



HAL
open science

Dépliage d'images omnidirectionnelles optimisé pour la détection de visages

Yohan Dupuis, Arturo Mendoza Quispe, Pascal Vasseur, Benjamin Castaneda,
Nicolas Ragot

► **To cite this version:**

Yohan Dupuis, Arturo Mendoza Quispe, Pascal Vasseur, Benjamin Castaneda, Nicolas Ragot. Dépliage d'images omnidirectionnelles optimisé pour la détection de visages. *Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014*, Jun 2014, Rouen, France. hal-00988791

HAL Id: hal-00988791

<https://hal.science/hal-00988791>

Submitted on 9 May 2014

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.

Dépliage d'images omnidirectionnelles optimisé pour la détection de visages

Y. Dupuis¹

A. Mendoza Quispe²

P. Vasseur³

B. Castaneda²

N. Ragot⁴

¹ CEREMA, 76120 Le Grand Quevilly, France, yohan.dupuis@cerema.fr

² Pontificia Univ. Catolica del Peru, Lima, Pérou, a20094311@pucp.edu.pe, castaneda.b@pucp.edu.pe

³ LITIS Lab, Université de Rouen, 76821 Saint-Etienne-du-Rouvray, France, pascal.vasseur@univ-rouen.fr

⁴ ESIGELEC-IRSEEM, 76801 Saint-Etienne-du-Rouvray, France, nicolas.ragot@esigelec.fr

Résumé

Cet article aborde la problématique de détection de visages basée sur l'algorithme de Viola-Jones dans des images omnidirectionnelles dépliées. Notre approche se fonde sur une heuristique qui optimise le dépliement de l'image omnidirectionnelle pour les attributs rectangulaires de Haar. Une comparaison des performances obtenues entre notre approche et celles d'un dépliement sphérique confirme la pertinence et l'efficacité de notre méthode. L'impact du seuil de décision et de la densité de fenêtres détectées sont également étudiés.

Mots Clef

Détection de visages, vision omnidirectionnelle, dépliement optimisé, Viola-Jones

1 Introduction

La vision omnidirectionnelle, tant d'un point de vue théorique qu'en termes d'applications, est un champ de recherche très actif. Les capteurs omnidirectionnels (catadioptrique, fish-eye, etc.) sont utilisés dans des domaines d'applications variés, en robotique notamment pour des problématiques de navigation ou d'estimation de mouvement [1, 2]. Leur large champ de vue, sans rupture géométrique et temporelle, est également apprécié pour des applications en reconstruction 3D d'environnements [3]. Pour ces domaines d'application, les traitements (détection de points d'intérêt et analyse des huit voisins) sont directement appliqués dans l'image omnidirectionnelle. Il s'agit donc d'une approche locale qui ne souffre pas des distorsions importantes liées à la nature du capteur [4].

Plus récemment, des applications en détection, reconnaissance et suivi de formes ont été publiées. Pour ces dernières, deux catégories de travaux se distinguent :

- La première fait référence à l'utilisation de l'image dépliée, c'est-à-dire une représentation pseudo-perspective de l'image omnidirectionnelle. Les techniques et algorithmes développés pour des images perspectives peuvent alors être appliqués. Les résultats récents obtenus en détection de visage à partir de l'algorithme de Viola et Jones sont prometteurs [5, 6].
- La seconde fait référence à l'utilisation de l'image omni-

directionnelle, sans qu'aucun traitement supplémentaire ne soit appliqué. Masselli et al. proposent un algorithme de détection de balles basé sur les paramètres de HoG [7]. Les auteurs utilisent l'invariance en rotation de l'objet. Dès lors, l'objet possède une représentation unique quelque soit sa position dans l'image. Dans [8], Dupuis et al. s'intéressent aux problématiques de détection de visages dans l'image omnidirectionnelle. La représentation d'un visage dans ce type d'images est modifiée en fonction de sa position selon le rayon et l'axe optique. Les résultats soulignent une augmentation des performances de détection pour un détecteur spécifiquement entraîné qui tient compte des distorsions radiales de l'image omnidirectionnelle. Néanmoins, et ce malgré des résultats prometteurs, cette approche n'intègre pas encore la rotation dans le plan.

Ces deux catégories d'approches possèdent leurs avantages et leurs inconvénients :

- Les approches utilisant l'image omnidirectionnelle dépliée est séduisante puisqu'elle ne nécessite pas l'entraînement d'un détecteur spécifique. Cependant, différentes limitations apparaissent : la projection de l'image omnidirectionnelle sous forme pseudo-perspective introduit un temps de calcul supplémentaire ; des problèmes de détection apparaissent lorsque les objets sont situés aux limites de l'image dépliée.
- Les approches directes fournissent de meilleurs résultats, tant en taux de détection qu'en temps de calcul. Néanmoins, un entraînement spécifique doit être réalisé afin de prendre en compte les distorsions radiales de l'image omnidirectionnelle. Plusieurs détecteurs sont également nécessaires pour intégrer la rotation dans le plan. Ce dernier point nécessite des temps de calcul importants et un réglage fin.

Dans ce papier, nous proposons une nouvelle approche de dépliement qui améliore significativement les performances de l'algorithme de Viola et Jones en comparaison des déploiements sphérique et cylindrique. De fait, le déploiement de l'application n'est pas contraint par le processus d'entraînement. L'article est structuré comme suit : le paragraphe 2 détaille notre méthodologie, le paragraphe 3 présente nos résultats. Finalement, une analyse des ré-

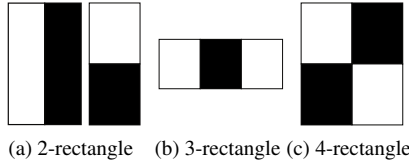


FIGURE 1 – Exemples d'attributs de Haar

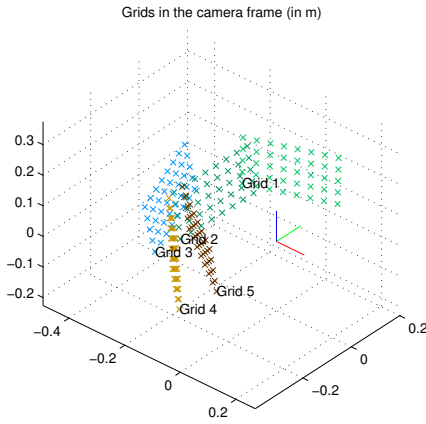


FIGURE 2 – Positions 3D des mires

sultats ainsi qu'une conclusion sont données dans le paragraphe 4.

2 Méthodologie

Le détecteur de visages de Viola et Jones utilise les attributs de Haar comme descripteurs (cf. Figure 1). La somme des pixels des régions claires est soustraite à la somme des pixels des régions sombres. Chaque attribut est défini par sa taille et sa position dans la région d'intérêt. De fait, un nombre important d'attributs existe. Une procédure de type Adaboost est alors appliquée pendant la phase d'apprentissage et permet ainsi de réduire le nombre d'attributs à évaluer [9]. Des arbres de décision sont alors utilisés comme classifieur faible pour déterminer le seuil optimal de chaque attribut. Les détecteurs disponibles "sur étagère", sont entraînés sur des images perspectives. Dès lors, le valeur de l'attribut sera particulièrement sensible aux distorsions présentes dans l'image omnidirectionnelle [8]. A titre d'exemple, un rectangle dans une image perspective n'est pas un rectangle dans une image omnidirectionnelle. Malgré le fait que le dépliement vise à recréer une image perspective, ce dernier ne garantit pas qu'un rectangle dans une image omnidirectionnelle se projette en un rectangle de facteur de forme homogène dans l'image dépliée (cf. Figure 5a).

2.1 Dépliement d'images omnidirectionnelles

Geyer [10] et Barreto [11] ont proposé un modèle générique de formation d'images omnidirectionnelles qui dissocie les non-linéarités introduites par le miroir (cf. figure 3). Il s'agit du modèle de la sphère d'équivalence qui se fonde sur une représentation sphérique de l'image. La projection d'un point 3D est alors décrite par :

1. Un point 3D $X = [x \ y \ z]^T$, est projeté sur la sphère unitaire de centre C_m , tel que :

$$(X_s)_{C_m} = \frac{X}{\|X\|} \quad (1)$$

2. $(X_s)_{C_m} = [x_s \ y_s \ z_s]^T$ est translaté en $C_p = [0, 0, \zeta]$, tel que :

$$(X_s)_{C_p} = \begin{bmatrix} x_s \\ y_s \\ z_s + \zeta \end{bmatrix} \quad (2)$$

3. $(X_s)_{C_p}$ est ensuite projeté sur le plan normalisé en M :

$$M = \begin{bmatrix} \frac{x_s}{z_s + \zeta} & \frac{y_s}{z_s + \zeta} & 1 \end{bmatrix}^T \quad (3)$$

4. Finalement, M est ensuite projeté en P via la matrice de projection généralisée, composée des paramètres γ_u, γ_v (distances focales horizontale and verticale) et u_0, v_0 (les coordonnées du point principal de l'image). Le paramètre ζ est défini comme le paramètre du modèle unifié [12].

$$P = \begin{bmatrix} \gamma_u & \gamma_u * \alpha & u_0 \\ 0 & \gamma_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} * M \quad (4)$$

Déplier l'image omnidirectionnelle se résume alors en un échantillonnage de cette sphère unitaire. En théorie, une image omnidirectionnelle peut être reprojétée sur n'importe quel type de surface. Les dépliements les plus courants sont les dépliements cylindrique et sphérique. Pour le dépliement cylindrique, l'angle d'azimut de la sphère est échantillonné régulièrement. Celui d'élévation est échantillonné de sorte à ce que le pas en hauteur soit constant. Pour le dépliement sphérique les angles d'azimut et d'élévation sont échantillonnés régulièrement.

L'échantillonnage régulier selon l'angle d'élévation φ se traduit par une fonction linéaire $F : [\varphi_{min}, \varphi_{max}] \rightarrow [0, N_{rows}]$. Une normalisation de cette fonction permet une indépendance du dépliement aux gammes de valeurs de φ et au nombre de lignes souhaitées pour l'image panoramique ; F se définit alors par : $F : [0, 1] \rightarrow [0, 1]$ (cf. Figure 6).

2.2 Fonction coût

Nous proposons d'utiliser des mires en damier afin de définir notre dépliement optimisé, c'est-à-dire un dépliement qui conserve les proportions géométriques : un rectangle dans l'image omnidirectionnelle est un rectangle

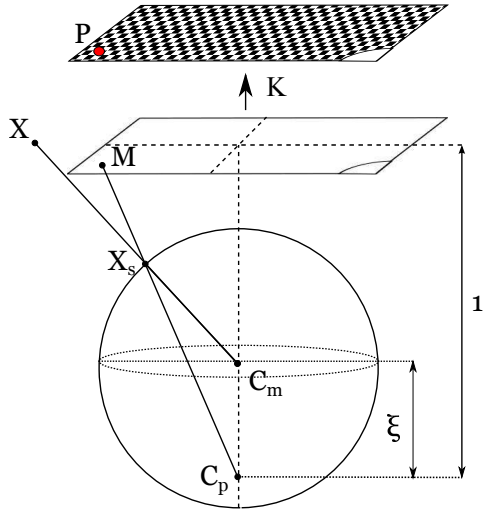


FIGURE 3 – Modèle de la sphère d'équivalence décrivant le processus de formation d'une image omnidirectionnelle

dans l'image panoramique dépliée. Les mires sont placées tout autour du capteur, à différentes positions permettant ainsi de couvrir l'ensemble du champ d'observation (cf. Figure 2). Plusieurs paramètres sont utilisés pour définir un rectangle (cf. Figure 4). Ces paramètres sont définis et combinés comme suit :

$$f_1 = \frac{\theta_1 - \theta_3}{\theta_1 + \theta_3}, f_2 = \frac{\theta_2 - \theta_4}{\theta_2 + \theta_4}, f_3 = \frac{\theta_1 - \theta_2}{\theta_1 + \theta_2}$$

f_1, f_2, f_3 sont évalués pour chaque rectangle. L'objectif visé étant de faire en sorte qu'un rectangle reste un rectangle, ces trois paramètres doivent tendre vers 0. f_3 permet de s'assurer que deux côtés consécutifs sont orthogonaux. Le dénominateur est utilisé pour garantir que chaque paramètre soit borné entre $[-1, 1]$. La fonction coût est alors définie par :

$$L = \alpha \times \underbrace{\frac{1}{3m} \sum_{i=1}^m \sum_{j=1}^3 f_j^{i2}}_{\textcircled{1}} + (1 - \alpha) \times \underbrace{\frac{1}{3} \sum_{j=1}^3 \sigma_{f_j}^2}_{\textcircled{2}} \quad (5)$$

Où :

f_j^i : paramètre j pour le rectangle i .

σ_{f_j} : écart-type du paramètre j sur l'ensemble des paramètres calculés.

α : poids

m : nombre de rectangles utilisés dans le processus d'optimisation

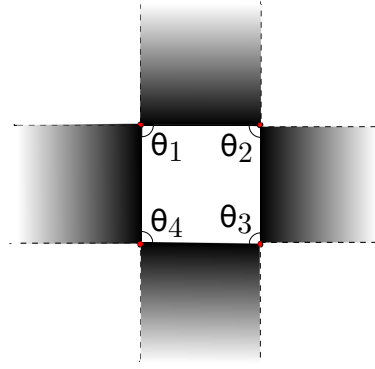


FIGURE 4 – Paramètres pour la fonction coût

La partie 1 de la fonction coût garantit une convergence des paramètres vers 0. Elle garantit également l'obtention de régions rectangulaires dans l'image dépliée. Le paramètre f_j^{i2} varie entre $[0, 1]$ puisqu'il est élevé au carré.

La partie 2 contrôle la dispersion. En effet, même si le paramètre ne peut pas être forcé à 0, nous souhaitons obtenir une forme du rectangle projeté homogène, c'est-à-dire avec un écart-type nul pour l'ensemble des motifs des mires. α est utilisé comme pondérer les parties 1 et 2 de la fonction coût.

2.3 Processus d'optimisation

L'objectif est de chercher un dépliement qui minimise la fonction coût. Nous définissons la fonction d'échantillonnage $G : [0, 1] \rightarrow [0, 1]$ comme suit :

$$\begin{aligned} G(x) &= A^T X \\ &= a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x \end{aligned} \quad (6)$$

$$\text{Telle que } G(1) = 1; \sum_{i=1}^4 a_i = 1; x = \frac{\varphi - \varphi_{min}}{\varphi_{max} - \varphi_{min}}$$

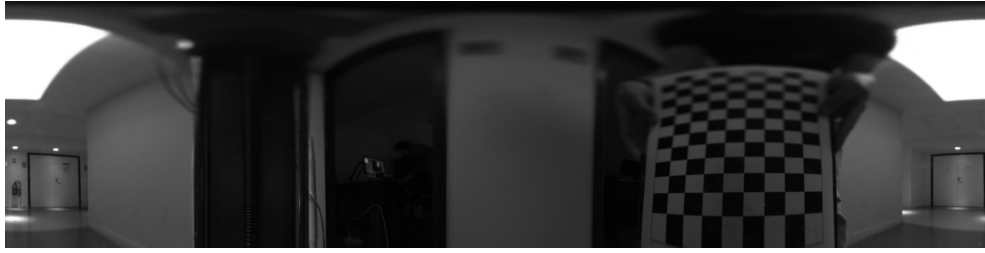
A est notre inconnue. Nous utilisons une heuristique pour estimer une solution. Nous initialisons le domaine de recherche tel que $-5 \leq a_i \leq 5$. Le pas d'échantillonnage initial est fixé à 0.1. Les combinaisons respectant les conditions définies dans l'équation 6 sont évaluées. Les vingt meilleures solutions sont retenues et utilisées pour borner le nouvel espace de recherche. Le pas d'échantillonnage est alors divisé par 10. Le processus est répété successivement jusqu'à atteindre une précision de l'ordre du milliè.

3 Résultats

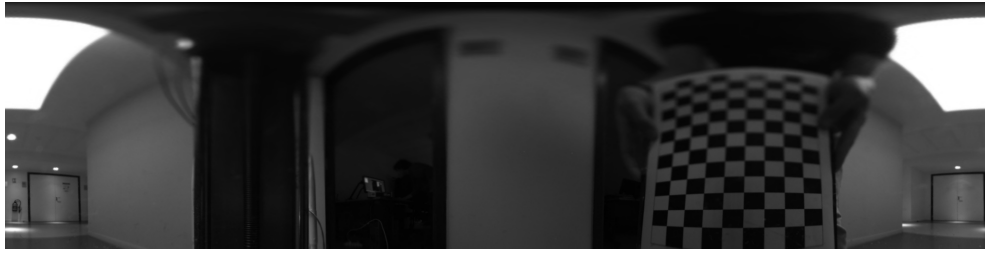
3.1 Fonction de dépliement

Les attributs de Haar étant rectangulaires, nous avons fait le choix de $\alpha = 0.5$.

Notre processus d'optimisation est implémenté dans un environnement Matlab et s'exécute en 5 minutes sur un processeur i7-3720QM. Ce temps d'exécution est significativement réduit en comparaison du temps nécessaire



(a) Sphérique



(b) Polynomial

FIGURE 5 – Exemples de dépliement : sphérique et polynomial. Le dépliement polynomial permet de conserver une forme homogène des motifs de la mire, notamment dans le bas de l’image dépliée.

pour entraîner un ensemble de détecteurs de visages spécifiques aux images omnidirectionnelles.

Nous avons réalisé notre entraînement sur des images obtenues à partir d’un capteur catadioptrique constitué d’une caméra CCD possédant une résolution de 1280×1024 .

Nous avons obtenu la solution suivante $A = [-0.390; 1.510; -1.804; 1.674]^T$ conduisant à une valeur de la fonction coût égale à : $L = 0.00327$. En comparant les Figures 5b et 5a, nous constatons que notre approche de dépliement est plus homogène puisqu’elle permet une meilleure conservation de la taille et de la forme des rectangles projetés. Les résultats attendus en termes de taux de détection avec l’algorithme de Viola et Jones devraient donc être sensiblement meilleurs à partir de notre dépliement (cf. Figure 5b).

Dans la Figure 6, nous comparons les fonctions de dépliement. Comme le souligne Dupuis et al. [6], le dépliement sphérique conduit à de meilleurs résultats que le dépliement cylindrique. Ainsi, notre dépliement polynomial n’est à comparer qu’avec les performances d’un dépliement sphérique. Le dépliement sphérique correspond à la fonction identité.

En considérant les orientations des mires utilisées dans notre processus de minimisation (cf. Figure 2), nous aurions pu nous attendre à une solution proche d’un dépliement cylindrique. En fait, les deux courbes sont très éloignées l’une de l’autre, ce qui indique que notre solution possède d’importantes non-linéarités avec une approche basée sur un dépliement cylindrique.

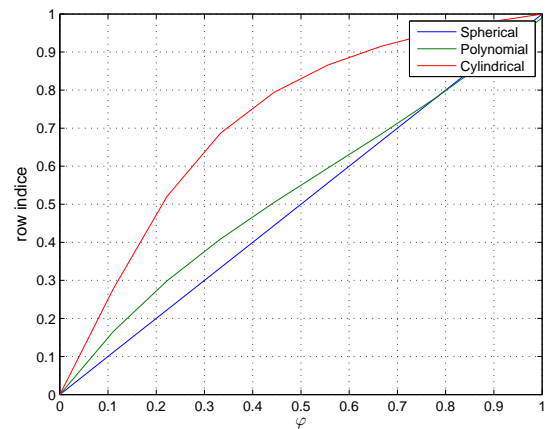
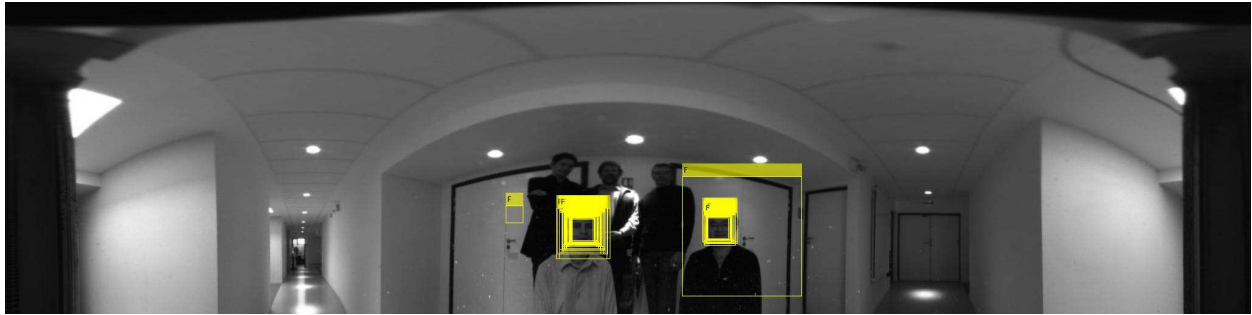
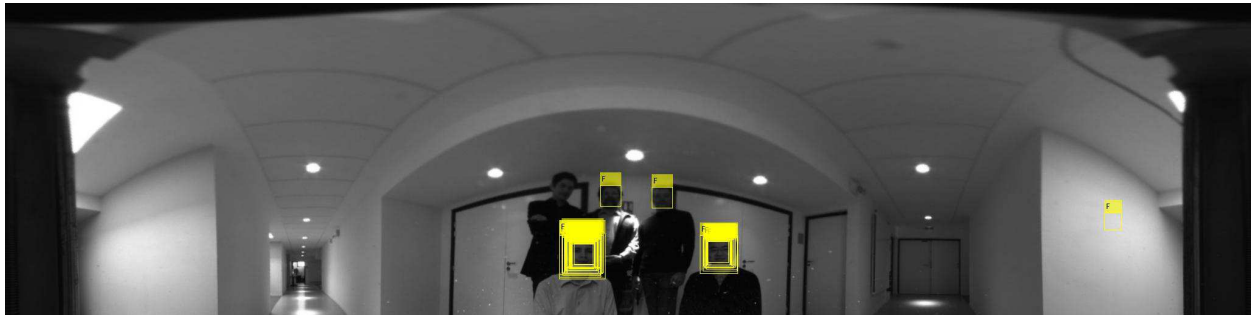


FIGURE 6 – Fonctions de dépliement : cylindrique, sphérique, polynomiale

Nous remarquons également que pour un pas d’échantillonnage identique (cf. axe des abscisses) dans la partie basse de l’image, la fonction de dépliement polynomial possède un angle d’élévation plus important que celle du dépliement sphérique. Ceci induit que les rectangles dans l’image dépliée sont réguliers (possédant des distorsions réduites). Dans la partie haute de l’image, le pas d’échantillonnage est plus faible ce qui réduit l’impression de tassement des rectangles une fois projetés.



(a) Sphérique



(b) Polynomial

FIGURE 7 – Exemples de détections

$$\tau_d = 0$$

3.2 Performances du détecteur de visages

Notre base de test est constituée de soixante six images contenant soixante quinze visages à détecter. Cette base de test est identique à celle utilisée par Dupuis et al. [8].

Nous comparons les performances des dépliements sphérique et polynomial. La Figure 8 illustre les courbes ROC pour les deux approches. Le taux de vrais positifs (True Positive Rate, TPR) est défini comme une fonction du nombre de faux positifs. Un visage est défini comme vrai positif si 70% de son aire est partagée avec un visage tel qu'annoté manuellement dans la base de test.

Nous évaluons également l'impact du seuil de fusion de détections multiples noté τ_d . Ce dernier représente le nombre de sous-fenêtres candidates nécessaires pour définir si la sous-fenêtre de chevauchement est à considérer comme un visage.

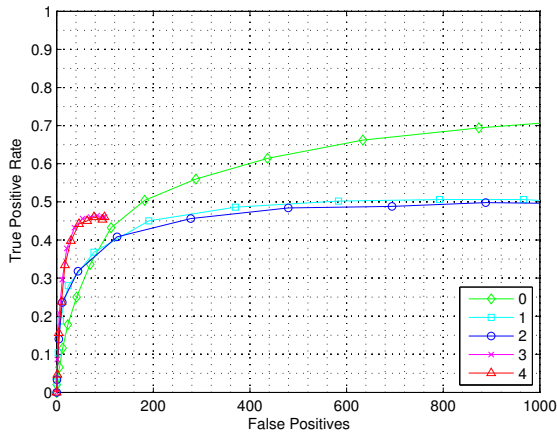
Pour chaque valeur de τ_d , le dépliement polynomial fournit de meilleurs résultats que le dépliement sphérique. En comparant les courbes ROC dans leurs parties asymptotiques, pour $\tau_d = \{0, 1, 2\}$ le dépliement polynomial présente un écart absolu de +15% par rapport au dépliement sphérique. Cet écart est d'autant plus significatif pour $\tau_d = \{3, 4\}$ et atteint +25%. Ces figures mettent en avant l'apport significatif du dépliement polynomial en comparaison d'un dépliement sphérique. Quelque soit le taux de détection, le nombre de faux positifs pour le dépliement polynomial est toujours plus faible que celui obtenu pour le dépliement sphérique. Il s'agit d'un point fort de notre méthode de dépliement.

τ_d impacte également les performances de détection. En effet, τ_d vise à réduire le nombre de faux positifs tout en maintenant une valeur correcte pour le taux de vrais positifs. Une comparaison des performances obtenues pour $\tau_d = 0$ et $\tau_d = 4$ met en avant notre approche : la Figure 8a illustre une variation importante de la courbe ROC sur sa partie asymptotique de 70% à 45% pour le dépliement sphérique. Cette variation est nettement moins prononcée (de 85% à 70%) pour le dépliement polynomial.

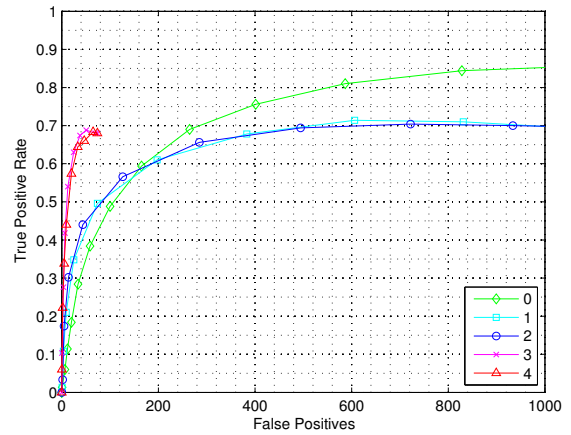
Pour un taux de vrais positifs de 45% pour le dépliement polynomial et 70% pour le dépliement sphérique, les deux méthodes donnent chacune quatre-vingt faux positifs.

La Figure 7 illustre des exemples de détections réalisées sur une image test dépliée par les deux approches (sphérique et polynomial). Comme cela est pressenti, notre dépliement polynomial fournit de meilleurs résultats puisqu'un nombre plus important de visages a été détecté. Le nombre de faux positifs est également meilleur dans le cas du dépliement polynomial (2 pour le dépliement sphérique et 1 pour le dépliement polynomial).

Dans la Figure 9, nous analysons l'influence des frontières de décision sur le taux relatif de vrais positifs. La frontière de décision par défaut est indiquée par la ligne rouge. Le taux de vrais positifs du dépliement sphérique commence à diminuer avant cette frontière de décision. Le dépliement polynomial reste quant à lui constant au delà de cette frontière pour ensuite décroître. Ceci signifie que quelque soit la dernière étape de la cascade attentionnelle, le centroïde du visage est plus proche de la frontière de décision par défaut



(a) Sphérique



(b) Polynomial

FIGURE 8 – Courbes ROC

Les valeurs de τ_a sont indiquées dans la légende des figures

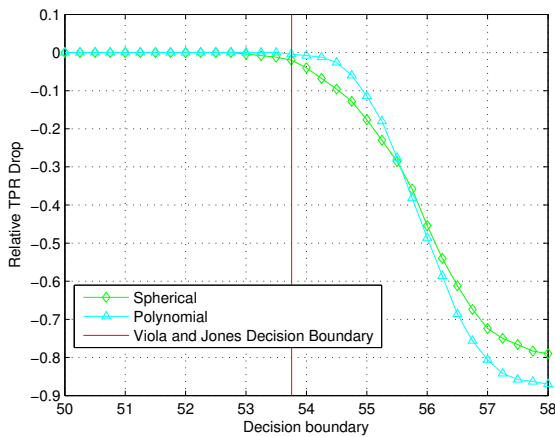


FIGURE 9 – Analyse de la confiance de classification

$$\tau_a = 0$$

dans le cas d'un dépliement sphérique que pour notre dépliement polynomial. En d'autres termes, les visages dans l'image dépliée polynomiale ont une apparence qui se rapproche plus de visages d'une image perspective dans l'espace des attributs de Haar que pour un dépliement sphérique.

4 Conclusion

Dans cet article, nous avons présenté et détaillé une approche qui permet une amélioration significative des performances de l'algorithme de Viola et Jones pour la détection de visages dans des images omnidirectionnelles dépliées. Notre approche ne nécessite aucun entraînement spécifique. Nos résultats mettent en avant qu'un dépliement optimisé par rapport aux attributs de Haar induit une amélioration significative des performances de détection.

Nos prochains travaux s'intéresseront tout d'abord à l'amélioration du processus de minimisation : une définition de paramètres globaux sera alors étudiée. Nous testerons également les performances des classificateurs pour d'autres attributs et détecteurs de Haar.

Remerciements

Le projet COALAS a été sélectionné dans le cadre du programme européen de coopération transfrontalière INTERREG IVA France (Manche) - Angleterre, cofinancé par le FEDER.

Références

- [1] J. Bazin, C. Demonceaux, P. Vasseur, and I. Kweon, "Motion Estimation by Decoupling Rotation and Translation in Catadioptric Vision," *Computer Vision and Image Understanding*, vol. 114, pp. 254–273, 2009.
- [2] C. Mei, E. Sommerlade, G. Sibley, P. Newman, and I. Reid, "Hidden View Synthesis using Real-Time Visual SLAM for Simplifying Video Surveillance Analysis," in *IEEE International Conference on Robotics and Automation*, May 2011, pp. 4240–4245.
- [3] M. Lhuillier, "Toward automatic 3D modeling of scenes using a generic camera model," in *CVPR*, 2008.
- [4] P. Sturm, S. Ramalingam, J. Tardif, S. Gasparini, and J. Barreto, "Camera Models and Fundamental Concepts Used in Geometric Computer Vision," *Foundations and Trends® in Computer Graphics and Vision*, vol. 6, no. 1–2, pp. 1–183, 2011.
- [5] A. L. C. Barczak, J. J. Okamoto, and V. J. Grassi, "Face Tracking Using a Hyperbolic Catadioptric

Omnidirectional System,” *Res. Lett. Inf. Math. Sci.*, vol. 13, pp. 55–67, 2009.

- [6] Y. Dupuis, X. Savatier, J.-Y. Ertaud, and G. Hoblos, “A Framework for Face Detection on Central Catadioptric Systems,” in *IEEE International Workshop on Robotic and Sensors Environments*, 2010.
- [7] A. Masselli, R. Hanten, and A. Zell, “Robust real-time detection of multiple balls on a mobile robot,” in *Mobile Robots (ECMR), 2013 European Conference on*, 2013, pp. 355–360.
- [8] Y. Dupuis, X. Savatier, J.-Y. Ertaud, and P. Vasseur, “Robust Radial Face Detection for Omnidirectional Vision,” *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1808–1821, 2013.
- [9] M. Viola, M. J. Jones, and P. Viola, “Fast Multi-view Face Detection,” in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2003.
- [10] C. Geyer and K. Daniilidis, “A unifying theory for central panoramic systems and practical implications,” in *ECCV*, 2000.
- [11] J. P. Barreto and H. Araujo, “Issues on the geometry of central catadioptric image formation,” in *CVPR*, 2001.
- [12] C. Mei and P. Rives, “Single View Point Camera Calibration from Planar Grids,” in *IEEE International Conference on Robotics and Automation*, 2007, pp. 3945–3950.