ω' le point de fuite associé à la direction de (a', b') sur le plan image. Puisque la projection centrale conserve le rapport anharmonique (appelé aussi cross-ratio) [Cox03], les rapports anharmoniques de (a, b, m, ∞) et de $(a'_i, b'_i, m'_i, \omega')$ sont égaux. Nous obtenons :

$$\frac{m'_i a'_i}{m'_i b'_i} \div \frac{\omega' a'_i}{\omega' b'_i} = \frac{ma}{mb} \tag{4.1}$$

Puisque m est le milieu de $[ab]$:

$$\frac{m'_i a'_i}{m'_i b'_i} \div \frac{\omega' a'_i}{\omega' b'_i} = 1 \tag{4.2}$$

Donc, le point m'_i n'est pas le milieu de $a'_i b'_i$ et sa position dépend de celle de $a'_i b'_i$ relativement à ω'. Dans la figure 4.5, puisque a'_i est proche de ω', m'_i sera ainsi proche de a'_i plutôt que de b'_i. En raison de l'équation 4.2, nous concluons que $m'_i a'_i$ n'est pas symétrique à $m'_i b'_i$.

Si nous considérons que m est le centre de symétrie de ab, alors les pixels du segment $a'_i b'_i$ peuvent satisfaire une symétrie partielle. Cependant, dans ce cas le centre de symétrie n'est plus le milieu de $a'_i b'_i$ mais sera m'_i. La symétrie va alors concerner le segment $m'_i a'_i$ et $m'_i d'_i$ où d'_i se trouve entre m'_i et b'_i tel que $m'_i a'_i = m'_i d'_i$ (voire Figure 4.5).

Nous concluons que plus l'ampleur du mouvement pan est grande, plus la taille de la région du visage qui contient une symétrie bilatérale devient petite.

4.1.3 Le mouvement tilt

Le mouvement tilt représente la rotation de la tête autour de l'axe des X (haut/bas). Nous analysons la déformation de la projection de la région symétrique du visage en présence du mouvement tilt.

Soit (δ) une droite verticale sur le visage, A et B deux points du visage appartenant à (δ). Leurs projections sur le plan image sont a_i et b_i (*cf.* figure 4.6). Lorsque la tête effectue un mouvement tilt, le segment $(a_i b_i)$ devient $(a'_i b'_i)$ de taille plus petite. Plus l'angle du mouvement est grand, plus la distance entre la projection des deux points est petite. (*cf.* figure 4.7).

Nous considérons que les points A et B délimitent la région symétrique. La hauteur de cette région alors, diminue au fur et à mesure que le mouvement tilt prend de l'ampleur.

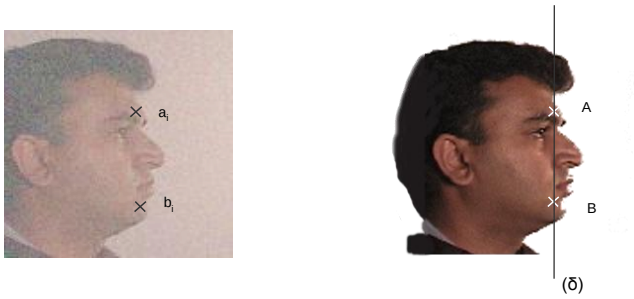

FIGURE 4.6: Projection d'un axe vertical appartenant au visage sur le plan image.

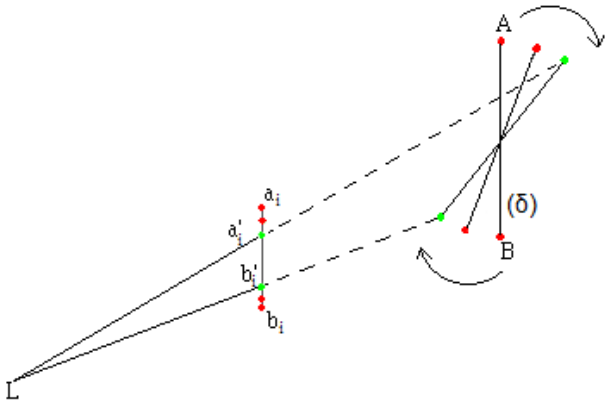

FIGURE 4.7: Diminution de la taille de la projection du segment $[AB]$ après le mouvement tilt vers le haut.

4.2 Caractérisation de la symétrie

Dans la section précédente, nous avons montré comment la symétrie présente sur le visage indiquait la déformation de la projection de la tête due à une rotation hors plan. Nous allons extraire des caractéristiques afin de discriminer les poses de la tête dues à des mouvements hors plan (pan et tilt). La rotation de la tête dans le plan sera déduite à partir de l'inclinaison de l'axe de symétrie.

Pour ce faire, nous détectons l'axe de symétrie (*cf.* paragraphe 3.2.3) ainsi que la région symétrique. La zone de symétrie n'est autre que l'enveloppe convexe qui englobe les pixels symétriques appartenant à la région du haut du visage. Le haut du visage est sélectionné au lieu de tout le visage car il est plus affecté par le mouvement de la tête que le visage entier. Le changement de la forme et de la taille de la région symétrique du haut du visage est plus significatif que le changement relevé sur tout le visage. Le haut du visage contient donc, plus d'information symétrique.

Afin de définir les pixels symétriques qui vont être utilisés pour délimiter la région symétrique, nous testons la correspondance entre chaque pixel et son symétrique, selon l'axe détecté. Contrairement à l'étape de détection de l'axe de symétrie, tous les pixels de la région d'intérêt sont considérés sans exclure les ensembles homogènes. Nous avons imposé la condition d'utiliser des ensembles non homogènes dans l'étape de détection de l'axe de symétrie afin de pouvoir exclure les symétries insignifiantes sur les régions uniformes. Dans cette étape de définition de la région symétrique, les pixels sont testés un par un en utilisant toute la texture du visage. De cette manière, la présence de points ou de régions locales spécifiques du visage, n'est pas nécessaire. Nous calculons la position d'un pixel symétrique à un pixel donné en utilisant l'équation 3.5. Si la différence en intensité entre deux pixels est plus grande qu'un certain seuil, alors les deux pixels ne sont pas considérés comme étant symétriques.

La figure 4.8 illustre les étapes qui mènent à la détection de la région symétrique. La première étape étant la détection du visage et la définition de la zone de recherche de l'axe de symétrie (*c.f* figure 4.8 (a)). Ensuite, la détection de l'axe de symétrie (*c.f* figure 4.8 (b)) puis la détection des points symétriques par rapport à cet axe et de la région symétrique (*c.f* figure 4.8 (c)). La zone de symétrie est définie comme étant l'enveloppe qui englobe les pixels symétriques du haut du visage.

Nous pouvons observer sur la figure 4.9 des exemples qui représentent des zones de symétrie extraites à partir d'images de la base de poses que nous avons synthétisée (les détails sur la construction de la base synthétique sont dans le chapitre suivant). A partir du constat que les régions symétriques correspondant aux mêmes poses, possèdent la même forme géométrique, nous avons défini deux vecteurs de caractéristiques à partir de

ces régions.

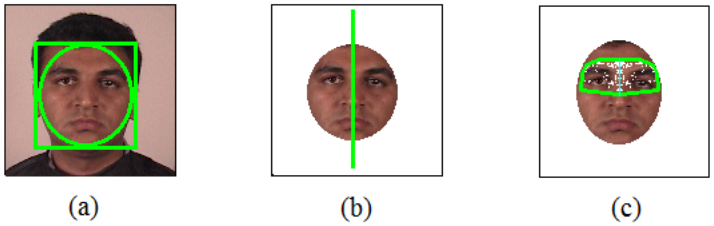

FIGURE 4.8: Les étapes allant de la détection du visage jusqu'à la définition de la région symétrique.

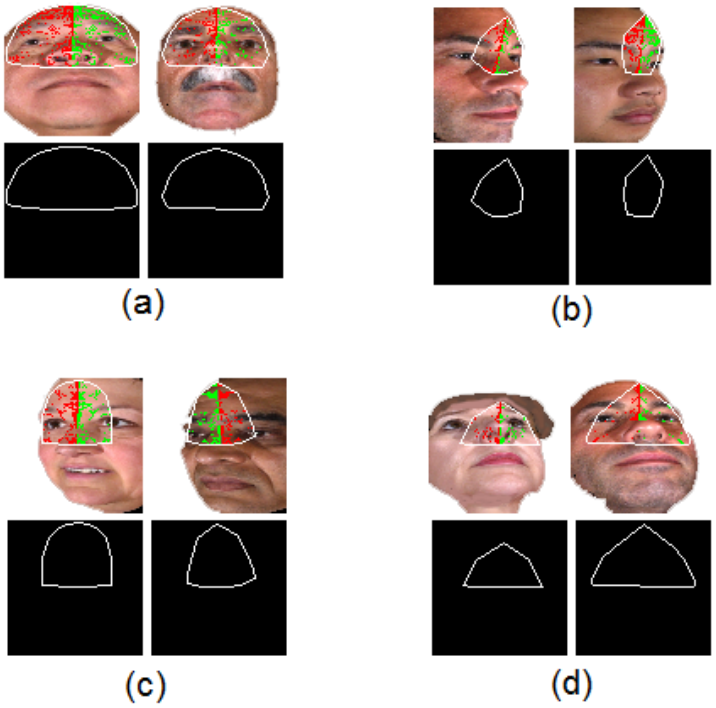

FIGURE 4.9: Les zones de symétrie correspondantes aux poses : (a) pan= 0°, tilt= 30°. (b) pan= 45°, tilt= 0°. (c) pan= −15°, tilt= 0°. (d) pan= 15°, tilt= 30°.

4.2.1 Caractéristiques basées sur la taille de la région symétrique (Vecteur TRS)

Le vecteur de caractéristiques TRS caractérise principalement le mouvement pan. La principale caractéristique qui nous permet d'estimer le mouvement pan est la taille de la région symétrique (*cf.* paragraphe 4.1.2). Nous avons relevé plusieurs mesures concernant la région symétrique (*cf.* figure 4.10). Après avoir testé la pertinence de ces mesures à discriminer la pose de la tête, en utilisant différents types de distances, nous avons retenu deux caractéristiques définies par des distances euclidiennes :

- la largeur de l'enveloppe qui englobe les pixels symétriques. Elle est définie par la distance euclidienne entre les deux pixels symétriques les plus éloignés de l'axe de symétrie.
- la distance moyenne qui sépare les pixels symétriques de l'axe de symétrie.

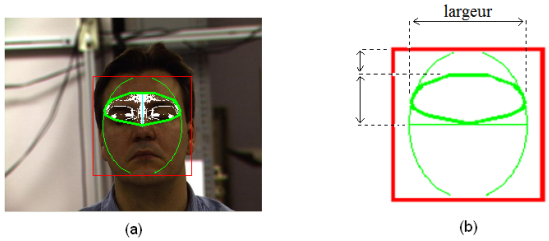

FIGURE 4.10: Extraction des caractéristiques. (a) Définition de l'enveloppe qui contient les pixels symétriques. (b) Différentes mesures relatives à la région symétrique.

Cependant, le vecteur de caractéristiques TRS est invariable à la réflexion. En effet, une pose du côté droit est symétrique à sa correspondante du côté gauche. Nous regroupons donc, les poses de gauche et de droite ayant la même valeur dans la même pose. La figure 4.11 présente un exemple qui regroupe des visages de gauche et de droite dans la même pose.

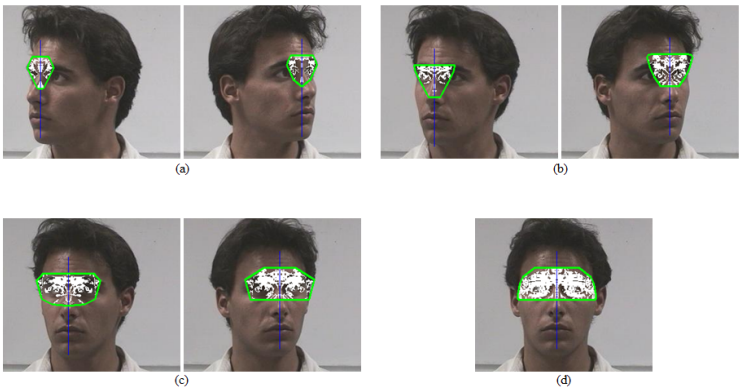

FIGURE 4.11: Regroupement des poses de gauche et de droite. (a) $-45°$ et $45°$, (b) $-30°$ et $30°$, (c) $-15°$ et $15°$, (d) $0°$

Différencier entre les poses de gauche et de droite

Nous avons besoin de différencier entre un visage de gauche et son correspondant à droite lors de l'estimation de la pose de la tête. Ceci dans le cas de l'utilisation du vecteur TRS. Pour ce faire, nous utilisons la différence en intensité entre la peau et l'arrière plan. Notre hypothèse est qu'un pixel du visage est plus similaire à un autre pixel du visage qu'à un pixel de l'arrière plan.

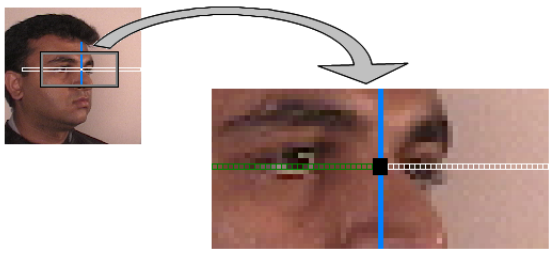

FIGURE 4.12: Différence entre les poses de gauche et de droite : Le pixel référence est comparé à ceux se trouvant à sa gauche (en blanc) ainsi qu'à ceux à droite (en vert).

Nous prenons un pixel référence situé à l'intersection de l'axe de symétrie avec la ligne

située à 1/4 de la hauteur du visage à partir du haut, afin de s'assurer que ce pixel fait partie du visage. Nous calculons l'intensité moyenne des pixels entourant et considérons cette valeur comme référence. Si l'axe de symétrie est plus proche du contour gauche (resp. contour droit) du visage, alors le visage est orienté vers la gauche (resp. droite). Nous calculons deux valeurs : la différence entre la valeur référence et l'intensité moyenne des pixels sur la gauche de l'axe et la même différence avec le côté droit de l'axe de symétrie.Si la différence est plus grande du côté gauche (resp. droit), nous concluons que le visage sur l'image est orienté vers la gauche (resp. droite). La figure 4.12 montre la différence entre les pixels de peau et ceux de l'arrière plan situés à gauche du visage, ce qui signifie que la pose est orientée vers la gauche.

La différence des moyennes ne donne pas une égalité parfaite même si le visage est frontal car un visage n'est jamais parfaitement symétrique. Pour cette raison, lorsque le résultat de la classification est 0° (c'est à dire une pose frontale), la différentiation entre la droite et la gauche n'est pas prise en considération, seul le résultat de la classification est pris en compte.

Avec cette méthode, nous déterminons de quel côté la pose est orientée. Cette information est combinée avec la prédiction afin d'obtenir la pose de la tête.

4.2.2 Caractéristiques basées sur l'orientation de la région symétrique (Vecteur ORS)

Le vecteur de caractéristiques ORS que nous avons défini, ne privilégie pas un degré de liberté. Contrairement au premier, il vise à caractériser l'ensemble des poses. Ses caractéristiques concernent la forme globale de la région symétrique. Nous utilisons un descripteur qui a fait ses preuves dans la détection d'objets et qui est l'histogramme de l'orientation du gradient [DT05]. La forme de la région symétrique est décrite par les directions de ses contours afin de discriminer les régions et par conséquent les poses.

Après la détection de la région symétrique, nous considérons une image binaire reproduisant sa silhouette. Pour caractériser cette silhouette et construire l'histogramme des orientations, nous calculons les dérivées selon les axes vertical et horizontal en utilisant le

filtre de Sobel. Nous obtenons deux images du gradient dx et dy. L'orientation de chaque point (x, y) du contour est alors,

$O(x, y) = \arctan(dy[x, y], dx[x, y])$

Nous construisons un histogramme de 18 bins pour les orientations de 0° à 180° afin de couvrir toute la forme du contour. Le vecteur de caractéristiques contient alors 18 attributs.

Ces deux vecteurs de caractéristiques vont être utilisés pour apprendre la pose de la tête. La classification ainsi que la régression sont testées afin de valider puis déployer le système d'estimation de la pose. Les tests seront effectués sur des bases d'images et de vidéos publiques.

4.3 Apprentissage de la pose

Afin de valider les caractéristiques extraites de la symétrie et vérifier leur efficacité pour l'estimation de la pose hors plan de la tête, nous construisons un modèle de poses par l'apprentissage. La pose est ensuite prédite et validée sur les données test.

Nous construisons un modèle avec des vecteurs de caractéristiques issus d'images de plusieurs personnes enregistrées sous différentes poses. L'ensemble d'images de poses utilisées pour l'apprentissage représente les angles pour lesquels l'axe de symétrie est proprement détecté. Il y a deux manières d'identifier les poses de la tête : une pose peut représenter un seul degré de liberté (pan ou tilt) ou rassembler les deux (pan et tilt).

Dans ce qui suit, nous détaillons les deux approches (poses simples et combinées) et présentons une architecture par routeur pour fusionner plusieurs modèles afin d'estimer les deux degrés de liberté, le pan et le tilt. Nous présentons aussi les méthodes d'apprentissage utilisées.

4.3.1 Poses simples

Lorsque nous voulons estimer chaque degré de liberté indépendamment de l'autre, chaque donnée utilisée pour l'apprentissage représente une pose selon un axe (l'axe des

X pour le tilt et l'axe des Y pour le pan). La figure 4.13 montre un exemple de poses discrètes avec l'intervalle $[-30°, 30°]$ et un pas égal à 15°. Pour le mouvement pan, les poses sont de droite à gauche et pour le tilt, de bas en haut.

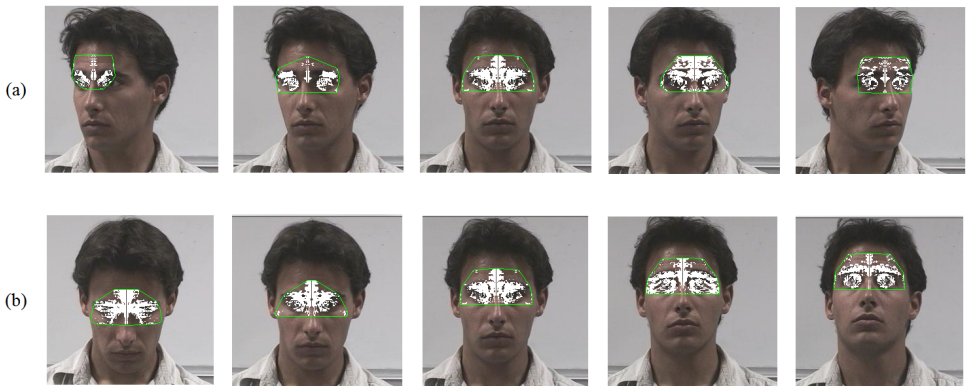

FIGURE 4.13: Configuration de poses simples, (a) pour le mouvement pan et (b) pour le mouvement tilt.

La pose de la tête est représentée par une classe dans le cas de l'apprentissage par classification. Les valeurs sont alors, discrètes. Par contre, avec l'apprentissage par régression, les valeurs prédites sont continues. Dans le cas discret, si l'intervalle du mouvement est $[Angle^{inf}, Angle^{sup}]$ et le pas entre deux poses est $Angle^{diff}$, alors le nombre de classes est $(Angle^{sup} - Angle^{inf})/Angle^{diff} + 1$. Dans nos expérimentations, nous avons exploré un nombre variable de poses (*cf.* chapitre 5).

4.3.2 Poses combinées

Il est possible de considérer une pose qui combine deux mouvements, l'horizontal et le vertical. Dans ce cas, chaque pose est une codification d'un pan et d'un tilt. Dans le cas discret, nous obtenons un nombre de classes égal à $NbrPosePan \times NbrPoseTilt$. La figure 4.14 montre un exemple de configuration de poses en considérant deux degrés de liberté.

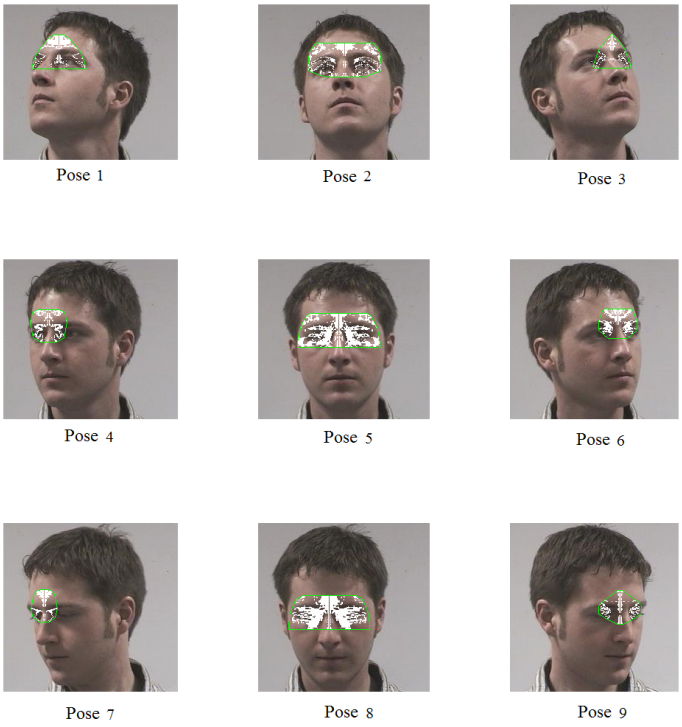

FIGURE 4.14: Configuration de poses combinées selon deux degrés de liberté (pan et tilt).

4.3.3 Poses routées

Afin d'estimer deux degrés de liberté moyennant des modèles séparés tout en exploitant le plus robuste pour guider le processus, nous mettons en place une architecture avec routeur (*cf.* figure 4.15). En effet, un modèle peut servir de routeur pour choisir entre une multitude d'autres modèles qu'ils soient classifieurs ou régresseurs. Pour chaque tranche de poses issue de la première estimation (premier degré de liberté), un modèle est construit pour continuer la prédiction et estimer le deuxième degré de liberté.

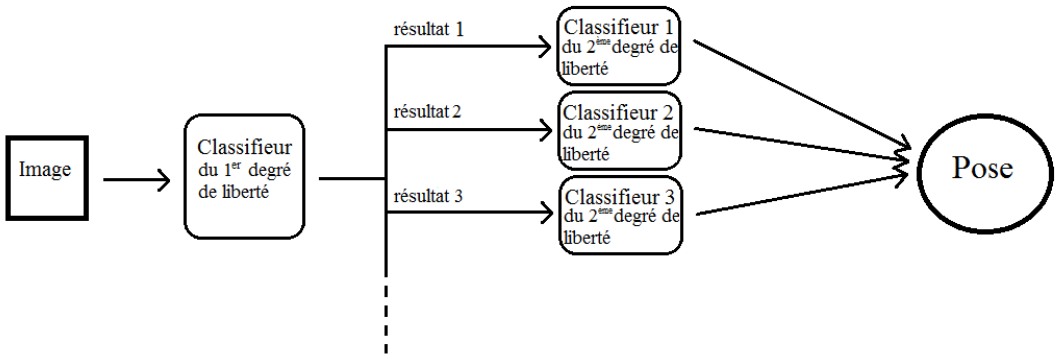

FIGURE 4.15: Architecture d'apprentissage avec routeur.

4.3.4 Méthodes d'apprentissage utilisées

La construction du modèle de pose peut se faire via la classification ou la régression. La classification des caractéristiques est un moyen d'apprentissage robuste. Dans ce cas de figure, la pose est représentée par une classe et prend des valeurs discrètes. Tandis que la régression prédit une pose continue. Deux classifieurs, ainsi que la régression linéaire, semblent convenir aux caractéristiques de la symétrie (définies dans le paragraphe 4.2). La classification par les séparateurs à vaste marge (SVM) est parvenue à discriminer les poses de la tête moyennant les caractéristiques de la symétrie faciale et atteint des taux de classification élevés. Cependant les arbres de classification sont une alternative efficace. Nous avons donc, introduit l'utilisation des arbres de décision alternatifs (LADTree) qui sont rapides et conviennent mieux lorsque le nombre d'attributs est petit comme tel est le cas avec le vecteur de caractéristiques TRS qui est basé sur la taille de la région symétrique (*cf.* paragraphe 4.2).

Nous présentons dans ce qui suit les classifieurs utilisés ainsi que la régression linéaire qui reste un moyen naturel pour étudier le mouvement de la tête.

4.3.4.1 L'arbre de décision alternatif

L'arbre de décision binaire a été inventé par Breiman et al. [BFOS84] qui l'appelle *Classification and Regression Tree (CART)*. L'arbre est constitué de nœuds internes et de feuilles. Les nœuds sont étiquetés par des tests et les feuilles par une classe. Le but est de définir une mesure qui est minimale lorsque la plupart des valeurs d'un nœud sont de la même classe. Les échantillons sont ainsi séparés selon leur classes. Il existe deux sortes d'arbres de décision. Le premier est construit dans un ordre de parcours en profondeur (depth-first), c'est l'arbre de décision standard. La deuxième manière est de construire l'arbre dans l'ordre du meilleur en premier (best-first). Dans ce cas, on choisit la "meilleure" branche à développer.

L'application du boosting aux arbres de décision leur donne plus de précision dans la classification. Le boosting dans l'apprentissage supervisé consiste à combiner plusieurs classifieurs. Le principe est d'apprendre itérativement plusieurs classifieurs faibles et de les ajouter à un classifieur final plus puissant. Les exemples sont pondérés selon qu'ils soient bien ou mal classifiés. A chaque itération, les exemples sont re-pondérés afin de focaliser sur ceux qui ont été mal classifiés. Il existe plusieurs algorithmes de boosting, tels que Adaboost [FS97] et LogitBoost [FHT98]. La combinaison de Adaboost avec les arbres de décision produit un très bon classifieur (par exemple, le progiciel C5.0). Cependant, ces classifieurs, en forme de vote majoritaire sur un nombre d'arbres de décision, sont, le plus souvent, larges et complexes ce qui les rend difficiles à interpréter. Pour toutes ces raisons et pour augmenter les performances de notre prédiction de la pose de la tête, nous utilisons l'arbre de décision alternatif (LADTree) [HPK$^+$01] qui est basé sur l'algorithme de boosting logistique LogitBoost et qui génère des règles généralement de petite taille, et donc, faciles à interpréter. LogitBoost a l'avantage sur AdaBoost d'être applicable directement sur des problèmes multi-classes. Un seul arbre est construit en estimant les valeurs de prédiction de toutes les classes simultanément.

L'arbre alterne entre nœuds de prédiction et nœuds de décision. La racine est un nœud de prédiction et contient une valeur pour chaque classe. Les valeurs de prédiction sont utilisées comme une mesure de confiance pour la prédiction. Le premier niveau de

l'arbre contient des nœuds de décision basés sur les valeurs des attributs du vecteur de caractéristiques pour chaque classe et ainsi de suite jusqu'aux feuilles. La somme des valeurs de prédiction (ou mesures de confidence) croisées en suivant tous les chemins pour lesquels tous les nœuds de décision sont à "vrai", est utilisée pour classifier une instance donnée. La classe ayant la plus grande valeur est la classe prédite.

L'arbre construit

Afin de déterminer l'orientation de la tête, un arbre de décision alternatif est construit en utilisant les caractéristiques extraites de la région symétrique relativement à l'amplitude de la pose. Chaque classe correspond à une pose discrète. La racine contient des valeurs nulles comme prédiction pour toutes les classes.

La figure 4.16 présente un exemple de classifieur construit pour prédire le mouvement pan sur la base FacePix [BGK+02] en utilisant le vecteur de caractéristiques TRS défini plus haut (*cf.* paragraphe 4.2) qui a deux attributs : "largeur" et "distance". Le nombre de classes choisies pour cet exemple est égal à 4 : $0°, 15°, 30°et45°$, afin d'obtenir un arbre de petite taille que nous pouvons schématiser.

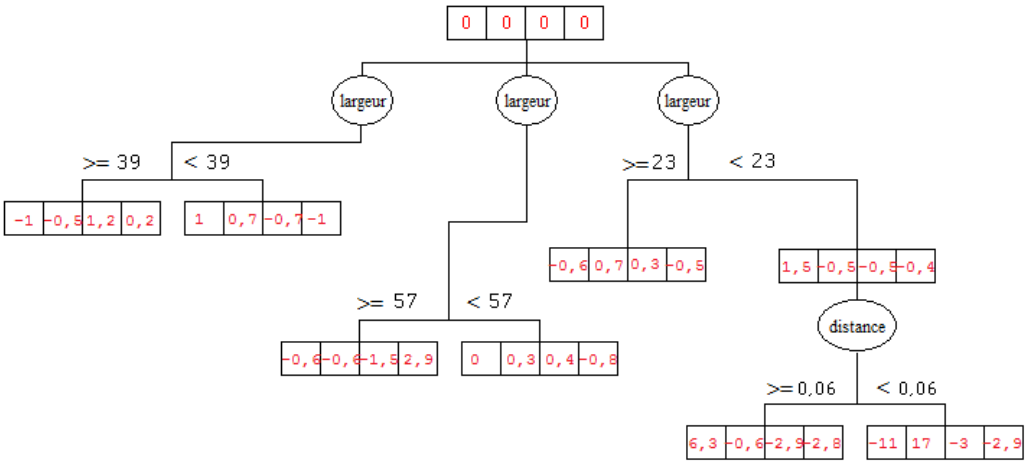

FIGURE 4.16: L'arbre de décision alternatif construit pour 4 poses discrètes moyennant le vecteur de caractéristiques TRS.

4.3.4.2 Les SVMs

Les séparateurs à vastes marges [BGV92] sont une classe d'algorithme d'apprentissage binaire basé sur la théorie de *Vapnik-Chervonenkis*. Le principe est de maximiser la marge qui sépare les données en deux classes distinctes. Le problème est formulé comme un problème d'optimisation quadratique pour trouver la frontière de séparation maximale à partir d'un ensemble d'apprentissage $\{(x1, y1), (x2, y2), ...\}$. La résolution de ce problème revient à construire une fonction f qui fait correspondre à chaque entrée x une sortie y, telle que $P(f(x) \neq y)$ soit minimale. Les données qui se trouvent sur la limite de la classe forment les vecteurs de support. L'hyperplan qui maximise la marge avec les vecteurs de supports est l'hyperplan optimal. La figure 4.17 représente un cas où le problème est linéairement séparable.

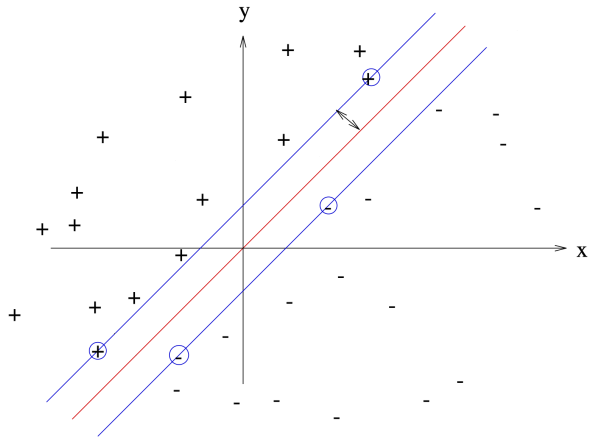

FIGURE 4.17: Séparateurs à vastes marges. Les vecteurs de supports sont en bleu et l'hyperplan optimal en rouge.

Le SVM construit

L'espace des données est transformé en un espace de plus grande dimension, où il est plus probable de trouver une séparation linéaire des données. La transformation se fait via une fonction noyau (*kernel* en anglais) qui représente un produit scalaire $< x_i, x_j >$ dans

l'espace de plus grande dimension appelé aussi espace de caractéristiques. Nous avons utilisé un SVM à noyau linéaire où la fonction noyau est $k(x_i, x_j) = x_i * x_j$.

Les séparateurs à vaste marge peuvent être combinés en un ensemble de *un contre un* ou *un contre tous* classifieurs afin de résoudre les problèmes multi-classes. Nous avons utilisé la stratégie *un contre un* qui consiste à transformer un problème à n classes en $n(n-1)/2$ classifieurs binaires où chaque classe est comparée à une autre. La classe gagnante est décidée par un vote majoritaire.

La figure 4.18 montre les limites des données de l'exemple utilisé dans le paragraphe précédent, où la classe représente l'angle pan, le vecteur de caractéristiques utilisé est le vecteur TRS et le nombre de classes est égal à 4. Nous avons choisi le vecteur TRS car il est de dimension 2 et est donc, facilement présentable en 2D.

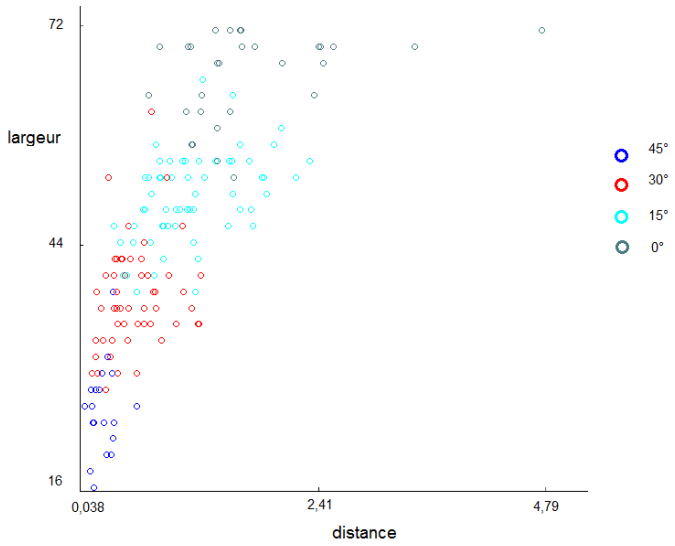

FIGURE 4.18: Limites des caractéristiques du vecteur TRS séparées en quatre classes.

4.3.4.3 La régression linéaire

Puisque un mouvement de tête est continu, la relation entre l'espace des caractéristiques et la pose de la tête peut être utilisée pour construire un modèle de régression. Le problème d'estimation de la pose se ramène alors, à un problème de régression linéaire simple ou multiple. De manière générale, la régression permet d'interpoler ou de prédire la valeur des données en cherchant une relation entre elles. Dans notre cas, il s'agit de suivre la déformation de la région symétrique du visage selon la pose de la tête (*cf.* figure 4.19).

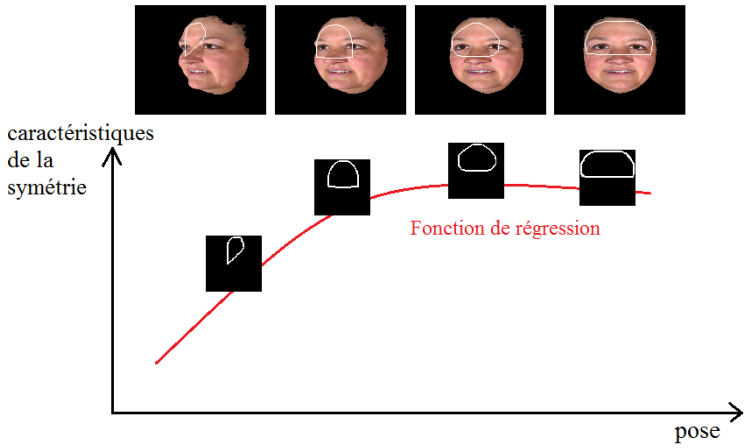

FIGURE 4.19: Suivi des caractéristiques de la symétrie faciale dans la régression linéaire.

Avec n caractéristiques symétriques $(S_1, S_2, ..., S_n)$, le calcul de la réponse P revient à résoudre un problème à $n+1$ dimensions, tel que :

$P = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + ... + \beta_n S_n + E$,

Les écarts entre la pose réelle P de chaque donnée de l'échantillon d'apprentissage et sa réponse P sont appelés résidus. Les paramètres $(\beta_0, \beta_1, ..., \beta_n, E)$ sont définis de manière à minimiser la somme des carrés des résidus. Ceci est le critère d'évaluation qui nous permet d'obtenir une solution unique à notre problème.

Nous avons proposé l'utilisation de la symétrie bilatérale du visage afin de caractériser la déformation de la projection du visage sur le plan image. Cette déformation est due à des mouvements hors plan de la tête. Dans ce cas de mouvements, la symétrie présente sur l'image change et devient partielle. Nous avons extrait des caractéristiques géométriques à partir de cette symétrie afin d'estimer la pose de la tête. Nous avons, aussi, cité les méthodes d'apprentissage que nous allons appliquer.
Dans le chapitre suivant, nous présentons les résultats obtenus.

Chapitre 5

Validation

Sommaire

Dans ce chapitre nous allons valider les résultats de nos deux principales contributions, la détection de la symétrie et l'estimation de la pose. Ces deux étapes du processus d'estimation de la pose de la tête sont validées séparément. La détection de la symétrie est testée en présence de rotations de la tête hors plan et dans le plan. Les paramètres qui permettent de détecter le meilleur axe de symétrie sont retenus afin d'être utilisés dans l'étape d'apprentissage. L'apprentissage des caractéristiques de symétrie, qui mène à l'estimation de la pose de la tête, est validée sur des bases d'images et de vidéos de l'état de l'art. Différentes conditions sont testées, comme le changement d'illumination et d'expression.

5.1 Bases de données utilisées

Il existe des bases d'images et de vidéos publiques dédiées à l'estimation de la pose. Nous utilisons des bases qui offrent une variabilité dans les conditions de prise de vue afin de valider notre approche. Nous avons, aussi, synthétisé une base d'images afin d'enrichir les configurations de poses dont nous disposons. Dans ce qui suit, nous présentons ces bases avec les caractéristiques de chacune d'elles. Les détails des protocoles de capture utilisés sont en annexe (*cf.* annexe A).

5.1.1 Base de données FacePix

La base FacePix [BGK+02] est constituée de trois ensembles d'images de visages : pose variable, éclairage variable sur fond noir et éclairage variable sur fond clair. Les ensembles de l'éclairage variable ne contiennent que des visages frontaux. Nous utilisons, donc, l'ensemble de poses variables qui est composé de 181 images de 30 personnes différentes. Les orientations des visages suivent un mouvement horizontal dans l'intervalle $\pm 90°$ à un pas de $1°$. Parmi les 181 poses, nous utilisons les poses qui varient de $-45°$ à $+45°$ car lorsque cet intervalle est dépassé, la symétrie bilatérale du visage disparait de l'image plane. La base FacePix a été capturée avec un mécanisme très précis. Elle nous a permis de tester plusieurs configurations de poses et surtout à construire un modèle robuste et

exploitable.

5.1.2 Base de données CMU PIE

La base CMU *Pose, Illumination and Expression* [SBB03] contient des images de 68 personnes sous différentes poses avec un intervalle de 22.5° entre les poses. Dans nos expérimentations, nous utilisons trois ensembles :

- CMU Expression, contient des visages avec 4 expressions différentes : neutre, yeux fermés, souriant et pour certaines personnes, avec des lunettes. Parmi les poses de cet ensemble, nous choisissons celles allant de $-45°$ à $+45°$.
- CMU Lighting, contient des images enregistrées sous un éclairage variable avec 21 orientations de flash.
- CMU Talking, contient des images de personnes en train de parler.

La figure 5.1 montre des images des trois ensembles de la base CMU PIE. Le tableau 5.1 résume les caractéristiques de chaque ensemble de la base. La base CMU PIE nous a permis de tester notre approche dans des conditions de changement d'expression et d'illumination.

Tableau 5.1: Caractéristiques de la base CMU PIE.

	nombre d'images par personne	nombre de poses	taille du visage
CMU Expression	20 / 15	5	re-dimensionné à 80×80
CMU Lighting	48	2	80×80
CMU Talking	120	2	80×80
CMU PIE (les trois ensembles)	188 / 183	5	80×80

5.1.3 Base de données Boston University

La base de vidéos *Boston University head pose* (BU) [VG09] consiste en 45 séquences vidéo. Cinq participants effectuent des mouvements aléatoires de la tête avec un émetteur

FIGURE 5.1: Exemple d'images de la base CMU PIE. La première ligne représente des images de l'ensemble *Expression*, la deuxième celles de l'ensemble *Lighting* et la troisième représente les images de l'ensemble *Talking*.

magnétique placé sur cette dernière. Les orientations selon les trois degrés de liberté sont ainsi enregistrées. Neuf séquences sont capturées pour chaque participant.

5.1.4 Base de données synthétique

Nous avons synthétisé une base d'images afin de disposer de visages sous différentes poses avec des angles pan et tilt précis. En effet, notre méthode requiert une bonne précision dans les données qui servent à créer un modèle de pose, afin d'assurer la robustesse du modèle et qu'il puisse être déployé sur des données réelles.

Nous avons créé la base à partir de la base d'images Texas 3DFRD [GMCB10]. Nous avons synthétisé un ensemble d'images 2D en faisant varier la pose de la tête. Texas 3DFRD fournit les modèles 3D ainsi que la texture à partir des quels OpenGL nous permet de synthétiser les images. Les poses varient comme ceci : l'angle pan va de $+45°$ à $-45°$ avec $5°$ d'intervalle entre chaque pose ce qui donne 19 orientations du pan et 3 angles tilt : $+30°$, $-30°$ et $0°$. Le nombre de poses totales combinées est de 57 pour

chaque personne. Le nombre de personnes est 37. La figure 5.2 présente quelques images d'une personne de la base synthétique.

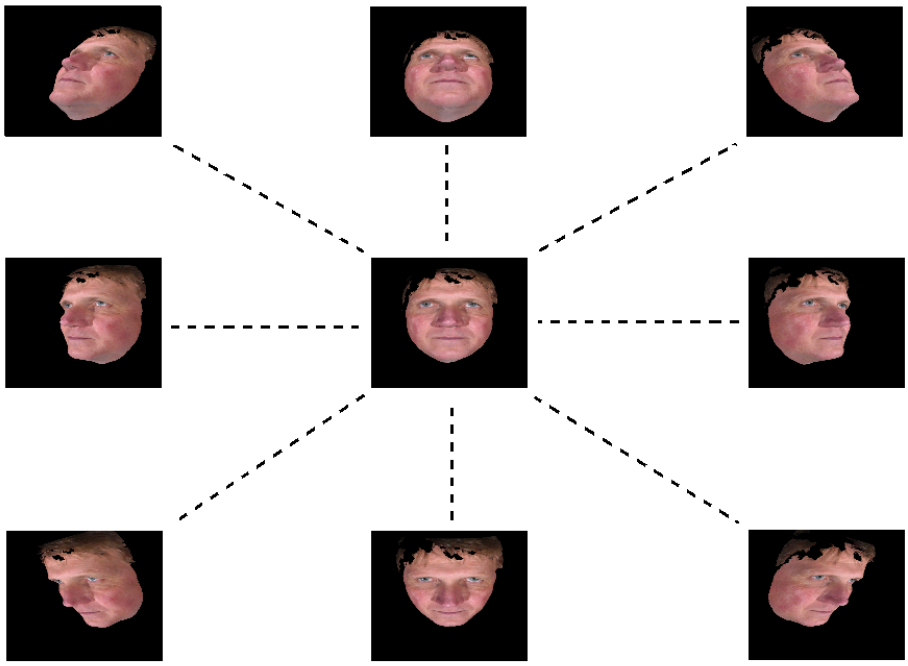

FIGURE 5.2: Exemple d'images d'une personne de la base synthétique.

5.2 Détection de la symétrie

Une bonne détection de la symétrie est nécessaire pour l'étape suivante qui est l'estimation de la pose de la tête. Nous testons l'algorithme de détection de l'axe de symétrie sur des images statiques. Les résultats sont comparés à la vérité terrain annotée. Deux

propriétés de l'axe de symétrie détecté sont validées : sa position et son inclinaison. La position de l'axe de symétrie est tout d'abord détectée sur des images possédant une symétrie verticale. Les paramètres optimaux utilisés sont alors retenus pour calculer l'inclinaison et la position de l'axe sur des images possédant cette fois un axe de symétrie bilatérale incliné dans le plan. Dans les paragraphes suivants, nous présentons les images utilisées pour ces expérimentations et nous citons la mesure utilisée pour évaluer l'erreur. La mesure de l'erreur nous permet de choisir les meilleurs paramètres pour l'algorithme de détection de la symétrie.

5.2.1 Jeux de données

Nous utilisons la base d'images FacePix [BGK+02] afin d'évaluer les résultats de la détection de l'axe de symétrie. Nous constituons deux ensembles d'images pour mesurer l'erreur sur la position de l'axe de symétrie ainsi que sur son inclinaison. Le premier ensemble représente des visages avec différentes orientations horizontales de $-45°$ à $+45°$ et un pas de $5°$. Cet ensemble va être utilisé pour mesurer l'erreur sur la position de l'axe de symétrie détecté. Pour ce faire, la position de l'axe de symétrie est manuellement annotée avec une inclinaison de $90°$. Afin de tester plusieurs résolutions, l'ensemble des images de visages a été re-dimensionné de 80×80 pixels à 60×60, 50×50, 40×40 et 30×30. Quelques images de personnes de cet ensemble sont présentées dans la figure 5.3.

Le deuxième ensemble d'images permet de mesurer l'erreur sur l'inclinaison de l'axe de symétrie détecté. Cet ensemble a été créé à partir d'images de visages de face que nous avons fait pivoter dans le plan jusqu'à $30°$ de chaque côté avec un pas de $5°$ (*cf.* figure 5.4).

5.2.2 Mesure de l'erreur

Pour calculer l'erreur sur la position de l'axe de symétrie détecté, la mesure la plus intuitive est la distance euclidienne moyenne entre les axes détectée Axe_i et la vérité terrain Axe_i^*. La distance est normalisée par la largeur du cadre englobant le visage

FIGURE 5.3: Exemples d'images de quelques personnes de l'ensemble 1. Images d'origine (a) et images re-dimensionnées à 60×60 (b), 50×50 (c), 40×40 (d) et 30×30 (e).

FIGURE 5.4: Exemples d'images d'une personne de l'ensemble 2.

$LargeurVisage$.

$$Erreur_{Position} = \frac{1}{Nb_{images}} \sum_i \frac{|Axe_i^* - Axe_i|}{LargeurVisage_i} \tag{5.1}$$

L'erreur sur l'inclinaison de l'axe de symétrie détectée est calculée de la même manière, avec l'erreur angulaire moyenne qui est la mesure la plus couramment utilisée pour évaluer la précision.

$$Erreur_{Angle} = \frac{1}{Nb_{images}} \sum_i |Angle_i^* - Angle_i| \tag{5.2}$$

5.2.3 Résultats et choix des paramètres

Nous rappelons les paramètres utilisés lors du processus de détection de la symétrie bilatérale du visage :

- L'intervalle $[AngleInf, AngleSup]$ de recherche de l'inclinaison α de l'axe de symétrie.
- La largeur r des cellules constituant la région d'intérêt.
- Le seuil $\varepsilon 1$ utilisé pour la mise en correspondance des pixels de la même cellule afin de déterminer si elle est homogène. Si au moins deux pixels d'une même cellule ne se correspondent pas, alors elle est considérée non homogène.
- Le seuil $\varepsilon 2$ utilisé pour la mise en correspondance des pixels d'une cellule avec sa cellule miroir afin de déterminer si elles sont symétriques ou pas (selon la propriété 3.3 *cf.* chapitre 3),

5.2.3.1 Erreur sur la position de l'axe de symétrie

Puisque l'axe de symétrie que nous avons annoté est vertical, dans cette section nous fixons α à 90° et nous étudions les paramètres ε_1, ε_2 et r. Le graphe de la figure 5.5 illustre l'erreur moyenne selon ces trois paramètres. Trois graphes correspondant chacun à une valeur de r, sont superposés dans un graphe 3D. Nous pouvons constater que lorsque le rayon est petit, l'erreur a tendance à rester plus stable en variant ε_1 et ε_2. La figure

5.6 montre le changement dans l'erreur pour chaque valeur r. Le seuil ε_1, qui détermine si une cellule n'est pas homogène, ne doit pas être très grand afin de ne pas exclure beaucoup de régions de l'image. Par contre, le seuil ε_2 utilisé pour faire correspondre deux cellules afin de déterminer si elles sont symétriques, doit être plus grand pour ne pas prendre en considération que les symétries parfaites. Les régions approximativement symétriques doivent être considérées comme des symétries potentielles dans ce processus. De cette manière, la détection de la symétrie ne repose pas sur des contours ou points caractéristiques mais devient plus globale et donc, plus robuste.

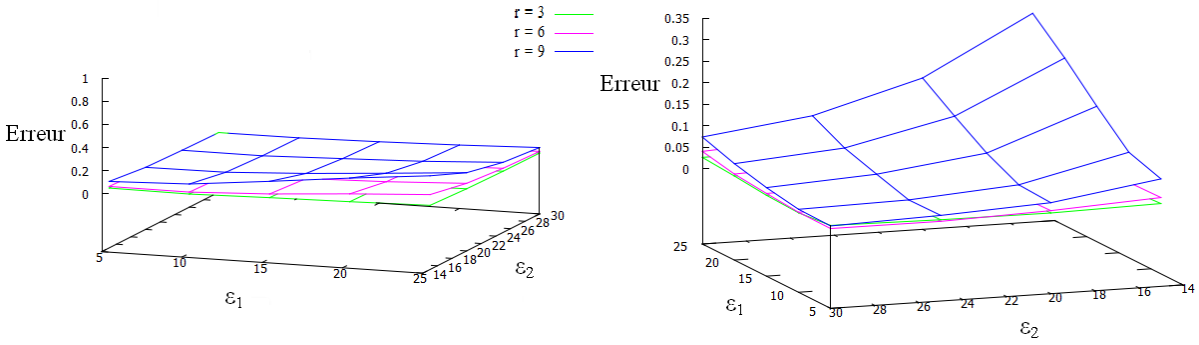

FIGURE 5.5: Deux vues 3D du graphe qui représente l'erreur selon les paramètres ε_1 et ε_2 pour trois valeurs différentes de r.

La tendance reste la même avec les images de plus petite résolution (*cf.* figure 5.7). Une légère augmentation de l'erreur est constatée pour les images de très petite résolution (30×30 pixels). Ceci est dû à la perte de beaucoup de symétries locales qui sont la base de la détection de la symétrie globale sur l'image.

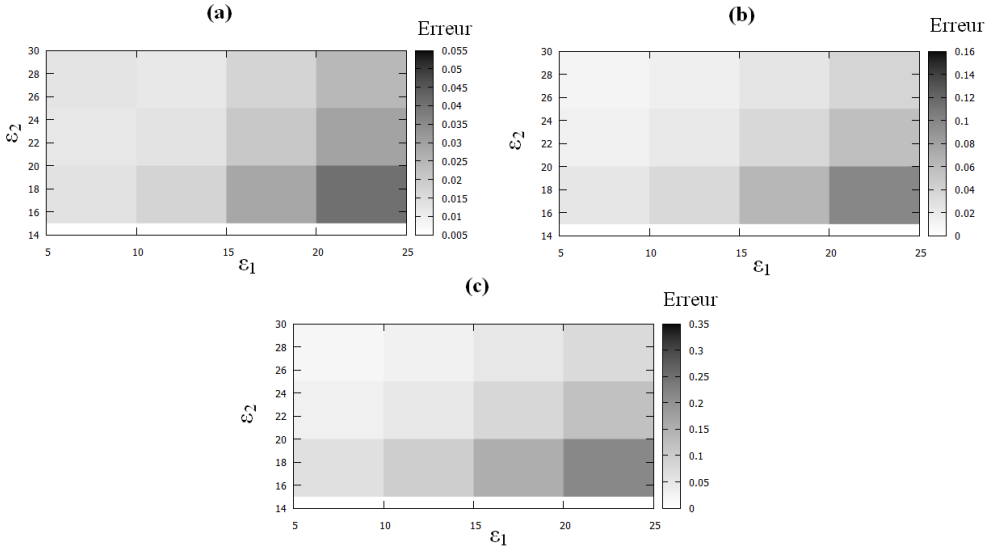

FIGURE 5.6: Diagramme de l'erreur selon les paramètres ε_1 et ε_2. (a) $r = 3$. (b) $r = 6$. (c) $r = 9$.

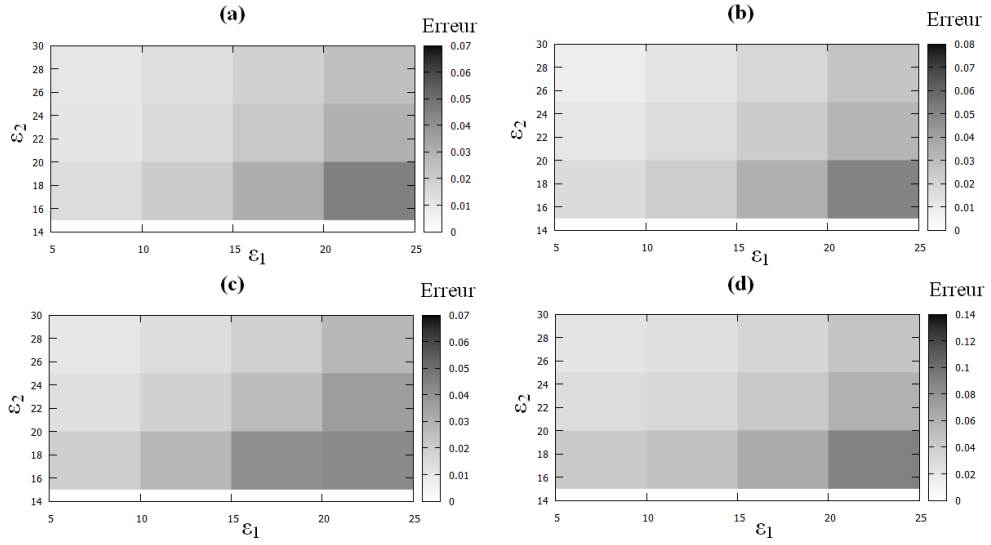

FIGURE 5.7: Diagramme de l'erreur selon les paramètres ε_1 et ε_2 avec $r = 3$. (a) Taille des visages 60×60. (b) Taille des visages 50×50. (c) Taille des visages 40×40. (d) Taille des visages 30×30.

Les figures 5.8 et 5.9 montrent des résultats de calcul de l'axe de symétrie sur quelques images choisies aléatoirement dans l'ensemble 1. Nous pouvons remarquer des cas où l'axe de symétrie n'est pas bien détecté (*cf.* figure 5.9 (a) 3ème ligne, 1ère colonne). En effet, d'autre axes de symétrie du visage peuvent l'emporter sur celui qui passe entre les yeux. L'axe de symétrie qui passe sur un œil et qui concerne la moitié du visage peut être considéré comme axe principal par le mécanisme de vote utilisé.Ceci est dû à la présence de plusieurs axes de symétrie locaux dans la moitié du visage. Nous avons limité ces cas d'erreurs en prenant en considération dans notre algorithme la distance entre les régions symétriques (*cf.* paragraphe 3.2.3.1). Celle ci ne doit pas être très petite afin d'exclure les symétries dans la même région, comme par exemple l'axe qui passe par un œil. Dans nos expérimentations nous avons fixé une distance minimale égale à $\frac{1}{6}$ de la largeur du visage.

Nous avons défini les meilleurs paramètres afin de détecter la position de l'axe de symétrie la plus proche de la vérité terrain. Cette position a été bien détectée indépendamment de la taille des visages. Nous utilisons ces résultats pour valider le processus complet de détection de la symétrie. Nous utilisons les paramètres retenus et complétons le processus à l'estimation de l'inclinaison de l'axe de symétrie. Les résultats sont présentés dans la section suivante.

FIGURE 5.8: Résultats de détection d'axe de symétrie sur des images de l'ensemble 1 de taille originale.

5.2.3.2 Erreur sur l'inclinaison de l'axe de symétrie

Nous utilisons l'ensemble 2 qui est constitué d'images de visages avec des rotations dans le plan (*cf.* figure 5.4). Nous fixons les paramètres selon les expérimentations précédentes, $r = 3$, $\varepsilon_1 = 10$ et $\varepsilon_2 = 30$. L'erreur angulaire moyenne lorsque $\alpha \in [60°, 120°]$ est de $2,95°$. La figure 5.10 montre des résultats d'axe de symétrie sur quelques images choisies aléatoirement dans l'ensemble 2.

La détection de la symétrie bilatérale du visage est la première étape dans notre

FIGURE 5.9: Résultats de détection d'axe de symétrie sur des images re-dimensionnées de l'ensemble 1.

approche d'estimation de la pose de la tête. Cette dernière dépend de la bonne détection de la symétrie. Pour cela, cette étape essentielle devait être validée à part. Cette validation est essentielle pour rendre efficaces et précises les différentes méthodes d'apprentissage.

FIGURE 5.10: Résultats de détection d'axe de symétrie sur des images de l'ensemble 2.

Dans ce qui suit, nous étudions la capacité des descripteurs extraits à refléter les mesures objectives de la symétrie.

5.3 Estimation de la pose de la tête

Nous avons validé les caractéristiques de la symétrie bilatérale du visage en vue d'estimer la pose de la tête. Des images et des séquences vidéos appartenant à des bases publiques annotées ont été utilisées. Les bases d'images utilisées présentent des variations d'expression et de conditions d'illumination. Elles ne traitent, cependant, qu'un seul degré de liberté qui est le pan. Les vidéos, quand à elles, comportent des mouvements selon les trois axes. Nous avons donc, synthétisé une base d'images de visages en présence des deux mouvements, pan et tilt afin de construire un modèle de prédiction des deux degrés de liberté. Ce modèle a été utilisé pour prédire le pan et le tilt sur les vidéos, puisque notre approche n'exige pas que les données d'apprentissage et de test proviennent de la même base d'images. Quand au troisième degré de liberté, le roll, il est défini pour les mouvements dans le plan par l'inclinaison de l'axe de symétrie (*cf.* section 4.1).

Les bases utilisées couvrent une grande variabilité en termes de pose et de conditions de prise de vue. Dans toutes les expérimentations, nous avons utilisé les paramètres $\varepsilon_1 = 10$, $\varepsilon_2 = 30$ et $r = 3$. L'intervalle de α est $[85°, 95°]$ pour les bases d'images avec mouvement pan (FacePix et CMU PIE) et $[60°, 120°]$ pour la base des vidéos avec mouvements aléatoires (BU dataset) ainsi que la base synthétique. Nous allons évaluer les deux vecteurs de caractéristiques définis dans le chapitre précédent. Nous présentons les résultats ci-après regroupés selon le vecteur de caractéristiques utilisé. Nous avons évalué aussi l'architecture en routeur qui permet de combiner les deux vecteurs.

5.3.1 Validation des caractéristiques de la symétrie sous différentes conditions

Nous étudions l'impact du changement dans les conditions externes, sur l'efficacité de la symétrie bilatérale du visage à discriminer les poses de la tête. Nous utilisons les bases FacePix et CMU PIE moyennant le vecteur de caractéristiques TRS (*cf.* section 4.2). Ce vecteur est basé sur la taille de la région symétrique et est spécifique au mouvement horizontal de la tête. La base FacePix est caractérisée par sa précision et la possibilité

d'en extraire un nombre variable de poses vu qu'elle dispose d'images à 1° d'intervalle. Elle nous a permis de tester des configurations avec un nombre de poses différent et des résolutions d'images différentes. L'influence de la précision dans la détection de l'axe de symétrie est aussi testée. La base CMU PIE, quand à elle, dispose de plusieurs ensembles d'images avec différentes expressions et différents angles et intensités d'illumination que nous testons également. Enfin, nous validons le classifieur construit avec des images statiques sur les séquences vidéo avec des mouvements continus appartenant à la base BU. Les expérimentations sont validées à chaque fois via le taux de bonne classification accompagné de l'erreur moyenne absolue (EMA). Cette erreur est calculée relativement aux annotations fournies avec les bases d'images et de vidéos.

5.3.1.1 Variation du nombre de poses utilisées

La base Face Pix [BGK+02] consiste en un ensemble de poses dans l'intervalle ±90° à un pas de 1°. Ceci nous permet de créer plusieurs ensembles à différentes configurations en changeant le nombre de poses à chaque fois (*cf.* paragraphe 4.3). Nous avons testé les trois configurations suivantes :

- 19 poses discrètes associées aux angles pan de $-45°$ à 45° avec un pas de 5°,
- 9 poses discrètes associées à l'intervalle $-40°$ à 40° avec un pas de 10°,
- 7 poses discrètes associées à l'intervalle $-45°$ à 45° avec un pas de 15°.

En exploitant la propriété du vecteur TRS qui est l'invariabilité à la réflexion (*cf.* paragraphe 4.2.1), nous obtenons un nombre de classes pour la première configuration égal à 10, 5 classes pour la deuxième configuration et 4 classes pour la troisième configuration. Le tableau 5.2 montre les résultats pour les trois modèles en utilisant les deux méthodes de classification : l'arbre de décision alternatif (LADTree) et les séparateurs à vaste marge (SVM) ainsi que la régression. Nous divisons les données en six ensembles de même taille et appliquons une validation croisée à $6-ensembles$. A chaque passage, 5 ensembles sont utilisés pour l'apprentissage et l'ensemble restant est utilisé pour le test. Le modèle à 7 poses possède le plus haut taux de bonne classification. Cependant, même pour le classifieur à 19 poses, la concentration reste autour de la diagonale (*cf.* figure 5.11). Les matrices de confusion des deux classifieurs LADTree et SVM montrent que

l'erreur est dans la classe voisine. Sachant que lorsque le nombre de classes est grand, la distance entre deux classes voisines est petite. Ceci est confirmé par l'erreur moyenne des deux classifieurs qui ne suit pas le taux de classification et reste petite dans les trois configurations.

La régression donne une erreur stable indifféremment du nombre de poses. En effet, un nombre de nœuds égal à 7 est suffisant pour construire la fonction de régression tout aussi bien que 9 ou 19 nœuds. Il y a une petite variation dans l'erreur vu que les données ne sont pas les mêmes. Nous allons voir l'intérêt de la régression et du modèle de classification à grand nombre de poses dans le traitement des séquences vidéos où les poses sont continues.

Tableau 5.2: Résultats d'estimation du pan sur la base FacePix avec un nombre variable de poses.

Données	Taux de classification (%)		EMA (°)		
	LADTree	SVM	LADTree	SVM	Régression
19 poses	31.99	30.72	5,09	5,25	4,81
9 poses	54.27	57.69	5,04	4,91	4,63
7 poses	79.63	79.01	3,14	3,70	4,65

Le tableau 5.3 présente les résultats des mêmes expérimentations en utilisant la base synthétique. La différence entre les deux bases est que FacePix possède un tilt nul alors que la base synthétique est créée avec un tilt variable. Nous testons les deux configurations avec la base synthétique. Les résultats sont proches de ceux de la base FacePix (*cf.* tableau 5.3) avec un avantage pour les modèles construits sous une orientation tilt nulle. En effet, les caractéristiques du pan sont mieux discriminées lorsque le tilt est uniforme. Les poses, dans ce cas, rassemblent plus de caractéristiques.

5.3.1.2 Précision de la détection de l'axe de symétrie

Nous testons la robustesse de la méthode par rapport à la précision de la détection de l'axe de symétrie. Les expérimentations de la section précédente se sont déroulées de manière complètement automatique. Maintenant, nous utilisons les annotations d'axes de

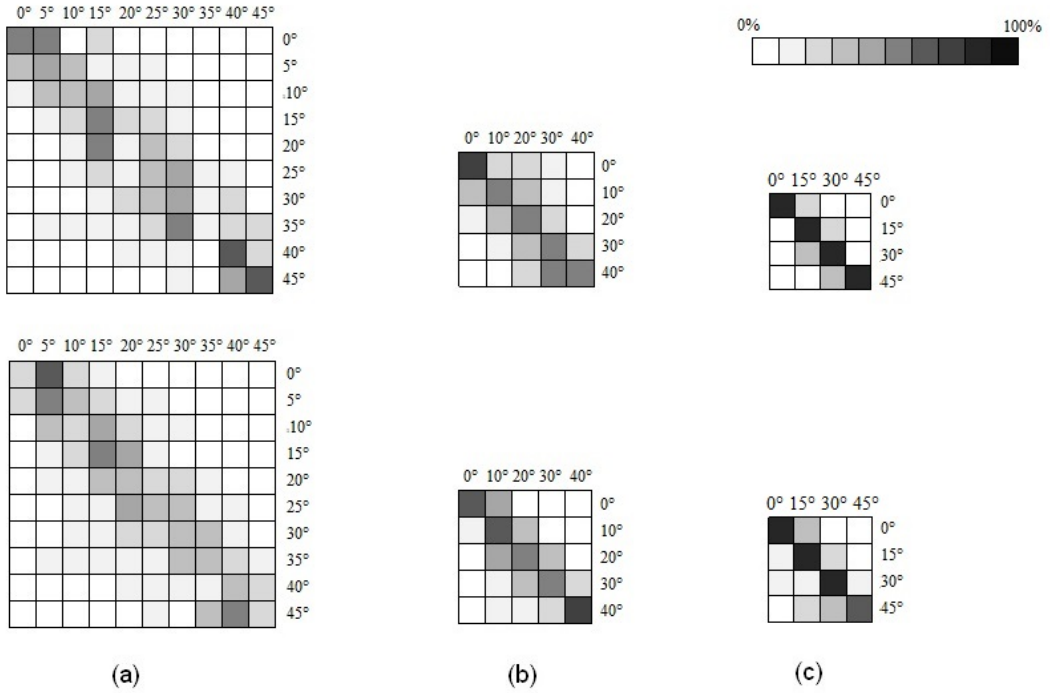

FIGURE 5.11: Matrices de confusion associées à : (a)un classifieur à 19 poses avec un pas de 5°. (b)un classifieur à 9 poses avec un pas de 10°. (c)un classifieur à 7 poses avec un pas de 15°. La première ligne concerne l'arbre de décision et la deuxième est celle du SVM.

symétrie exploitées dans la section 5.2 afin de comparer les résultats d'une configuration semi-automatique avec celle qui est complètement automatique. Un autre paramètre, qui est la précision de la détection du visage sur l'image, influence l'estimation de la pose. Nous testons, donc, la sensibilité à la détection du visage. Les détails des résultats obtenus avec le modèle de 7 poses en utilisant l'arbre de décision alternatif, sont présentés dans le tableau 5.4.

Nous remarquons que dans une configuration complètement automatique, le taux de classification diminue. Ceci nous indique l'inconvénient majeur de notre méthode, qui est la sensibilité à la détection du visage et à celle de la symétrie, dans la mesure où nous

Tableau 5.3: Résultats d'estimation du pan sur la base synthétique avec un nombre variable de poses.

Données		Taux de classification (%)		Erreur moyenne (°)		
		LADTree	SVM	LADTree	SVM	Régression
tilt variable	19 poses	29,96	31,41	5,57	5,55	5,55
	10 poses	57,36	58,26	4,79	4,59	5,47
	7 poses	67,95	70,96	4,99	4,47	6,05
tilt = 0	19 poses	35,13	38,41	4,67	4,23	4,26
	10 poses	56,45	63,68	4,71	3,84	4,49
	7 poses	76,45	80,69	3,53	2,89	4,6

Tableau 5.4: Résultats d'estimation du pan sur la base FacePix avec deux configurations semi-automatiques et une configuration complètement automatique.

Données	Taux de classification (%)	EMA (°)
Axe de symétrie annoté	82.38	2.71
Visage annoté et symétrie automatique	81.90	2.78
Détection du visage et de la symétrie automatique	79.63	3.14

cherchons à déployer un système automatique. En effet, l'erreur dans l'estimation de la pose intègre l'erreur dans la détection de l'axe de symétrie.

5.3.1.3 Résolution des images

Nous avons généré des images basse résolution à partir de la base FacePix. La taille des visages était de 80×80 à l'origine. Nous avons re-dimensionné pour obtenir deux ensembles. Dans le premier ensemble, la taille des visages est 40×40 pixels et 25×25 pixels dans le deuxième. Les résultats sont présentés dans le tableau 5.5.

La précision a diminué de $79,6\%$, pour les images originales, à $63,8\%$ pour les images à très basse résolution. La méthode est sensible à la détection de la symétrie. Sur les images à basse résolution, les symétries locales utilisées pour détecter l'axe ne sont pas très pertinentes.

Tableau 5.5: Résultats d'estimation du pan sur la base FacePix avec différentes résolutions d'images.

Résolution	Taux de classification (%)		Erreur moyenne (°)		
	LADTree	SVM	LADTree	SVM	Régression
80×80	79,63	79.01	3,14	3,70	4,65
40×40	74,1	75,24	4,57	4,00	5,67
25×25	63,8	62,31	7,24	6,30	8,06

5.3.1.4 Changement d'illumination et d'expression

Nous avons testé notre approche sur la base CMU PIE afin de la valider dans un contexte qui apporte plus de variabilité dans les conditions d'éclairage et d'expression. Nous utilisons les trois ensembles : CMU Expression, CMU Lighting et CMU Talking. Nous avons construit un modèle pour chaque ensemble afin de tester à part la robustesse à l'éclairage et aux expressions. Nous fusionnons également les images des trois ensembles dans un ensemble englobant. Dans l'ensemble *Expression*, la distance des différentes caméras par rapport aux utilisateurs n'est pas la même. Nous normalisons la taille de tous les visages extraits. Nous re-dimensionnons les visages à 80×80, qui est la taille moyenne dominante des visages des autres ensembles, pour ne pas laisser la différence en taille discriminer les poses.

Le défi auquel les caractéristiques de la symétrie sont confrontées est l'ensemble avec l'éclairage variable. En effet, le changement d'expression ne pénalise pas l'estimation de la pose car les caractéristiques sont géométriques et concernent le visage dans sa globalité. Elles sont extraites de la même manière indépendamment de l'expression. Cependant, lorsqu'un éclairage intense est appliqué sur un côté latéral du visage (*cf.* figure 5.1 ligne 2), la scène perd sa symétrie. Dans ce cas, l'égalisation d'histogramme n'est pas suffisante. Nous appliquons une normalisation basée sur la transformée en cosinus discrète DCT (de l'anglais *discrete cosine transform*). Un nombre de coefficients est tronqué pour minimiser la variation de l'illumination qui se trouve principalement dans les basses fréquences. Nous calculons le taux de classification et l'erreur moyenne pour chaque ensemble en utilisant la validation croisée à $6-ensembles$. Les résultats sont présentés dans le tableau

5.6.

La normalisation par DCT ajoute un certain bruit qui affecte le processus de mise en correspondance, lors de la détection de la région symétrique. Sur l'ensemble *Expression* la précision diminue de 72.57% à 49.81%. Contrairement à l'ensemble *Lighting* où la normalisation fait plus de bien que de mal. En effet, l'éclairage affecte la mise en correspondance plus que le fait le bruit ajouté par la normalisation. D'un autre côté, la normalisation donne de meilleurs résultats sur les ensembles avec un grand nombre d'images d'apprentissage. C'est le cas des ensembles *Lighting* et *Talking* ainsi que l'ensemble englobant toutes les images. Le grand nombre d'images utilisées pour l'apprentissage compense la perte due à la normalisation et permet même d'améliorer légèrement la précision de ces ensembles. Nous préconisons donc, l'utilisation de la normalisation par DCT seulement dans les conditions extrêmes d'éclairage moyennant un modèle construit avec un nombre important d'images d'apprentissage.

Tableau 5.6: Résultats d'estimation du pan sur la base CMU PIE.

Données	Pré-traitement	Taux de classification (%)		Erreur moyenne (°)		
		LADTree	SVM	LADTree	SVM	Régression
CMU Expression	Égalisation d'histogramme	72.57	69.54	7,03	7,63	9,95
CMU Lighting		72.51	70.54	6,18	6,62	10,12
CMU Talking		80,78	80,91	4,32	4,29	7,98
CMU PIE		73,40	74.17	6,04	5,86	9,2
CMU Expression	DCT	49,81	50,27	13,71	13,33	13,55
CMU Lighting		85,90	86,29	3,17	3,08	6,47
CMU Talking		87,63	88,56	2,78	2,57	6,81
CMU PIE		82,54	83,67	4,28	4,03	7,53

5.3.1.5 Résultats obtenus avec le vecteur TRS sur la base des vidéos BU

Nous testons également notre méthode sur des séquences vidéo de la base BU car notre but est d'utiliser la solution dans un environnement réel et avoir un retour d'information sur la pose.

Dans la base BU, les personnes effectuent des mouvements de tête aléatoires selon les trois degrés de liberté. Dans cette section, nous estimons le pan en utilisant les modèles appris avec la base d'images FacePix. Les images de cette base couvrent mieux l'intervalle des poses que celles de la base CMU PIE. En effet, les poses dans CMU PIE sont largement espacées (22.5° entre deux poses contiguës). Nous nous assurons d'abord que la taille des visages dans la base de vidéos est le même que dans la base avec laquelle le modèle a été créé (FacePix). Un redimensionnement est effectué si nécessaire. L'erreur calculée est exprimée en Erreur Absolue Moyenne (EAM) et en Écart Type (STD). Les résultats sont détaillés dans le tableau 5.7.

Les meilleurs résultats ont été obtenus avec les modèles à 19 poses discrètes (5° d'intervalle entre les poses). La performance de ces modèles, malgré leur taux de classification bas par rapport aux classifieurs à 7 poses (lors de l'apprentissage), se justifie par le fait que les instances mal classifiées se trouvent attribuées une pose dans le voisinage de la vérité terrain (*cf.* matrices de confusion au paragraphe 5.3.1.1). Aussi, l'intervalle entre les poses est petit et convient mieux aux mouvements continus comme ceux dans les séquences vidéo.

Tableau 5.7: Résultats de l'estimation du pan sur la base de vidéos BU.

Modèle utilisé	EAM (°)	STD (°)
LADTree - 5° d'intervalle	5,24	4,33
LADTree - 15° d'intervalle	5,12	5,62
SVM - 5° d'intervalle	5,35	5,18
SVM - 15° d'intervalle	6,63	6,00
Régression - 5° d'intervalle	6,75	6,29
Régression - 15° d'intervalle	7,76	6,04

5.3.1.6 Comparaison avec l'état de l'art

Nous avons comparé nos résultats avec les autres méthodes qui ont utilisé les mêmes bases. Le tableau 5.8 présente les résultats obtenus sur FacePix et CMU PIE, exprimés en taux de classification et en erreur moyenne absolue (EMA). Nous remarquons que nos

résultats sont comparables à l'état de l'art et dans certains cas meilleurs. Sur la base FacePix, les méthodes par intégration de variétés (manifold embedding) donnent de très bon résultats mais il n'y a pas de solution explicite pour intégrer de nouvelles données aux ensembles LLE et LE [MCT09]. Contrairement à notre approche, où les nouvelles données permettent la prédiction de la pose via un modèle construit auparavant.

Tableau 5.8: Comparaison des résultats du pan sur les bases FacePix et CMU PIE avec l'état de l'art.

Méthode	Face Pix		CMU PIE
	EAM (°)	Taux (%)	Taux (%)
K-manifold clustering [LLL10]	3,16	-	-
Regression [JLS$^+$11]	6,1	-	-
Biased Isomap [BYP07]	5,02	-	-
Biased LLE [BYP07]	2,11	-	-
Biased LE [BYP07]	1,44	-	-
Regression [lTBC$^+$03]	-	-	84,30
Caractéristiques de la symétrie	3,14	79,63	83,67

Dans ce qui suit, nous estimons les deux degrés de liberté pan et tilt sur les séquences vidéo de la base BU. Les résultats seront aussi comparés à l'état de l'art.

5.3.2 Résultats d'estimation des trois degrés de liberté

Nous entreprenons à présent l'estimation des trois degrés de liberté sur les séquences vidéo de la base BU. Nous disposons d'un modèle robuste de prédiction du pan. Le roll est défini par l'inclinaison de l'axe de symétrie lors d'un mouvement dans le plan. Il reste l'estimation du tilt qui n'est pas encore validée. Pour cela, nous utilisons le vecteur ORS (*cf.* paragraphe 4.2) qui caractérise la forme de la région symétrique propre à une pose. Les deux vecteurs peuvent aussi, être combinés moyennant l'architecture par routeur (*cf.* paragraphe 4.3.3).

Nous allons valider le vecteur ORS en utilisant la base d'images synthétiques afin de faire face aux deux mouvements pan et tilt. La base d'images synthétique a été créée

spécialement pour construire un modèle de prédiction du tilt. Le modèle qui va donner les meilleurs résultats sera interrogé pour prédire la pose sur les vidéos de la base BU.

5.3.2.1 Construction de modèles pour estimer le tilt

Contrairement au vecteur TRS qui est spécifique au mouvement horizontal de la tête, le vecteur ORS basé sur l'orientation de la région symétrique caractérise la forme globale de cette dernière. Chaque forme est spécifique à une pose de la tête exprimant un pan et un tilt. Nous utilisons, alors, le vecteur de caractéristiques ORS pour construire un modèle de prédiction du tilt. Ce vecteur peut être utilisé pour estimer une combinaison de pan et de tilt et aussi, pour estimer chaque degré de liberté seul.

Dans la configuration de poses combinées, un seul classifieur est utilisé pour estimer une pose qui combine le pan et le tilt. Chaque combinaison est codifiée et représentée par une pose. Lorsque nous utilisons 7 poses pan et 3 poses tilt, nous obtenons 21 poses qui combine chacune un pan et un tilt. Nous obtenons $55,47\%$ avec le LADTree contre $70,78\%$ avec les SVMs. Nous pouvons remarquer que le LADTree n'a pas bien réagi à un grand nombre d'attributs. En effet, le vecteur ORS contient 9 fois plus d'attributs que le vecteur TRS. Par contre, l'augmentation du nombre d'attributs n'a pas eu d'effet négatif sur les SVMs. Néanmoins c'est la qualité des données, linéairement séparables, qui a permis d'avoir un meilleur taux de classification.

Nous allons utiliser le classifieur le plus robuste afin d'évaluer les résultats du pan et du tilt à partir d'une pose combinée. Nous ne testons pas la régression car les valeurs dont nous disposons sont des codifications et non des angles d'orientation. Les poses prédites pour chaque instance sont ré-exprimées en pan + tilt afin que l'erreur moyenne soit calculée. Les résultats sont présentés dans le tableau 5.9. Nous testons deux configurations : 7 poses pan et 19 poses pan. De la première résulte 21 poses et de la deuxième 57. Nous utilisons la validation croisée à $10 - ensembles$. Les résultats sont comparables pour les deux modèles. La combinaison des deux poses pan et tilt en une seule pose n'est pas la configuration idéale. Le pan est toujours bien estimé contrairement au tilt.

Nous évaluons aussi l'estimation du pan et du tilt séparément en utilisant la classification et la régression. Le tableau 5.10 présente les résultats exprimés en taux de

Tableau 5.9: Résultats de l'estimation du pan et du tilt en poses combinées avec la base synthétique.

Données	Taux de classification (%)		Erreur moyenne (°)	
	pan	tilt	pan	tilt
21 poses	68,25	61,18	5,15	15,35
57 poses	30,12	40,81	5,85	16,37

classification et en erreur moyenne (EMA). Les résultats du tilt sont toujours moins bons que ceux du pan. Nous utilisons, donc, la précision obtenue avec le pan pour aider à mieux estimer le tilt. Nous construisons un modèle de tilt pour chaque pan, afin qu'il puisse être utilisé sur les séquences vidéo après un routage effectué avec la prédiction du pan. Les taux obtenus avec ces modèles de tilt sont présentés aussi dans le tableau 5.10. Nous remarquons que ces modèles donnent de meilleurs résultats que celui qui estime le tilt indépendamment du pan.

5.3.2.2 Résultats d'estimation de la pose sur la base de vidéos BU

D'après les tests réalisés jusqu'ici, nous utilisons l'architecture par routeur pour estimer la pan et le tilt. Le roll est calculé lors de la détection de l'axe de symétrie. L'erreur moyenne sur toutes les séquences est de $2,57°$ et l'écart type est $3,56°$.

Nous avons conçu une architecture qui permet de tirer profit des meilleurs modèles dont nous disposons afin de renforcer les moins robustes. Les résultats obtenus avec les deux vecteurs de caractéristiques nous indiquent que le vecteur TRS est robuste pour estimer le pan. Le vecteur ORS, quand à lui, est moins robuste mais a l'avantage de caractériser le tilt, chose que ne fait pas le premier. Nous allons exploiter les avantages de chacun en utilisant l'architecture avec routeur (*cf.* figure 4.15) et ceci pour améliorer les résultats de l'estimation du tilt. Dans cette architecture, un modèle est utilisé dans une première étape pour estimer un degré de liberté. Nous apprenons un modèle du deuxième degré de liberté pour chaque résultat du premier. En l'occurrence, notre routeur sera le modèle du pan puisque nous disposons de plusieurs modèles robustes. Pour chaque pan estimé, le modèle de tilt correspondant sera interrogé.

Tableau 5.10: Résultats de l'estimation du pan et du tilt en poses séparées avec la base synthétique.

			Taux de classification(%)	EMA (°)	
			SVM	SVM	Régression
Pan	tilt variable	19 poses	29,76	5,93	5,89
		7 poses	68,34	4,99	6,64
	tilt = 0	19 poses	30,01	5,29	5,19
		7 poses	77,22	3,3	4,82
Tilt	pan variable	3 poses	57,28	15,76	16,7
	pan = 0		32,38	13,00	9,98
	pan = 5		33,01	12,88	10,4
	pan = 10		38,35	11,72	10,39
	pan = 15		34,15	12,4	10,33
	pan = 20		44,1	11,23	10,19
	pan = 25		54,49	8,67	9,19
	pan = 30		64,94	6,03	7,51
	pan = 35		69,72	5,28	7,01
	pan = 40		65,85	5,24	6,65
	pan = 45		72,23	4,16	5,4

Nous utilisons le modèle de pan à 15° d'intervalle qui a donné de bons résultats (l'arbre alternatif via le vecteur de caractéristiques TRS). Nous apprenons pour les 7 poses, un modèle de tilt via le vecteur de caractéristiques ORS. Ces modèles auront l'avantage de rassembler plus de caractéristiques puisqu'ils ont en commun la même orientation du pan.

Les résultats de la prédiction du tilt en utilisant les 7 modèles associés aux résultats de l'estimation du pan sur la base BU sont présentés dans le tableau 5.11.

Tableau 5.11: Résultats de l'estimation du tilt sur la base BU avec une architecture en routeur.

Modèle utilisé	EAM (°)	STD (°)
SVM	9,13	8,05
Régression linéaire	5,82	6,41

Nous remarquons que la division de l'ensemble d'apprentissage en 7 ensembles a permis de construire une régression qui a donné une estimation plus précise sur l'orientation du tilt. En effet, la construction d'un régresseur de tilt pour chaque pan est plus fiable. La rotation est continue autour de l'axe des X suivant la même orientation par rapport à l'axe des Y.

5.3.2.3 Comparaison avec l'état de l'art

Le tableau 5.12 regroupe les résultats obtenus sur la base de vidéos BU. Nos résultats sont comparables à l'état de l'art et dans certains cas, ils sont meilleurs.

Tableau 5.12: Comparaison des résultats de la base BU avec l'état de l'art.

	Caractéristiques de la symétrie		[MWM10]	[VSG12]
	EMA (°)	STD (°)	EMA (°)	STD (°)
roll	2,57[b]	3,56[b]	2,91	2,82[a]
tilt	5,82	6,41	3,67	4,67[a]
pan	5,12	5,62	4,97	5,79[a]

[a] La pose est estimée seulement lorsque les yeux sont détectés.
[b] La pose est estimée seulement dans le cas de mouvement dans le plan.

5.3.3 Cas particulier d'occultations

L'utilisation des caractéristiques de la symétrie bilatérale du visage présente un avantage sur les autres méthodes géométriques qui utilisent des points spécifiques. En effet, une occultation partielle du visage peut être gérée du moment qu'il existe au moins deux pixels symétriques de part et d'autre du visage. Ceci est dû au fait que tous les pixels de la texture contribuent à la démarcation de la région symétrique. Les points spécifiques du visage peuvent être occultés ou mal détectés à cause des obstacles ou de l'amplitude des mouvements. Certains points caractéristiques peuvent disparaitre de l'image plane lorsque le mouvement sur le côté dépasse 30°.

Nous avons testé le processus d'extraction des caractéristiques de symétrie sur des séquences vidéos prises avec une web-cam. La figure 5.12 montre des captures de quelques séquences. Nous pouvons voir un cas d'occultation d'un œil. La région symétrique reste la même à travers les trames en dépit du mouvement de la main sur le visage (*cf.* figure 5.13).

FIGURE 5.12: Exemple de trames appartenant à des séquences vidéo prises dans le laboratoire.

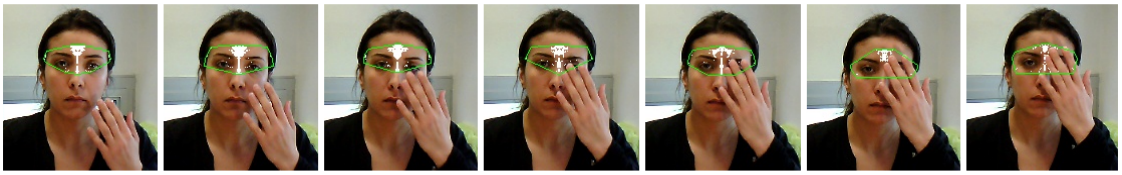

FIGURE 5.13: Exemple de trames illustrant une occultation partielle du visage.

5.3.4 Conclusion

Le principal avantage de notre méthode est que le calcul peut commencer à n'importe quelle pose, sans initialisation. A partir du moment où le visage et la symétrie sont détectés automatiquement pour les poses entre -45° et +45°. Le processus est complètement automatique mais sensible à la précision de la détection de la région symétrique sur le visage. Le système est robuste au changement dans les condition d'éclairage, au changement d'expression et aussi à l'identité des personnes puisque le fondement de la méthode est géométrique. Et, contrairement aux méthodes géométriques existantes, aucun point spécifique du visage n'est requis. Tous les pixels de la texture du visage sont testés pour faire part de la région symétrique. De ce fait, les yeux fermés ou les visages partiellement occultés donnent les mêmes résultats qu'un visage complet.

Chapitre 6

Conclusion et perspectives

Sommaire

6.1 Conclusion générale

Nous avons présenté une nouvelle approche pour l'estimation de la pose de la tête humaine. Nous en exploitons la symétrie bilatérale du visage afin de faire face aux mouvements dans le plan (roll) et hors plan (pan et tilt). Le roll est déduit de l'inclinaison de l'axe de symétrie. Le pan et le tilt sont estimés à partir des caractéristiques de la symétrie bilatérale du visage. Notre méthode exploite la simplicité des méthodes géométriques d'estimation de la pose ainsi que la généralité des méthodes globales. Les méthodes géométriques locales sont rapides et indépendantes de l'identité mais sont basées sur des points spécifiques du visage.

Notre approche est basée sur la détection de l'axe de symétrie sur le visage, à partir duquel la région symétrique est définie. La région symétrique est délimitée à partir de la texture globale du visage et non pas en suivant des points spécifiques. Des caractéristiques sont extraites de cette région afin de discriminer les différentes poses de la tête selon les orientations du pan et du tilt. La classification s'est avérée robuste pour les mouvements à grande ampleur spécialement le pan. De plus le pan a été mieux caractérisé par la symétrie bilatérale que le tilt. Les résultats sur des bases publiques ont montré l'efficacité de la classification sur les caractéristiques du mouvement pan. Le mouvement tilt, quand à lui a moins d'ampleur et reste dépendant du pan qui a plus d'influence sur la géométrie de la région symétrique du visage. En effet, les poses selon l'orientation du pan sont bien discriminées même avec un tilt variable. Ce qui n'est pas le cas du tilt, les meilleurs résultats ont été obtenus en construisant un régresseur de tilt, et ceci pour différentes orientations pan.

Le principal avantage de notre méthode est que le calcul peut commencer à n'importe quelle pose dans l'intervalle $\pm 45°$ horizontalement. L'initialisation n'est pas requise puisque l'axe de symétrie est détecté automatiquement. Le processus est, donc, complètement automatique et robuste au changement dans les conditions d'éclairage, au changement d'expression et aussi à l'identité des personnes puisque le fondement de la méthode est géométrique. Cependant, la méthode est sensible à la précision dans la détection de l'axe de symétrie. Aussi, l'amplitude des mouvements dont la pose est esti-

mée, est limitée à 45° de chaque côté latéral. Au delà de cet intervalle, l'axe de symétrie n'apparait pas sur le plan image. Ces inconvénients nous incitent à réfléchir à d'autres perspectives pour essayer de compléter notre travail.

6.2 Perspectives

Nous proposons ci-après, des orientations de recherche dans les deux axes principaux des travaux réalisés : la détection de la symétrie et l'estimation de la pose de la tête.

Exploitation de la symétrie

Nous avons analysé certaines limites de notre approche. Notamment, la sensibilité à la détection des visages ainsi que la détection de l'axe de la symétrie. Dans notre méthode, nous entamons la localisation de l'axe de symétrie après celle du visage. Nous proposons que la deuxième étape ne soit pas conditionnée par le succès de la première, mais que la localisation de la symétrie oriente la localisation du visage. Ceci peut s'avérer très utile, spécialement pour des images à très basse résolution. La détection de l'axe de symétrie sur des images à très basse résolution pourrait se faire sans passer par l'étape de détection du visage. En effet, la détection de visage sur ces images n'est pas une tâche facile à réaliser. La détection de la symétrie peut attirer l'attention sur l'emplacement potentiel d'un visage.

D'un autre côté, l'amélioration de l'algorithme de détection de la symétrie est envisagée car l'erreur dans l'estimation de l'orientation et de la position de l'axe de symétrie va être intégrée dans l'erreur de l'estimation de la pose. Effectivement, notre détection de l'axe de symétrie est basée sur les symétries locales. Sur des images à très basse résolution, ces symétries ne sont pas très pertinentes. Nous prévoyons de remplacer le processus de mise en correspondance entre des petites régions de l'image, que nous avons appelé *cellule*, par une mise en correspondance plus globale, en utilisant toutes les données qui peuvent être filtrées ou réduites dans un plus petit espace.

D'une estimation grossière de la pose vers une estimation plus fine

Notre système retourne une estimation grossière de la pose. Ce résultat est amplement suffisant pour inférer la direction de l'attention du regard d'une personne ou bien pour regrouper un ensemble d'images de visages selon la pose de ce dernier afin de soumettre ces images à d'autres algorithmes de reconnaissance de forme. Cependant, nous pouvons utiliser ce système comme première étape d'un système hybride d'estimation de la pose. La pose estimée sera considérée comme une initialisation de la deuxième étape. Durant cette dernière étape, le système hybride transforme les poses discrètes en poses continues et étend l'intervalle du mouvement. La corrélation temporelle issue du suivi de la tête peut être exploitée à cette fin. Des modèles plus robustes d'estimation de la pose peuvent êtres construits à partir de données continues. La symétrie peut être exploitée moyennant des méthodes avancées de régression ou d'apprentissage de variétés.

Bibliographie

[AKK11] Stylianos Asteriadis, Kostas Karpouzis, and Stefanos Kollias. Robust validation of visual focus of attention using adaptive fusion of head and eye gaze patterns. *ICCVW*, 2011.

[AP09] Jania Aghajanian and Simon J.D. Prince. Face pose estimation in uncontrolled environments. *BMVC*, 2009.

[ASGG09] Akshay Asthana, Conrad Sanderson, Tom Gedeon, and Roland Goecke. Learning-based face synthesis for pose-robust recognition from single image. *BMVC*, 2009.

[BEB08] Jan Bandouch, Florian Engstler, and Michael Beetz. Evaluation of hierarchical sampling strategies in 3d human pose estimation. *BMVC*, 2008.

[Ben10] Chiraz BenAbdelkader. Robust head pose estimation using supervised manifold learning. *European Conference on Computer Vision (ECCV)*, 2010.

[BETvG08] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding (CVIU)*, 110(3) :346–359, 2008.

[Bey94] David J. Beymer. Face recognition under varying pose. *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 756–761, 1994.

[BFOS84] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

[BGK+02] J. Black, M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan. A framework for performance evaluation of face recognition algorithms. In *ITCOM, Internet Multimedia Systems II, Boston*, 2002.

[BGV92] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152, 1992.

[BK32] G.D. Birkoff and O.D. Kellogg. Aesthetic measure. In *Harvard Univ. Press*, 1932.

[BM09] K. Bailly and M. Milgram. Head pan angle estimation by a nonlinear regression on selected features. In *International Conference on Image Processing (ICIP)*, pages 3589–3592, 2009.

[BmO04] Sileye O. Ba and Jean marc Odobez. A probabilistic framework for joint head tracking and pose estimation. In *International Conference on Pattern Recognition (ICPR)*, pages 264–267, 2004.

[BMP00] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context : A new descriptor for shape matching and object recognition. In *NIPS*, pages 831–837, 2000.

[BMX+04] Simon Baker, Iain Matthews, Jing Xiao, Ralph Gross, Takeo Kanade, and Takahiro Ishikawa. Real-time non-rigid driver head tracking for driver mental state estimation. In *in 11th World Congress on Intelligent Transportation Systems*, 2004.

[BN03] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6) :1373–1396, 2003.

[BO11] Sileye O. Ba and Jean-Marc Odobez. Multiperson visual focus of attention from head pose and meeting contextual cues. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 33 :101–116, 2011.

[BR08] Ben Benfold and Ian Reid. Colour invariant head pose classification in low resolution video. *BMVC*, 2008.

[BR09] Ben Benfold and Ian Reid. Guiding visual surveillance by tracking human attention. In *BMVC*, 2009.

[Bre01] Leo Breiman. Random forests. *Mach. Learn.*, 45(1) :5–32, October 2001.

[BYP07] Vineeth Nallure Balasubramanian, Jieping Ye, and Sethuraman Panchanathan. Biased manifold embedding : A framework for person-independent head pose estimation. *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[CET01] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 23(6) :681–685, June 2001.

[CEW06] Weilong Chen, Meng Joo Er, and Shiqian Wu. Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain. *IEEE Transactions on Systems, Man, and Cybernetics*, 36 :458–466, 2006.

[CFB05] Xin Chen, Patrick J. Flynn, and Kevin W. Bowyer. Bowyer : Fully automated facial symmetry axis detection. In *in Frontal Color Images, 4-th IEEE Workshop on Automatic Identification Advanced Technologies*, pages 106–111, 2005.

[CGX11] Ke Chen, Shaogang Gong, and Tao Xiang. Human pose estimation using structural support vector machines. *ICCV*, 2011.

[CL06] H. Cornelius and G. Loy. Detecting rotational symmetry under affine projection. In *International Conference on Pattern Recognition (ICPR)*, pages 292–295, 2006.

[Cox03] H.S.M. Coxeter. *Projective Geometry*. Fuzzy Systems and Knowledge Discovery. Springer-Verlag 2nd Revised edition, 2003.

[CSS+11] I. Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Okabe, Y. Sato, and A. Sugimoto. Appearance-based head pose estimation with scene-specific adaptation. *ICCV*, 2011.

[CTC10] Angela Caunce, Chris Taylor, and Tim Cootes. Improved 3d model search for facial feature location and pose estimation in 2d images. *BMVC*, 2010.

[CTCG95] Timothy F. Cootes, Christopher J. Taylor, David H. Cooper, and Jim Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1) :38–59, 1995.

[DBD14] Taner Danisman, Ioan Marius Bilasco, and Chabane Djeraba. Cross-database evaluation of normalized raw pixels for gender recognition under unconstrained settings. In *International Conference on Pattern Recognition (ICPR)*, page à paraître, 2014.

[DG04] Stephane Derrode and Faouzi Ghorbel. Shape analysis and symmetry detection in gray-level objects using the analytical fourier-mellin. *Signal Processing*, 84 :25 – 39, 2004.

[DH98] S.C. Dakin and A.M. Herbert. The spatial region of integration for visual symmetry detection. *Proceedings of the Royal Society of London. Series B : Biological Sciences*, 265(1397) :659–664, 1998.

[DLD10] Afifa Dahmane, Slimane Larabi, and Chabane Djeraba. Detection and analysis of symmetrical parts on face for head pose estimation. In *International Conference on Image Processing (ICIP)*, 2010.

[DLDB12] Afifa Dahmane, Slimane Larabi, Chabane Djeraba, and Ioan Marius Bilasco. Learning symmetrical model for head pose estimation. In *ICPR*, pages 3614–3617, 2012.

[DM11] Mohamed Dahmane and Jean Meunier. Object representation based on gabor wave vector binning : An application to human head pose detection. *ICCV*, 2011.

[DOCA11] Meltem Demirkus, Boris Oreshkin, James J. Clark, and Tal Arbel. Spatial and probabilistic codebook template based head pose estimation from unconstrained environments. *ICIP*, 2011.

[DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005.

[FA91] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 13(9) :891–906, 1991.

[FF95] B. Funt and G. Finlayson. Color constant color indexing. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 17(5) :522–529, 1995.

[FHT98] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression : a statistical view of boosting. *Annals of Statistics*, 28 :2000, 1998.

[Fis36] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7 :179–188, 1936.

[FS97] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Computer and System Sciences*, 55(1) :119–139, 1997.

[GACZE12] N. Ghadarghadar, E. Ataer-Cansizoglu, Peng Zhang, and D. Erdogmus. A sift-point distribution-based method for head pose estimation. In *Machine Learning for Signal Processing (MLSP)*, pages 1–4, 2012.

[GBK01] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many : Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 23(6) :643–660, 2001.

[GFDH08] Guodong Guo, Yun Fu, Charles R. Dyer, and Thomas S. Huang. Head pose estimation : Classification or regression ? *International Conference on Pattern Recognition (ICPR)*, 2008.

[GH04] Martin Gruendig and Olaf Hellwich. 3d head pose estimation with symmetry based illumination model in low resolution video. In *Lecture Notes in Computer Science*, volume 3175, pages 45–53. Springer, 2004.

[GHC04] N. Gourier, D. Hall, and J. L. Crowley. Estimating face orientation from robust detection of salient facial features. In *Pointing 2004, ICPR Workshop on Visual Observation of Deictic Gestures*, 2004.

[GMC96] Shaogang Gong, Stephen McKenna, and John J. Collins. An investigation into face pose distributions. In *International Conference on Automatic Face and Gesture Recognition (FG '96)*, pages 265–, 1996.

[GMCB10] Shalini Gupta, Mia K. Markey, Kenneth R. Castleman, and Alan C. Bovik. Texas 3d face recognition database. *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 97–100, May 2010.

[GMHC07] Nicolas Gourier, Jérôme Maisonnasse, Daniela Hall, and James L. Crowley. Head pose estimation on low resolution images. In *Proceedings of the 1st international evaluation conference on Classification of events, activities and relationships*, CLEAR'06, pages 270–280, 2007.

[HC12] Huy Tho Ho and Rama Chellappa. Automatic head pose estimation using randomly projected dense sift descriptors. In *ICIP*, pages 153–156, 2012.

[HCZZ04] Yuxiao Hu, Longbin Chen, Yi Zhou, and HongJiang Zhang. Estimating face pose by facial asymmetry and geometry. In *FGR*, pages 651–656, 2004.

[HGD12] Murad Al Haj, Jordi Gonz‘alez, and Larry S. Davis. On partial least squares in head pose estimation : How to simultaneously deal with misalignment. *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[HJ10] D.W Hansen and Q Ji. In the eye of the beholder : a survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32 :478–500, 2010.

[HN03] X. He and P. Niyogi. Locality preserving projections. *Advances in Neural Information Processing Systems*, 16, 2003.

[HPK+01] Geoffrey Holmes, Bernhard Pfahringer, Richard Kirkby, Eibe Frank, and Mark Hall. Multiclass alternating decision trees. In *ECML*, pages 161–172. Springer, 2001.

[HS81] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17 :185–203, 1981.

[HS88] C. Harris and M. Stephens. A Combined Corner and Edge Detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

[HSlTB11] Dong Huang, Markus Storer, Fernando De la Torre, and Horst Bischof. Supervised local subspace learning for continuous head pose estimation. *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[HSW98a] J. Huang, X. Shao, and H. Wechsler. Face pose discrimination using support vector machines (svm). In *International Conference on Pattern Recognition (ICPR)*, 1998.

[HSW98b] Jeffrey Huang, Xuhui Shao, and Harry Wechsler. Face pose discrimination using support vector machines (svm). In *Proceedings of the 14th International Conference on Pattern Recognition-Volume 1 - Volume 1*, ICPR '98, pages 154–, Washington, DC, USA, 1998. IEEE Computer Society.

[JE10] Sam Johnson and Mark Everingham. Clustered pose and nonlinear appearance models for human pose estimation. *BMVC*, 2010.

[Jia09] Hao Jiang. Human pose estimation using consistent max-covering. *ICCV*, 2009.

[JLS+11] Hao Ji, Risheng Liu, Fei Su, Zhixun Su, and Yan Tian. Robust head pose estimation via convex regularized sparse regression. *ICIP*, 2011.

[JV03] Michael Jones and Paul Viola. Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96*, 3 :14, 2003.

[KPM97] N. Kruger, M. Potzsch, and C. Von Der Malsburg. Determination of face position and pose with a learned representation based on labelled graphs. *Image and Vision Computing*, 15(8) :665 – 673, 1997.

[KS02] Volker Kruger and Gerald Sommer. Gabor wavelet networks for efficient head pose estimation. *IVC*, 2002.

[Kue91] Andreas Kuehnle. Symmetry-based recognition of vehicle rears. *Pattern Recogn. Lett.*, 12(4) :249–258, April 1991.

[LCR01] Yanxi Liu, Robert Collins, and William E. Rothfus. Robust midsagittal plane extraction from normal and pathological 3d neuroradiology images. *IEEE Transactions on Medical Imaging*, 20(3) :175 – 192, March 2001.

[LE06] G. Loy and J. Eklundh. Detecting symmetry and symmetric constellations of features. In *European Conference on Computer Vision (ECCV)*, pages 508–521, 2006.

[LFG+01] S.Z. Li, Qingdong Fu, Lie Gu, B. Scholkopf, Yimin Cheng, and Hongjiag Zhang. Kernel machine based learning for multi-view face detection and pose estimation. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 674 –679 vol.2, 2001.

[LFY+07] Z. Li, Y. Fu, J. Yuan, T. Huang, and Y. Wu. Query driven localized linear discriminant models for head pose estimation. In *International Conference on Multimedia and Expo (ICME)*, pages 1810–1813, 2007.

[LGSL04] Yongmin Li, Shaogang Gong, Jamie Sherrah, and Heather Liddell. Support vector machine based multi-view face detection and recognition. *Image and Vision Computing*, 22(5) :413 – 427, 2004.

[LHT04] Stephen R.H. Langton, Helen Honeyman, and Emma Tessler. The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception and Psychophysics*, 66(5) :752–771, 2004.

[LLL09] Xiangyang Liu, Hongtao Lu, and Heng Luo. A new representation method of head images for head pose estimation. *ICIP*, 2009.

[LLL10] Xiangyang Liu, Hongtao Lu, and Wenbin Li. Multi-manifold modeling for head pose estimation. *ICIP*, 2010.

[LMH+09a] T. Luhandjula, E. Monacelli, Y. Hamam, B.J. van Wyk, and Q. Williams. Visual intention detection for wheelchair motion. *ISVC, LNCS*, 2009.

[LMH+09b] T. Luhandjula, E. Monacelli, Y. Hamam, B.J. Van Wyk, and Q. Williams. Visual intention detection for wheelchair motion. In *International Symposium on Visual Computing (ISVC)*, pages 407–416, 2009.

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60 :91–110, 2004.

[LS99] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755) :788–791, October 1999.

[LS10] Andreas Launila and Josephine Sullivan. Contextual features for head pose estimation in football games. *ICPR*, 2010.

[lTBC+03] Ying li Tian, Lisa Brown, Jonathan Connell, Sharat Pankanti, Arun Hampapur, Andrew Senior, and Ruud Bolle. Absolute head pose estimation from overhead wide-angle cameras. In *In IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.

[LvWKW06] K.T. Luhandjula, B.J. van Wyk, K. Kith, and M.A. Van Wyk. Eye detection for fatigue assessment. *Proceeding of the Seventeenth International Symposium of the Pattern Recognition Society of South Africa*, 2006.

[LWD10] Yali Li, Shengjin Wang, and Xiaoqing Ding. Person-independent head pose estimation based on random forest regression. *ICIP*, 2010.

[LYT11] Ce Liu, J. Yuen, and A. Torralba. Sift flow : Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5) :978–994, 2011.

[LZ03] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8) :959 – 973, 2003.

[LZD08] A. Lablack, Zhongfei Zhang, and C. Djeraba. Supervised learning for head pose estimation using svd and gabor wavelets. In *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*, pages 592–596, 2008.

[MCDT07] E. Murphy-Chutorian, A. Doshi, and M.M. Trivedi. Head pose estimation for driver assistance systems : A robust algorithm and experimental evaluation. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 709–714. IEEE, 2007.

[MCT09] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision : A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(4) :607–626, 2009.

[MGM05] M. Mancas, B. Gosselin, and B. Macq. Fast and automatic tumoral area localisation using symmetry. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 725–728, 2005.

[MGP06] N. J. Mitra, L. Guibas, and M. Pauly. Partial and approximate symmetry detection for 3d geometry. In *ACM Transactions on Graphics*, volume 25, pages 560–568, 2006.

[MGP07] N. J. Mitra, L. Guibas, and M. Pauly. Symmetrization. In *ACM Transactions on Graphics*, volume 26, pages 1–8, 2007.

[MKK+06] Yong Ma, Yoshinori Konishi, Koichi Kinoshita, Shihong Lao, and Masato Kawade. Sparse bayesian regression for head pose estimation. *International Conference on Pattern Recognition (ICPR)*, 3 :507–510, 2006.

[ML04] Sinjini Mitra and Yanxi Liu. Local facial asymmetry for expression classification. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 889–894, 2004.

[MWM10] Louis-Philippe Morency, Jacob Whitehill, and Javier Movellan. Monocular head pose estimation using generalized adaptive view-based appearance model. *Image Vision Comput.*, 28(5) :754–761, May 2010.

[MZS+06] Bingpeng Ma, Wenchao Zhang, Shiguang Shan, Xilin Chen, and Wen Gao. Robust head pose estimation using lgbp. *ICPR*, 2006.

[OFL07] Mustafa Ozuysal, Pascal Fua, and Vincent Lepetit. Fast keypoint recognition in ten lines of code. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[OGX09] Javier Orozco, Shaogang Gong, and Tao Xiang. Head pose classification in crowded scenes. *BMVC*, 2009.

[OPH96] Timo Ojala, Matti Pietikinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, pages 51–59, 1996.

[PBT11] Maria Paterakiy, Haris Baltzakisy, and Panos Trahanias. Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation. *ICCV*, 2011.

[PDG08] Vinod Pathangay, Sukhendu Das, and Thomas Greiner. Symmetry-based face pose estimation from a single uncalibrated view. *8th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008),The Netherlands*, pages 1–8, 2008.

[Pea01] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11) :559–572, 1901.

[PS10] Raymond Ptucha and Andreas Savakis. Pose estimation using facial feature points and manifold learning. *ICIP*, 2010.

[PZJ05] Y. Pan, H. Zhu, and R. Ji. *3-D Head Pose Estimation for Monocular Image*. Fuzzy Systems and Knowledge Discovery. Springer, 2005.

[RBK98] H. A. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 38–, 1998.

[RBN10] Rui Rodrigues, Joao P. Barreto, and Urbano Nunes. Camera pose estimation using images of planar mirror reflections. *ECCV*, 2010.

[RO09] Elisa Ricci and Jean-Marc Odobez. Learning large margin likelihoods for realtime head pose tracking. *ICIP*, 2009.

[RR98] R. Rae and H. Ritter. Recognition of human head orientation based on artificial neural networks. *IEEE Transactions on Neural Networks*, 9(2) :257–265, 1998.

[RR06] Neil Robertson and Ian Reid. Estimating gaze direction from low-resolution faces in video. In *European Conference on Computer Vision (ECCV)*, volume 3952, pages 402–415. 2006.

[RS00] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE*, 290 :2323–2326, 2000.

[RY08] A. Ranganathan and M.-H. Yang. Online sparse matrix gaussian process regression and vision applications. *ECCV*, 2008.

[Sal94] StevenL. Salzberg. C4.5 : Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning*, 16(3) :235–240, 1994.

[SBB03] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(12) :1615–1618, 2003.

[SBH10] S. Sedai, M. Bennamoun, and D. Q. Huynh. Localized fusion of shape and appearance features for 3d human pose estimation. *BMVC*, 2010.

[SC00] Bernt Schiele and James L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *Int. J. Comput. Vision*, 36(1) :31–50, January 2000.

[SG01] Jamie Sherrah and Shaogang Gong. Fusion of perceptual cues for robust tracking of head pose and position. *Pattern Recognition*, 34(8) :1565–1572, 2001.

[SGO99] Jamie Sherrah, Shaogang Gong, and Eng-Jon Ong. Understanding pose discrimination in similarity space. *BMVC*, 1999.

[SS99] C. Sun and D. Si. Fast reflectional symmetry detection using orientation histograms. *Real-Time Imaging*, 5(1) :63 – 74, 1999.

[ST96] Yoshinobu Sato and Shinichi Tamura. Detecting planar and curved symmetries of 3d shapes from a range image. *Computer Vision and Image Understanding*, 64(1) :175–187, 1996.

[Ste05] Fred Stentiford. Attention based facial symmetry detection. In *Proc. ICAPR*, 2005.

[Sti04] R. Stiefelhagen. Estimating head pose with neural networks. In *Pointing 2004, ICPR Workshop on Visual Observation of Deictic Gestures*, 2004.

[SYW02] Rainer Stiefelhagen, Jie Yang, and Alex Waibel. Modeling focus of attention for meeting indexing based on multiple cues. *IEEE Transactions on Neural Networks*, 13 :928–938, 2002.

[TFHH06] J. Tu, Y. Fu, Y. Hu, and T. Huang. Evaluation of head pose estimation for studio data. In *International Workshop Classification of Events Activities and Relationships (CLEAR)*, 2006.

[THM95] C. W. Tyler, L. Hardage, and R. T. Miller. Multiple mechanisms for the detection of mirror symmetry. *Spatial Vision*, 9 :79100, 1995.

[TPPK11] T. Theoharis, P. Perakis, G. Passalis, and I. A. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3d face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(10) :1938–1951, 2011.

[TSL00] J. B. Tenenbaum, V. Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500) :2319–2323, 2000.

[VBB07] T. Vatahska, M. Bennewitz, and S. Behnke. Feature-based head pose estimation from images. In *7th International Conference on Humanoid Robots*, 2007.

[VG09] Roberto Valenti and Theo Gevers. Robustifying eye center localization by head pose cues. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, 2001.

[VNS06] M. Voit, K. Nickel, and R. Stiefelhagen. Neural network-based head pose estimation and multi-view fusion. In *International Workshop Classification of Events Activities and Relationships (CLEAR)*, 2006.

[VSG12] Roberto Valenti, Nicu Sebe, and Theo Gevers. Combining head pose and eye location information for gaze estimation. *IEEE Transactions on Image Processing*, 21(2) :802–815, 2012.

[WFLC00] H. R. Wilson, F.Wilkinson, L. Lin, and M. Castillo. Perception of head orientation. *Vision Research*, 40(5) :459–472, 2000.

[WHGY08] Xianwang Wang, Xinyu Huang, Jizhou Gao, and Ruigang Yang. Illumination and person-insensitive head pose estimation using distance metric learning. *ECCV*, 2008.

[WPP+04] J. Wu, J. Pedersen, D. Putthividhya, D. Norgaard, and M. M. Trivedi. A two-level pose estimation framework using majority voting of gabor wavelets and bunch graph analysis. In *Pointing 2004, ICPR Workshop on Visual Observation of Deictic Gestures*, 2004.

[WS07] Jian-Gang Wang and Eric Sung. Em enhancement of 3d head pose estimated by point at infinity. *Image Vision Comput.*, 25(12) :1864–1874, December 2007.

[WT00] Y. Wu and K. Toyama. Wide range illumination insensitive head orientation estimation. In *International Conference on Automatic Face and Gesture Recognition (AFGR)*, pages 183–188, 2000.

[WT08] Junwen Wu and Mohan M. Trivedi. A two-stage head pose estimation framework and evaluation. *Pattern Recogn.*, 41(3) :1138–1158, March 2008.

[WYS12] Jian-Gang Wang, Wei-Yun Yau, and E. Sung. Head pose estimation by bootstrapping generalized discriminant analysis with sift flow alignment criterion. In *International Conference on Biometrics (ICB)*, pages 32–39, 2012.

[XBMK04] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 535–542, 2004.

[XMKC03] Jing Xiao, Tsuyoshi Moriyama, Takeo Kanade, and Jeffrey F. Cohn. Robust full-motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology*, 13(1) :85–94, 2003.

[ZF03] Yan Zhang and Jufu Feng. Eliminating variation of face images using face symmetry. *LNCS*, 2003.

[ZHLH07] Zhenqiu Zhang, Yuxiao Hu, Ming Liu, and Thomas Huang. Head pose estimation in seminar room using multi view face detectors. In *Multimodal Technologies for Perception of Humans*, volume 4122, pages 299–304. 2007.

Annexe A

Acquisition de bases de données pour l'estimation de la pose de la tête

Différentes techniques sont utilisées pour l'acquisition et l'annotation des bases dédiées à l'estimation de la pose de la tête. Chacune des bases publiques que nous avons utilisées utilise sa propre méthode.

A.1 La base FacePix

Une plateforme dédiée a été construite. Cette plateforme est constituée de deux anneaux rotatifs de manière indépendante qui sont chacun capables de supporter l'éclairage et l'équipement de l'appareil, tout en le déplaçant de manière contrôlée avec précision autour du participant. Le dispositif est illustré dans la figure A.1. Le participant est assis dans la partie centrale qui ne tourne pas de la plate forme, focalisé sur un point immobile. La caméra (non représentée sur la figure) ainsi qu'une toile qui fait office d'arrière plan blanc, sont posés de manière opposée sur un anneau rotatif.

Les anneaux sont motorisés, permettant à une caméra vidéo de balayer autour de l'avant du participant. Ceci permet une capture rapide sur toute la plage d'angles de pose dans un

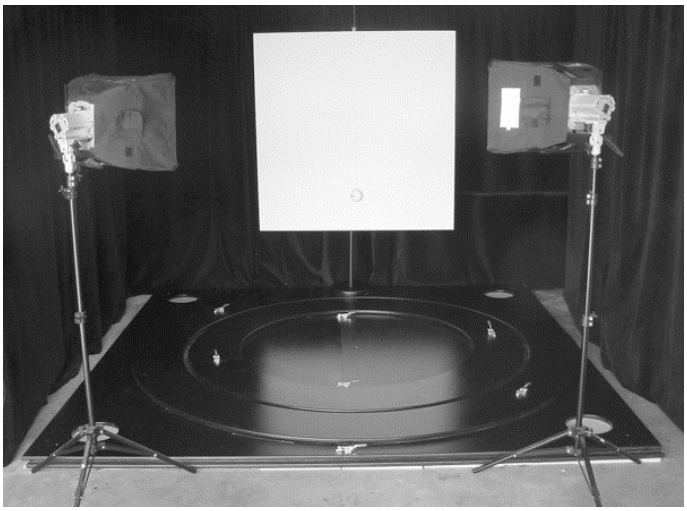

FIGURE A.1: Plateforme utilisée pour capturer les images de la base FacePix.

court laps de temps, réduisant ainsi le temps que le participant doit rester immobile. Le taux de trame rapide de la caméra vidéo veille à ce que, même avec les clins d'œil, il sera possible de capturer une image avec des yeux ouverts pour chaque angle. Le résultat est une image pour chaque angle de rotation dans l'intervalle $\pm 90°$ avec un pas de $1°$.

A.2 La base CMU PIE

Une pièce 3D a été aménagé afin de capturer des images sous différentes poses. Les participants sont assis et restent immobiles. Plusieurs caméras fixées à des emplacements précis, capturent des images simultanément. Chaque caméra correspond à une orientation de la tête. La figure A.2 représente la pièce 3D utilisée.

A.3 La base Boston University

Des séquences vidéos des participants effectuant des mouvements aléatoires de la tête, sont enregistrées. La pose de la tête est enregistrée par un capteur FoB (*de l'anglais Flock of Birds*). L'information est obtenue d'un émetteur magnétique placé sur la tête des participants. Le

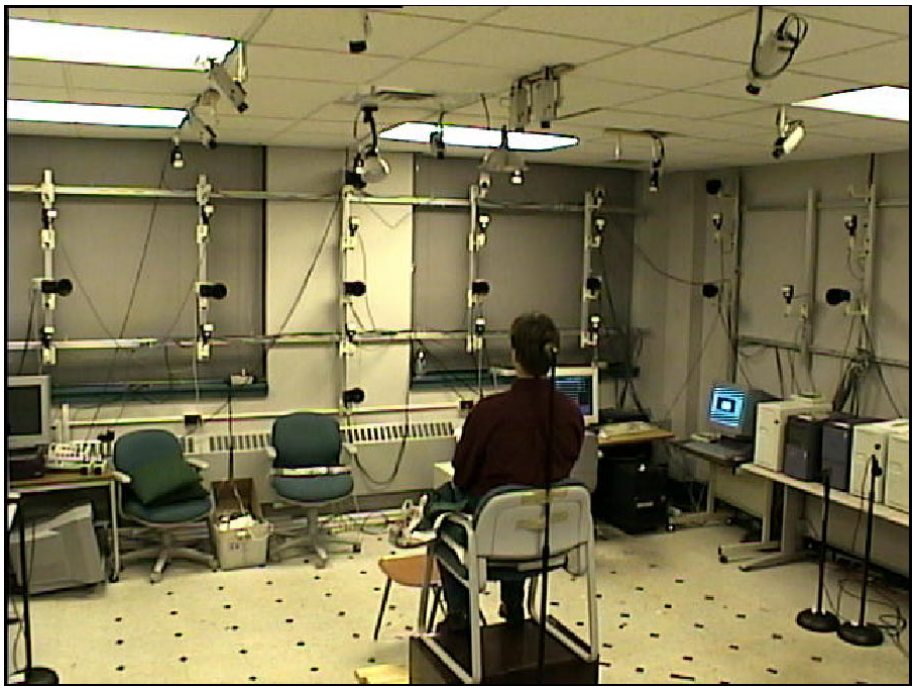

FIGURE A.2: Système de capture utilisé pour la base CMU PIE.

système est illustré dans la figure A.3.

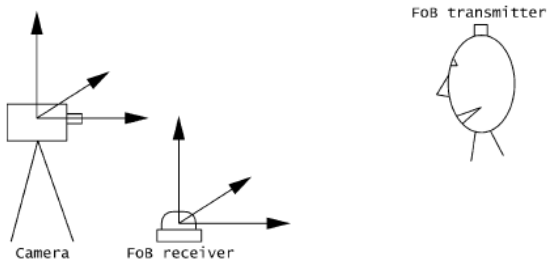

FIGURE A.3: Système de capture utilisé pour la base BU.

Annexe B

Détection du visage

Nous avons utilisé dans nos expérimentations, l'implémentation du détecteur de Viola et Jones disponible dans OpenCV, afin de localiser le visage dans une image. Ce détecteur permet de localiser des visages de face et de profil. Toutefois, les rotations dans le plan ne dépassent pas 15° de chaque coté.

Nous avons dû adapter l'algorithme afin de pouvoir traiter le mouvement roll dans un intervalle plus grand que ±15°. Lorsqu'aucun visage n'est détecté, nous effectuons une rotation de l'image de 20°. Si la détection n'aboutit toujours pas, nous effectuons la rotation du coté opposé (*cf.* figure B.1). Ceci nous permet de mettre le visage à l'intérieur de l'intervalle ±15°. Si un visage est détecté sur l'image retournée, sa position est utilisée pour calculer une position relative du visage sur l'image originale.

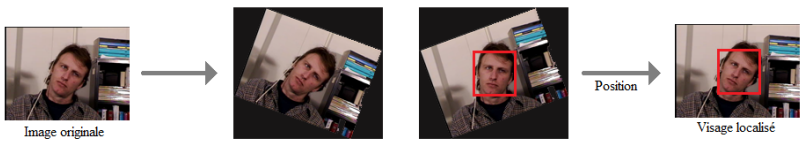

FIGURE B.1: Détection de visage avec rotation importante dans le plan.

Annexe C

Publications

Revue internationale

Afifa Dahmane, Slimane Larabi, I. M. Bilasco et Chabane Djeraba *Head Pose Estimation Based on Face Symmetry Analysis*, Journal of Signal, Image and Video Processing 8(5), 2014.

Conférences internationales

Afifa Dahmane, Slimane Larabi, et Chabane Djeraba *Detection and analysis of symmetrical parts on face for head pose estimation*, International Conference on Image Processing (ICIP), Septembre 26-29, 2010 Hong Kong.

Afifa Dahmane, Slimane Larabi, I. M. Bilasco, Chabane Djeraba *Estimation discrète de l'angle Pan de la tête*, Compression et Représentation des Signaux Audiovisuels (CORESA), Lille, France, 24-25 Mai 2012.

Afifa Dahmane, Slimane Larabi, I. M. Bilasco, Chabane Djeraba *Learning symmetrical Model for Head Pose Estimation*, 21st International Conference on Pattern Recognition (ICPR), November 11-15, 2012 Japon.