
Détection radiale de visages sur images omnidirectionnelles

**Yohan Dupuis¹, Xavier Savatier², Jean-Yves Ertaud²,
Pascal Vasseur³**

1. Cerema DTerNC
Chemin de la Poudrière
F-76120 Le Grand Quevilly
yohan.dupuis@cerema.fr
2. ESIGELEC-IRSEEM
Avenue Galilée
F-76800 Saint Etienne du Rouvray
prenom.nom@esigelec.fr
3. LITIS-Université de Rouen
Avenue de l'Université
F-76800 Saint Etienne du Rouvray
pascal.vasseur@univ-rouen.fr

RÉSUMÉ. Les capteurs de vision omnidirectionnelle sont aujourd'hui couramment utilisés pour l'interprétation géométrique de scènes. Cependant, peu de travaux portent sur la détection d'objets à partir de ces capteurs. Les travaux existants passent par un dépliement des images omnidirectionnelles pour obtenir des images pseudo-perspectives. L'algorithme de détection de visages sur images perspectives est ensuite appliqué directement sur les images dépliées. Dans ces travaux, nous investiguons la manière dont les images omnidirectionnelles doivent être traitées pour être interprétées telles que fournies par le capteur de vision omnidirectionnelle. Nos résultats montrent qu'une attention particulière doit être portée sur le choix des descripteurs lors de l'adaptation d'algorithmes développés pour la vision perspective à la vision omnidirectionnelle.

ABSTRACT. Omnidirectional vision sensors are mainly used for geometrical interpretation of scenes. However, few researchers have investigated how to perform object detection with such systems. The existing approaches require a geometrical transformation prior to the interpretation of the omnidirectional images. The face detection algorithm trained on perspective images is then applied on the unwrapped image. In this paper, we focus on how to process the omnidirectional images as provided by the sensor. While adapting algorithms developed for perspective images to omnidirectional images, our results suggest that the choice of descriptors is a critical step .

MOTS-CLÉS: boosting, détection de visages, vision omnidirectionnelle.

KEYWORDS: boosting, face detection, omnidirectional vision.

DOI:10.3166/TS.31.143-173 © 2014 Lavoisier

Extended Abstract

Face detection algorithms are now widely used in many real-time applications. This has been enabled by the breakthrough achieved by Viola and Jones in the early 2000's. Their method is based on a tradeoff between high face detection performance and highly computationally optimized operations. However, their work has been developed for perspective images.

Perspective cameras are interesting as they provide images close to the representation given by human eyes. However, perspective cameras have a limited field of view (FOV), which is a drawback when applications such as video surveillance are considered for instance. Omnidirectional vision systems provide an elegant alternative to perspective vision systems. Their large FOV is really attractive property, which explains that they are being widely used in mobile robotics. Omnidirectional vision systems are used to tackle challenges involving robot navigation, movement estimation or 3D reconstruction. In these applications, omnidirectional images can be processed directly as they involve point matching.

Most of the face detection algorithms, based on Viola and Jones framework, use region-based feature descriptors. As a result, the distortions induced by the omnidirectional image formation process do not allow the application of face detectors trained on perspective images. As a consequence, approaches found in the state-of-the-art use an intermediate image representation that aims at recreating the geometry found in perspective images.

In this paper, we propose to investigate how to detect faces on omnidirectional images without requiring an intermediate representation. Our contributions can be divided into two parts.

First of all, we highlight the consequences of omnidirectional image processing as compared to perspective image processing. We show that processing omnidirectional images directly has a lot of conceptual and practical benefits. Still, the distortions that exist on omnidirectional images introduce new challenges especially caused by an increased dispersion of the face class with respect to the feature space.

Secondly, we propose to train a face detector from synthesized omnidirectional-like image face patches. The face patches are obtained from faces detected on perspective images. The detector evaluation is performed on natural omnidirectional images. We investigated the influence of the feature descriptor used as well as the strong classifier. Our results suggest that a particular effort should be focused on the design of feature descriptors that take into account omnidirectional image distortions. Moreover, the performance achieved in our work indicates that our detector for omnidirectional

tional images may be trained from synthesized image patches taken from perspective images.



Results. Continuous lines indicate faces exclusively detected by our approach

1. Introduction

La détection de visages est devenue au fil du temps un sujet de recherche de plus en plus important avec de nombreuses applications en indexation d'images, interface homme-machine, robotique ou encore biométrie par exemple. Des avancées majeures ont été atteintes dans le cas des caméras perspectives où la maturité et les performances des algorithmes de traitement ont rendu possible leur transfert dans notre vie quotidienne.

Paul Viola et Micheal Jones, dans leurs travaux fondateurs, ont permis de réaliser ce qui était encore un rêve à leur époque (Viola, Jones, 2004). Un ordinateur ou une machine équipés d'une caméra sont maintenant capables de détecter un visage en temps réel. Viola et Jones n'étaient pas des pionniers en la matière (Schneiderman, Kanade, 2000 ; Rowley *et al.*, 1996). Cependant, leur démarche était motivée par un compromis entre des performances de détection et des considérations matérielles. Il en résulte une approche qui a atteint des performances sans précédent tout en utilisant des étapes de calculs fortement optimisées. Cette approche continue aujourd'hui d'inspirer de nombreux chercheurs dans leur quête d'un détecteur de visages optimal.

Cependant, ces travaux ont été développés pour des caméras perspectives. De par leur champ de vue limité, les caméras perspectives requièrent que les individus soient dans le champ de vue pour être détectés. Cette limitation est très contraignante pour de nombreuses applications : vidéoconférences, vidéo surveillance, etc. L'observation d'une scène sans direction privilégiée peut se faire en utilisant un réseau de caméras ; cette solution peut cependant se révéler contraignante dans sa mise en œuvre : coût du matériel et de son installation, problèmes de synchronisation, acquisition et traitement de plusieurs flux vidéo. Une alternative est l'utilisation de caméras omnidirectionnelles.

Les caméras omnidirectionnelles sont aujourd'hui principalement utilisées en robotique mobile. Elles permettent de répondre aux problématiques de navigation, d'estimation de mouvements (Bazin *et al.*, 2009 ; Mei *et al.*, 2011) ou bien de reconstruction 3D de scènes (Lhuillier, 2008). Les outils mathématiques utilisés reposent sur l'utilisation de la géométrie projective. Dans les applications citées précédemment, les images omnidirectionnelles sont traitées telles quelles. Ceci est possible puisque les méthodes impliquent la mise en correspondance des points et des petites régions à proximité du pixel considéré (Sturm *et al.*, 2011). De plus, des détecteurs d'amers usuels, comme les SIFT ou le détecteur de Harris, ont été modifiés avec succès pour prendre en compte la géométrie particulière des images omnidirectionnelles (Arican, Frossard, 2010 ; Demonceaux *et al.*, 2011 ; Lourenço *et al.*, 2010). Ces derniers résultats montrent que ces images peuvent être traitées directement.

Bien que la détection d'objets puisse impliquer la mise en correspondance de points d'intérêt, les méthodologies aujourd'hui privilégiées utilisent des descripteurs régionaux. La non-linéarité de la résolution des images omnidirectionnelles ne permet plus l'utilisation de détecteurs entraînés spécifiquement pour des images perspectives. L'approche classique est donc de re-crée une image pseudo-perspective en utilisant l'anti-anamorphose (Strauss, Comby, 2007) (c.f. figure 1). Une fois la représentation intermédiaire de l'image omnidirectionnelle obtenue, le détecteur entraîné pour des images perspectives est appliqué.

Dans cet article, nous proposons de montrer que la détection de visages est faisable sur une image omnidirectionnelle. Nous mettons en avant le fait que des résultats de détection satisfaisants peuvent être obtenus sans requérir une représentation intermédiaire de l'image omnidirectionnelle. L'élément clé est le choix de descripteurs adaptés à la nature de l'image omnidirectionnelle. Dans une première partie, nous nous attacherons à mettre en avant les conséquences du dépliement d'images omnidirectionnelles dans le contexte de la détection d'objets. Dans un second temps, nous évaluerons la robustesse de notre approche.

1.1. Travaux antérieurs

Le première tentative d'utilisation des capteurs de vision omnidirectionnelle pour la détection de visages a été proposée par Douxchamp et Campbell (Douxchamps, Campbell, 2007) en 2007. Dans ces travaux, les auteurs supposent que les visages sont tous alignés sur un cercle de rayon r sur le plan image omnidirectionnelle. La largeur l de l'image dépliée est ensuite fonction de ce rayon. Pour réduire le nombre de faux positifs, un filtre utilisant la teinte de peau est ensuite appliqué. Les principales limitations sont l'utilisation d'une connaissance a priori de la position des personnes ainsi que l'utilisation du filtre qui peut se révéler bruité et sensible aux conditions d'éclairage.

En 2009, Barczak *et al.* utilisent un capteur de vision omnidirectionnelle pour effectuer le suivi de visages (Barczak *et al.*, 2009). Les auteurs s'appuient sur un modèle géométrique de projection ad-hoc en supposant que les éléments du capteur sont par-

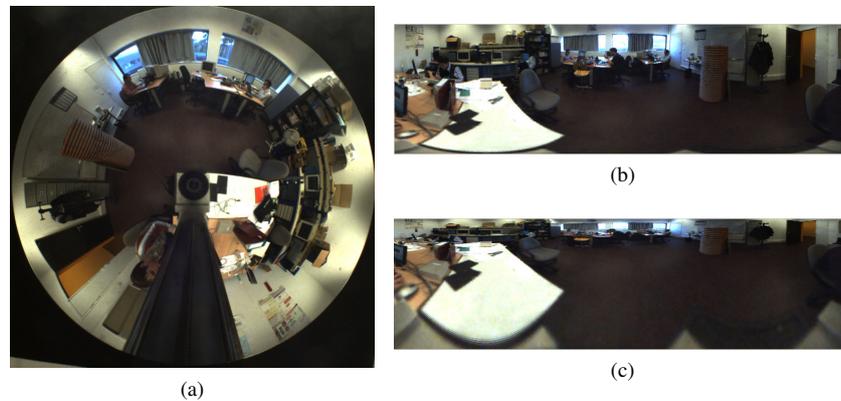


Figure 1. Exemples de dépliements (a) Image originale (b) Projection sphérique (c) Projection cylindrique

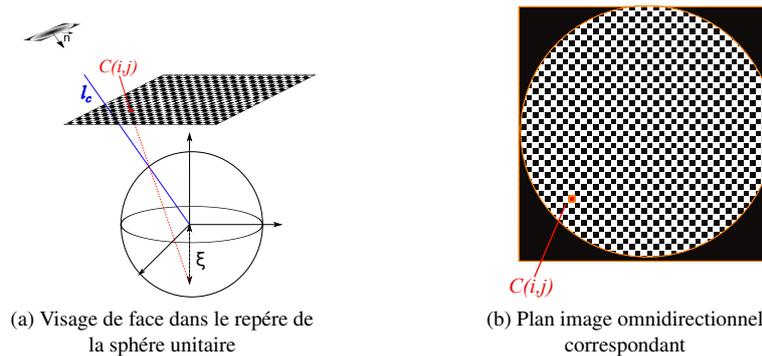


Figure 2. Visage de face dans le cadre des capteurs de vision omnidirectionnelle

faitement alignés et que la partie optique n'induit pas de distorsion. Pour atteindre un taux de détection convenable, les auteurs utilisent deux projections : un dépliement global de l'image ainsi qu'une projection perspective locale de chaque région d'intérêt. Il en résulte un détecteur qui ne fonctionne plus en temps réel. Ces travaux ont confirmé une faisabilité sans toutefois apporter une étude quantitative de la performance du détecteur de visages sur des images dépliées.

En 2010, nous proposons de répondre sur ce point (Dupuis *et al.*, 2010). Nous avons mesuré l'impact des deux approches de dépliement communément utilisées. De plus, nous avons évalué les performances spatiales et angulaires des méthodes de dépliement. Nous avons montré que le dépliement sphérique permet d'obtenir de meilleures performances.

Dans les travaux précédemment cités, le détecteur de visage de face de Viola et Jones a été utilisé (Bradski, Kaehler, 2008). Cependant, la notion de visage de face pour image omnidirectionnelle diffère du cas perspectif. En effet, un visage est dit de face lorsque le plan englobant le visage est quasi parallèle au plan image perspectif.

Afin d'illustrer ce point, nous définissons les variables suivantes (c.f. figure 2) :

- $C(i, j)$ centre de la région d'intérêt sur le plan image.
- l_C rayon 3D correspondant à la génératrice de $C(i, j)$.
- \mathbf{n} vecteur normal au plan visage.

Dans le contexte des images omnidirectionnelles, on définit alors un visage comme étant de face lorsque l_C et \mathbf{n} sont quasiment colinéaires. (c.f. figure 2). Ainsi toute rotation du visage sera définie par rapport à l_C .

Malgré les avancées récentes dans la détection de visages sur images omnidirectionnelles, nous pensons qu'il est erroné de considérer l'image dépliée comme une image perspective. L'objectif du dépliement est de retrouver des propriétés géométriques qui sont proches des images perspectives. Dans ce cas, un détecteur entraîné sur les images perspectives peut être directement appliqué. Cette approche pose cependant plusieurs problèmes. Premièrement, le dépliement implique une étape d'interpolation pour obtenir une image panoramique non éparse. Deuxièmement, une image omnidirectionnelle n'a théoriquement pas de limite. En pratique, ce type d'image ne possède qu'une limite. Dans le plan image omnidirectionnelle, si un objet est partiellement visible, cela signifie qu'il se situe dans la partie aveugle du capteur. Ainsi, une personne peut librement se déplacer autour du capteur sans toutefois rencontrer une limite liée au système de capture. Cela constitue l'aspect le plus attractif des capteurs de vision omnidirectionnelle. Lors du dépliement, trois bordures supplémentaires sont introduites. D'un point de vue omnidirectionnel, ces trois nouvelles limites sont conceptuellement incorrectes. De plus, ces nouvelles limites complexifient la recherche de visages. Ainsi, un objet entièrement observé par le capteur omnidirectionnel, et se trouvant sur une des limites définies pour le dépliement, se retrouvera découpé en plusieurs parties complémentaires présentes à différentes extrémités de l'image. La situation devient même plus complexe lorsque le visage se retrouve au centre de l'image omnidirectionnelle (c.f. figure 4). La solution proposée est alors de répéter les zones proches des bordures de l'image, approche qui a pour conséquence, notamment, d'augmenter de façon significative le temps de calcul nécessaire au traitement complet d'images omnidirectionnelles dépliées. De fait, nous pensons que les images omnidirectionnelles doivent être traitées directement. C'est ce que nous attachons à démontrer dans le reste de cet article.

Pour aider le lecteur, qu'il soit de la communauté de la vision omnidirectionnelle ou de la détection d'objets, à mieux appréhender nos choix, nous proposons, dans un premier temps, une présentation brève de la théorie de la vision omnidirectionnelle (c.f. section 2). Dans la section 3, nous présentons une analyse quantitative du bruit induit par le processus de formation de l'image omnidirectionnelle. La section 4 se focalise sur notre méthodologie. Enfin les sections 5 et 6 présentent respectivement nos résultats et concluent notre discussion.

2. Théorie de la vision omnidirectionnelle

2.1. Introduction

Les caméras perspectives sont aujourd'hui considérées comme conventionnelles. Cependant, leur champ de vue est particulièrement limité, *i.e.* 30 à 60 degrés. A titre comparatif, le champ de vue de l'œil humain est de 170 degrés de large pour 135 degrés de haut. Les caméras peuvent être classifiées en fonction de leur *sphère de vue*. La sphère de vue est utilisée pour représenter le champ de vue de la caméra. Le centre de la sphère correspond au centre optique de l'appareil de capture. Il existe quatre catégories. Une caméra est dite *directionnelle* si son champ de vue couvre une partie limitée de la sphère. Les caméras perspectives sont donc des caméras directionnelles. Lorsque le champ de vue couvre l'ensemble de la sphère, on parle de caméra *omnidirectionnelle*. Les deux types de caméras restant correspondent aux caméras *panoramiques* et aux caméras *large champ*. Un capteur panoramique a un champ de vue d'au moins 360 degrés sur un des axes de la sphère. Un système optique large champ a un champ de vue relativement plus grand que les caméras directionnelles. En pratique, il n'existe pas réellement de caméras omnidirectionnelles. Ainsi, les capteurs de vision panoramique sont communément appelés capteurs de vision omnidirectionnelle.

2.2. Systèmes de vision panoramique

Les capteurs de vision panoramique peuvent être regroupés en trois catégories (Bunschoten, 2003) :

- Caméras rotatives

Une caméra perspective est placée sur un système rotatif. Il existe deux types d'approches : les caméras cylindriques qui tournent autour d'un degré de liberté et les caméras sphériques qui autorisent deux degrés de liberté. Les vues prises à des positions distribuées de manière discrète sont ensuite fusionnées. Les meilleurs résultats sont obtenus lorsque l'axe de révolution coïncide avec le centre optique de la caméra. L'inconvénient de ce type d'approche est le temps d'acquisition de la scène ainsi que le fait que celle-ci doit être statique.

- Matériel dédié

Le problème évoqué précédemment peut être pallié en utilisant un réseau de caméras. Les champs de vue des caméras sont distribués de manière à couvrir la sphère de vue. Les acquisitions peuvent être synchrones ou asynchrones. Le système le plus connu est la Ladybug de la société Pointgrey. Un autre approche consiste à modifier l'optique de la caméra en remplaçant les lentilles conventionnelles par des lentilles *fisheye*.

- Capteurs catadioptriques

Un miroir, bien souvent convexe, est combiné à une caméra perspective pour en augmenter le champ de vue. Les caméras perspectives sont modélisées par le biais du modèle pinhole. Ainsi, tous les rayons lumineux impactant le plan image passent par un point unique : le centre de projection. Une famille limitée de miroir, définie dans les travaux de (Baker, Nayar, 1999) répondent à cette contrainte, aussi connue sous le

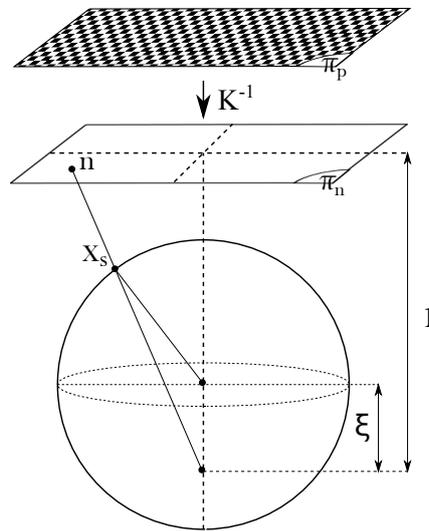


Figure 3. Modèle de la sphère unitaire

terme de *contrainte du point de vue unique*. Cette contrainte permet l'utilisation de la géométrie épipolaire.

Grâce au respect de la contrainte du point de vue unique, les capteurs catadioptriques sont largement utilisés dans la communauté de la vision omnidirectionnelle. Une vue panoramique de la scène est obtenue en une unique prise de vue. Il n'y a pas de partie mobile ni de synchronisation nécessaire. Les capteurs catadioptriques présentent donc des avantages qui les distinguent des autres approches.

2.3. Modélisation mathématique

Comme tout système de mesure, les capteurs de vision omnidirectionnelle doivent être calibrés. La calibration permet d'assurer la véracité de la projection de la scène sur le plan image. La calibration peut être géométrique ou radiométrique. Cette dernière est bien souvent ignorée par la communauté de la vision omnidirectionnelle. Cependant, le lecteur intéressé par ce propos peut se reporter aux travaux de (Lin *et al.*, 2004). La calibration géométrique consiste à exprimer la relation qui existe entre un point 3D et un point 2D sur le plan image. Les méthodes de calibration peuvent être regroupées en deux familles : paramétrique et non paramétrique. Les approches paramétriques impliquent l'utilisation d'un modèle dont on estime les paramètres. Les méthodes non paramétriques réduisent le problème à l'estimation de l'origine et de la direction du rayon lumineux engendrant chaque pixel (Sturm *et al.*, 2011). Le principal inconvénient des méthodes non paramétriques est l'absence de modèle. De plus,

une méthodologie complexe est bien souvent nécessaire pour obtenir une précision convenable. Par conséquent, les modèles paramétriques sont les plus utilisés par la communauté. Il existe trois types de modèles :

– Ad-hoc

Le processus physique de projection ainsi que sa géométrie constitue la base du modèle. Les paramètres optiques ainsi que les paramètres de la caméra sont utilisés pour modéliser la formation de l'image. Ce modèle fut introduit par (Svoboda, 1999). Les modèles ad-hoc impliquent l'estimation de nombreux paramètres. Ces problèmes de minimisation sont souvent enclins aux minima locaux. Enfin, comme son nom l'indique, le modèle est spécifique à chaque configuration.

– Approximation polynomiale

Pour contourner les défauts du modèle ad-hoc, les modèles polynomiaux ont été utilisés (Scaramuzza *et al.*, 2006). L'objectif est de caractériser la fonction de projection sous une forme polynomiale.

– Modèle unifié

Un modèle généralisable à tous les capteurs centraux a été proposé par (Geyer, 2003 ; Barreto, 2003). Un capteur est dit *central* s'il respecte la contrainte du point de vue unique. Le modèle de projection est basé sur la sphère unitaire. Les points 3D sont dans un premier temps projetés sur la sphère puis sur le plan image. Bien que simple, ce modèle comprenait toujours de nombreux paramètres à estimer. Mei et Rives (Mei *et al.*, 2011) ont simplifié le modèle de Barreto et Geyer en considérant que l'erreur due à l'architecture du capteur était négligeable. Ce modèle peut être utilisé pour les capteurs catadioptriques aussi bien que les optiques *fisheyes*.

Les nombreux avantages associés au modèle unifié expliquent que nous l'avons choisi pour la suite de nos travaux. De par le modèle, nous sommes capables d'exprimer les coordonnées d'un pixel p dans un repère sphérique 3D et vice-versa. Le modèle est basé sur cinq paramètres¹ : u_0 et v_0 , les coordonnées du point principal²; γ_u et γ_v , les focales généralisées, respectivement horizontale et verticale; ξ , le paramètre lié au miroir (cf. figure 3). u_0 , v_0 , γ_u et γ_v sont utilisés dans la matrice de projection généralisée :

$$K = \begin{bmatrix} \gamma_u & 0 & u_0 \\ 0 & \gamma_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Ces paramètres peuvent être utilisés soit pour la projection d'un point 3D sur le plan image soit l'opposé. L'opération de *lifting* permet de projeter un point du plan image sur la sphère unitaire. Considérons le point $p = [i \ j \ 1]^T$ sur le plan π_p . Sa projection $n = [x \ y \ 1]^T$ sur le plan π_n est donnée par la relation suivante :

$$n = K^{-1}p \quad (2)$$

1. Si les distorsions et la non-orthogonalité des pixels sont considérées comme nulles.

2. Position sur le plan image où l'axe optique de la caméra intersecte le plan image.

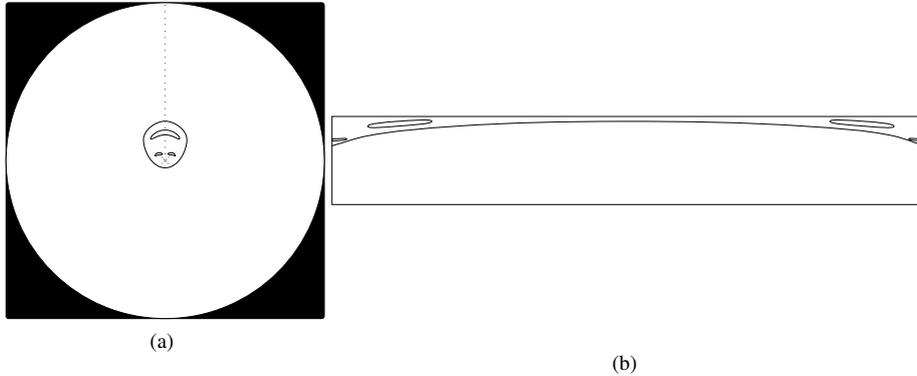


Figure 4. Illustration de la segmentation de visages. (a) Dessin omnidirectionnel
(b) Image dépliée résultante

Enfin, ses coordonnées dans le repère sphérique est obtenues en appliquant l'opérateur de lifting sur n :

$$X_s = h^{-1}(n) = \begin{bmatrix} \frac{\xi + \sqrt{1 + (1 - \xi^2)(x^2 + y^2)}}{x^2 + y^2 + 1} x \\ \frac{\xi + \sqrt{1 + (1 - \xi^2)(x^2 + y^2)}}{x^2 + y^2 + 1} y \\ \frac{\xi + \sqrt{1 + (1 - \xi^2)(x^2 + y^2)}}{x^2 + y^2 + 1} - \xi \end{bmatrix} \quad (3)$$

2.4. Déplie ment d'images omnidirectionnelles et ses conséquences sur la détection de visage

Premièrement, comme présenté dans la figure 1, deux types de projections, analogues aux caméras rotatives, existent. Une image omnidirectionnelle est à l'origine circulaire. Elle ne comporte qu'une seule limite : la limite physique du capteur. Un objet peut être déplacé dans l'image sans rencontrer de limite liée au processus de formation de l'image. Cette propriété intrinsèque constitue le point le plus attrayant des capteurs omnidirectionnels. Le déplie ment, il est vrai, permet de reconstituer une image géométriquement proche de la représentation perspective. Cependant, elle demande l'ajout de trois bordures supplémentaires, la limite naturelle de l'image étant conservée. D'un point de vue omnidirectionnel, l'ajout de ces trois limites est conceptuellement erroné. De plus, comme illustré sur la figure 4, la recherche de région d'intérêt s'avère plus complexe. En effet, celle-ci peut se retrouver coupée, voire totalement distordue lorsqu'elle se trouve au centre de l'image.

Deuxièmement, même si l'anti-anamorphose vise à retrouver des propriétés géométriques proches des images perspectives, cela s'avère impossible. En effet, comme

nous pouvons le constater sur la figure 1, l'image devient de plus en plus floue et distordue lorsque l'on s'approche du bas. Ce phénomène est expliqué par l'utilisation de l'interpolation et la résolution non linéaire le long de l'azimut des images omnidirectionnelles. Par conséquent, seule une sous-partie de l'image dépliée est exploitable. Le dépliement requiert des transformations géométriques et colorimétriques. Elles représentent des étapes supplémentaires nécessaires en amont de tout traitement supplémentaire.

Enfin, la détection de visage implique une recherche exhaustive sur l'image considérée. Une fenêtre glissante permet d'évaluer la plupart des sous-régions constituant l'image à tester. Le nombre de fenêtres \mathcal{N} évaluées par le détecteur de visages pour une image $M \times N$ est donnée par :

$$\mathcal{N} = \sum_{20 \leq i \leq n, i \in \mathcal{I}} (M - i) \times (N - i) \quad (4)$$

pour une sous-région de taille minimale 20 x 20 pixels, \mathcal{I} l'ensemble des tailles évaluées et $n = \min(M, N)$. Appliquons maintenant cette formule à deux cas. Dans un premier cas, nous traitons l'image omnidirectionnelle directement. L'image omnidirectionnelle correspond à un disque sur le plan image. En conséquence, M est égal à N. L'équation 4 devient donc :

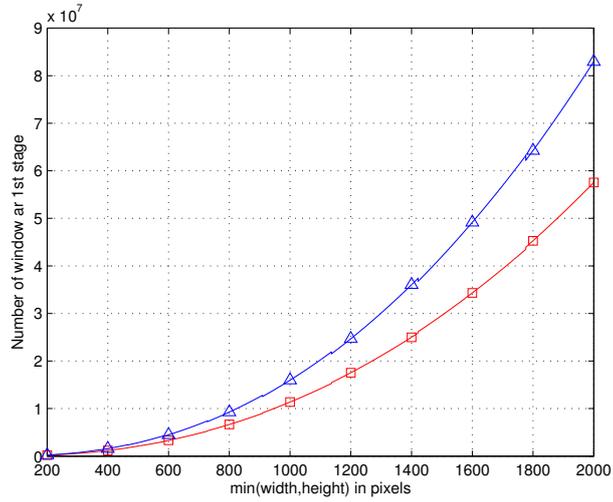
$$\mathcal{N}_{square\ image} = \sum_{20 \leq i \leq \frac{n}{2}, i \in \mathcal{I}} (n^2 - 2ni + i^2) \quad (5)$$

Dans le second cas, nous déplaçons l'image en utilisant le dépliement sphérique. Il en résulte une image rectangulaire de hauteur $\frac{n}{2}$ et largeur πn . L'équation 4 devient alors :

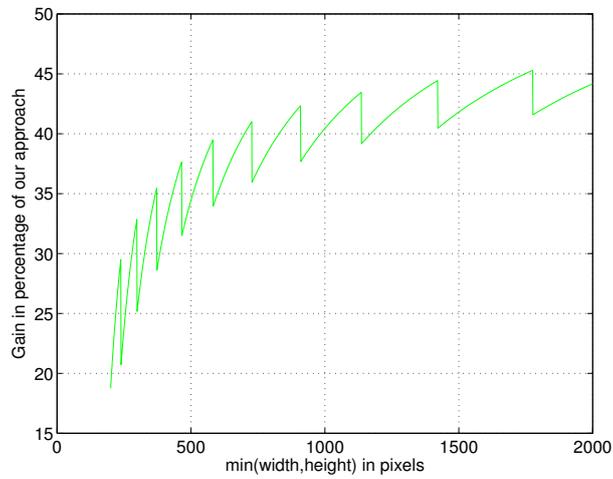
$$\mathcal{N}_{panoramic\ image} = \sum_{20 \leq i \leq \frac{n}{2}, i \in \mathcal{I}} \left(\frac{\pi}{2} n^2 - \frac{2\pi + 1}{2} ni + i^2 \right) \quad (6)$$

En pratique, la taille des ROI testées évolue selon une suite géométrique. La raison de cette suite est un des paramètres du détecteur d'objets et impacte principalement la vitesse de détection. La figure 5 donne un aperçu de l'évolution des équations 5 et 6 lorsque n varie de 200 à 2 000. L'équation 5 est théorique si un seul détecteur est considéré. Cependant en adoptant l'approche utilisée par (Viola, Jones, 2003), plusieurs détecteurs entraînés pour une pose spécifique peuvent être fusionnés pour permettre une détection robuste à la rotation dans le plan. En conséquence, l'équation 5 est totalement applicable. Comme indiqué, la réduction du nombre de sous-régions à évaluer atteint plus de 40 % pour une image 1 024 x 1 024 pixels et une suite géométrique de raison 1,25.

Dans cette première section, nous avons montré que considérer une image omnidirectionnelle dépliée comme perspective est, à la fois, erroné mais aussi source de



(a)



(b)

Figure 5. (a) Nombre de fenêtres évaluées au premier étage sans dépliement (rouge/carrés) et avec dépliement (bleu/triangles) (b) Gain en % d'une approche directe comparée aux méthodes existantes

calculs supplémentaires. De plus, l'ajout de bordures supplémentaires imposées par la désanamorphose fait apparaître des problèmes sous-jacents. Cependant, le traitement direct d'images omnidirectionnelles apporte lui aussi de nouvelles difficultés. Nous allons donc dédier la prochaine section à leur description et quantification.

3. Problématiques liées à la détection de visages sur images omnidirectionnelles

3.1. Base d'entraînement

Les approches modernes pour apprendre aux machines à détecter les visages impliquent l'utilisation de milliers, voire centaines de milliers d'échantillons. Alors que des échantillons d'arrière-plans sont faciles à obtenir, il apparaît souvent difficile de collecter une base d'échantillons de qualité pour la classe d'objets à détecter. Pour les images perspectives, les chercheurs peuvent cependant utiliser les bases de données des visages existantes et même les fusionner pour obtenir une base encore plus grande. Ils peuvent aussi naviguer sur internet pour trouver une multitude d'images comprenant des visages. Il est ainsi possible d'obtenir une base comprenant une diversité importante de teintes de peaux, de conditions d'illuminations ce qui aura pour conséquence d'obtenir un détecteur plus robuste aux conditions réelles (Jain, Learned-Miller, 2010).

Les amers des visages ont tendance à être stables parmi une population d'images perspectives par rapport aux images omnidirectionnelles qui souffrent de fortes distorsions. Les distorsions géométriques trouvées sur les images omnidirectionnelles ont trois origines. Tout d'abord, le déplacement d'un objet le long d'un rayon allant du centre de l'image à sa bordure produira des distorsions variées. En effet, la résolution des capteurs de vision omnidirectionnelle est bien souvent non linéaire dans cette dimension. Ensuite, une rotation dans le plan image apparaît. Elle est causée par le processus de formation des images omnidirectionnelles. Ainsi, si nous voulons être en mesure de traiter complètement une image omnidirectionnelle, cette rotation dans le plan devra être prise en compte. Enfin, les distorsions causées par la pose du visage lui-même causera des distorsions bien plus importantes. En effet, l'apparence d'un visage orienté perpendiculairement à l'axe optique du capteur de vision est totalement différente de celle obtenue si ce même visage était placé de manière fronto-parallèle à l'axe optique.

Tableau 1. Paramètres du capteur de vision omnidirectionnelle

| Paramètre | Valeur |
|------------|---------|
| ξ | 1.00609 |
| γ_u | 391 |
| γ_v | 389 |
| u_0 | 512 |
| v_0 | 512 |

Dans ces travaux, nous nous intéressons uniquement aux distorsions causées par la non-linéarité de la résolution le long d'un rayon. Nous devons par conséquent regrouper un nombre significatif de visages englobant une dispersion colorimétrique et géométrique importante. Comme nous n'étions pas en mesure de collecter une base aussi large, nous avons synthétisé ces visages grâce à la variante du modèle unifié proposée par (Mei, Rives, 2007) lui-même basé sur l'approche de Barreto et Geyer (Geyer, Daniilidis, 2001 ; Barreto, 2006). Après calibration de notre capteur de vision omnidirectionnelle, nous obtenons les paramètres présentés dans le tableau 1.

Le modèle étant paramétrique, il est possible de retrouver la direction du rayon associé à chaque pixel. Un visage issu d'une image perspective peut être vu comme un plan 3D mis à l'échelle. Comme indiqué précédemment, nous nous intéressons uniquement à la détection frontale de visages sur images omnidirectionnelles. Par conséquent, nous considérons le point principal X_{pp} de notre imagerie de visage comme un paramètre de notre fonction de projection. Sa projection sur le plan image x_{pp} correspond au centre de notre région d'intérêt. Nous approximons la taille réelle du visage avec un rectangle de dimension 20 x 20 cm. Pour éviter tout biais géométrique dans notre base d'entraînement, nous considérons toutes les transformations géométriques possibles qui amèneraient le plan visage 3D Π_f à voir la projection de son point principal x_{pp} le long du rayon r . Nous définissons π_f la région d'intérêt correspondante et \mathbb{F} l'ensemble des π_f pour un rayon donné. Pour une image 1024 x 1024, \mathbb{F} comprend environ 100 000 éléments. Pour chaque π_f , grâce au modèle de la sphère unitaire, nous sommes capables de définir la fonction bijective suivante :

$$\Pi_f \xrightarrow{\sim} \pi_f \quad (7)$$

Par conséquent, nous sommes en mesure de synthétiser tous les π_f à partir d'une image perspective Π_f . Nous pouvons donc obtenir les dispersions colorimétriques et géométriques requises pour notre base d'entraînement de visages. Notre approche peut être utilisée pour tout apprentissage demandant une base importante. Nous souhaitons maintenant quantifier les difficultés supplémentaires induites par les capteurs omnidirectionnels.

3.2. Influence des distorsions sur le détecteur

Pour quantifier les distorsions géométriques en fonction de la taille et de la position de la région d'intérêt π_f , nous utilisons le pictogramme de visage présenté dans la figure 6a. Nous mesurons la distance de Jaccard entre la figure 6a et les imagerie projetées. La distance de Jaccard J_δ est définie comme suit :

$$J_\delta = \frac{M_{01} + M_{10}}{M_{01} + M_{10} + M_{11}} \quad (8)$$

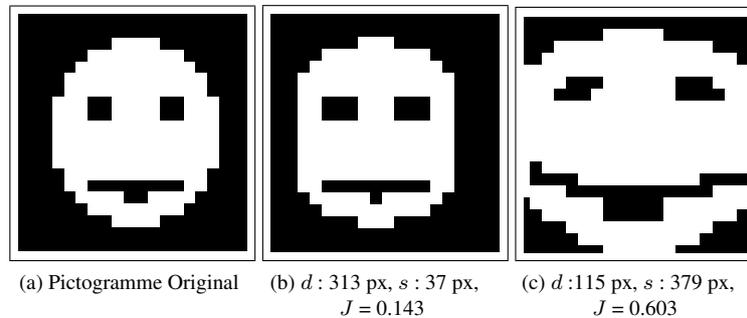


Figure 6. Illustration de la mesure de distance de Jaccard; d : distance au point principal, s : taille de la région d'intérêt, J : distance de Jaccard

où:
 M_{ij} nombre de pixels appartenant à la classe i dans l'image de référence alors qu'il appartient à la classe j dans l'image de test.

Dans notre pictogramme, chaque pixel appartient soit à la classe 1 (en blanc) ou classe 0 (en noir). Dans l'équation 8, le numérateur mesure le nombre de pixel dont la classe change entre les deux images. Le dénominateur prend aussi en compte les pixels restant dans la région blanche dans les deux images.

Les attributs utilisés dans la détection de visage permettent de détecter des régions de forts contrastes. Ils sont paramétrés par leur taille et position. Par conséquent lors de l'entraînement du détecteur, il faut s'assurer de l'alignement de ces régions. L'utilisation du pictogramme binaire et de la distance de Jaccard nous permet d'évaluer les modifications qu'elles subissent en forme, position et taille. Ces variations peuvent être interprétées comme la dispersion du visage d'un individu sur l'ensemble \mathbb{F} . Les figures 6b et 6c illustrent deux exemples de distorsions ainsi que la distance de Jaccard associée. Nous avons mesuré la distance de Jaccard pour toutes les régions $\pi_f \in \mathbb{F}$. La figure 7 présente la dispersion de la distance de Jaccard. Les deux sous-figures représentent les histogrammes normalisés, exprimés en pourcentage, selon l'axe des y . Comme nous pouvons le constater les variations sont conséquentes entre l'image de référence et leur projection. Les variations sont aussi bien engendrées par la position de la région d'intérêt que par la taille de celle-ci. Ainsi lorsque π_f devient grand, J_δ peut atteindre une valeur de 0,6 (c.f. figure 7a). Cela correspond à un changement de classe de 55 % des pixels. Une telle variation est en soi très importante. Comme on peut le constater sur la figure 7b, la dispersion est aussi importante pour une position donnée le long du rayon. La figure 8 s'intéresse à la distribution de la distance de Jaccard. Elle met en exergue que la distance de Jaccard n'est pas distribuée de manière homogène. Environ la moitié des régions d'intérêt a une distance de Jaccard inférieure à 0,3. Par conséquent, un détecteur unique devrait être en mesure de traiter les distorsions géométriques.

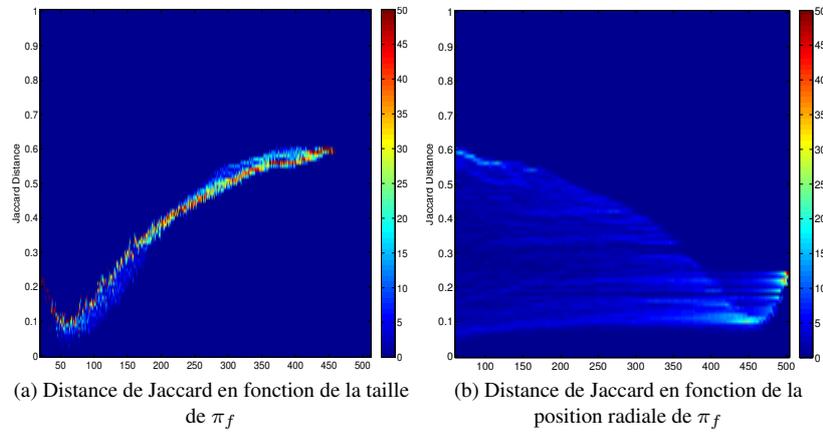


Figure 7. Distance de Jaccard- Histogrammes normalisés en %
 Normalisation appliquée dans la dimension y. La légende indique la fréquence relative

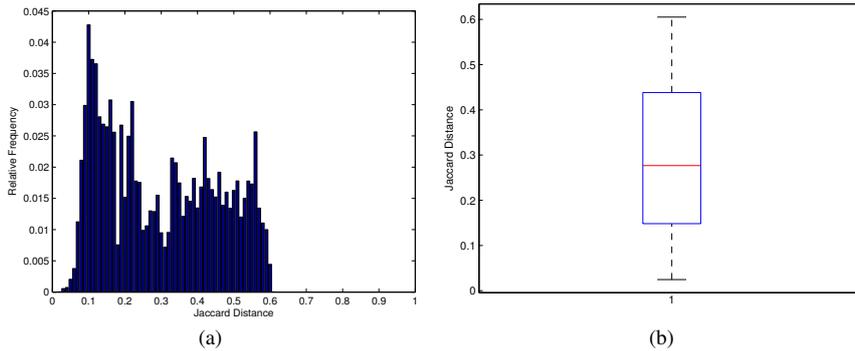


Figure 8. Distance de Jaccard (a) Histogramme normalisé (b) Boîte à moustaches

La distance de Jaccard nous donne une information qualitative à l'échelle d'une région dans l'image et non du pixel. Combinées aux distorsions géométriques, la dispersion colorimétrique de chaque pixel aura un impact sur la performance du classifieur final. En effet, en apprentissage statistique, les meilleures performances de classification sont atteintes lorsque la distance intra classe est faible alors que la distance inter classe est plus conséquente. Ainsi, on souhaite que la dispersion, de l'intensité d'un pixel $I_{x,y}$, à la position $[x, y]^T$, soit la plus faible possible à travers la population d'échantillons. Dans un contexte probabiliste, cela revient à avoir une fonction de masse avec une variance réduite. En d'autres termes, l'incertitude est faible. L'entropie est une mesure communément utilisée pour la mesure d'incertitude. Nous proposons d'utiliser cette mesure pour mesurer l'incertitude de l'intensité de chaque pixel pour

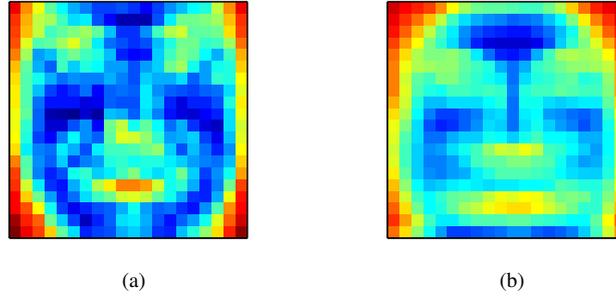


Figure 9. Entropie pixellique (a) Images perspectives (b) Images omnidirectionnelles

des régions d'intérêts projetées sur le plan image omnidirectionnelle ainsi que des visages correspondants dans le plan image perspectif (c.f figure 9). Nous avons utilisé 5 000 visages issus de la base LFW (Jain, Learned-Miller, 2010) que nous avons projetés aléatoirement le long d'un rayon sur le plan image omnidirectionnelle.

(Viola, Jones, 2004) proposent de prétraiter les imagettes en normalisant leur intensité par la variance comme suit :

$$I_{I_p} = \frac{I_p}{\sigma_w} \quad (9)$$

où:

I_{I_p} intensité du pixel à la position p après normalisation
 I_p intensité originale du pixel à la position p
 σ_w écart type des intensités dans l'imagette w

Pour augmenter la robustesse aux variations d'illumination, nous avons centré et réduit la distribution des intensités des pixels. Cette opération est appliquée avant le calcul d'entropie pour mesurer la variabilité à laquelle devra faire face le classifieur. Inspirés de la règle des trois sigma, nous avons donc divisé notre fonction de masse en huit intervalles discrets B_j :

$$\forall I_{(x,y)} \in B_j, \tau_j \leq I_{(x,y)} < \tau_{j+1} \quad (10)$$

où $\tau_j \in \{-\infty, -3\sigma, -2\sigma, -\sigma, 0, \sigma, 2\sigma, 3\sigma, \infty\}$

L'entropie pour un pixel $H_{(x,y)}$ est donnée par :

$$H_{(x,y)} = \sum_{j=1}^8 -p_{B_j} \times \log p_{B_j} \quad (11)$$

où p_{B_j} est la probabilité que $I_{(x,y)} \in B_j$

Nous avons appliqué l'équation 11 sur les deux jeux de données. Nous obtenons ainsi les résultats présentés sur la figure 9. Les couleurs froides correspondent à une entropie faible alors que les couleurs chaudes sont associées à une entropie forte. Comme on peut le constater sur la figure 9a, la forme du visage reste elliptique. De plus on distingue facilement les yeux et la bouche. Sur la figure 9b, c'est le contraire. En effet, le visage a maintenant une forme plutôt ronde et on distingue plus difficilement les régions des yeux et de la bouche qui sont plus floues. L'entropie moyenne est 13 % supérieure dans les imagerie omnidirectionnelles. Cette entropie va devoir être ensuite traitée par l'algorithme d'apprentissage. Elle aura pour conséquence de requérir une structure de classifieur plus complexe que dans le cas des images perspectives. La variabilité en termes de forme est elle liée aux observations du paragraphe précédent.

Dans cette section, nous avons montré qu'il va être plus difficile d'apprendre à une machine à détecter un visage lorsque l'on considère uniquement les distorsions dues à la résolution radiale non linéaire. La difficulté devient encore plus importante lorsque la rotation dans le plan sera introduite. Nous nous intéressons maintenant à la présentation de notre méthode pour contrecarrer la non linéarité de la résolution radiale. Nous proposons aussi d'en évaluer les limites.

4. Méthodologie

4.1. Méthode d'apprentissage

La méthode d'apprentissage la plus utilisée pour la détection de visages est l'Ada-boost. Elle est fondée sur la notion de *boosting*. L'objectif du *boosting* est de combiner plusieurs classifieurs faibles h_k pour obtenir un comité d'experts forts, soit un classifieur fort H . La décision finale est une somme pondérée des votes de tous les membres du comité :

$$H(X) = \sum_{k=1}^n \alpha_k h_k(x) \quad (12)$$

L'Ada-boost implique deux pondérations. Les classifieurs faibles, comme indiqué dans l'équation 12, sont pondérés tout comme les exemples de la base d'entraînement. Les deux pondérations sont intrinsèquement liées puisque le poids de chaque classifieur faible est proportionnel à son erreur sur la base d'entraînement étant donnée la

Algorithm 1 Adaboost discret

- 1: Considérons: $(x_1, y_1), \dots, (x_m, y_m); x_i \in \mathcal{X}, y_i \in \{-1, +1\}$
 - 2: Initialisation des poids: $D_1(i) = \frac{1}{m}$
 - 3: **pour** $t = 1, \dots, T$ **faire**
 - 4: $h_t = \arg \min_{h_j \in \mathcal{H}} \epsilon = \sum_{i=1}^m D_t(i) \mathcal{I}(y_i \neq h_j(x_i))$ {où \mathcal{I} est la fonction caractéristique}
 - 5: **si** $\epsilon_t \geq \frac{1}{2}$ **alors**
 - 6: **BREAK**
 - 7: **sinon**
 - 8: $\alpha_t = \frac{1}{2} \log \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$
 - 9: $D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$ {où Z_t est le facteur de normalisation}
 - 10: **fin si**
 - 11: **fin pour**
 - 12: $H(x) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(x))$
-

Algorithm 2 Adaboost réel

- 1: Considérons: $(x_1, y_1), \dots, (x_m, y_m); x_i \in \mathcal{X}, y_i \in \{-1, +1\}$
 - 2: Initialisation des poids: $D_1(i) = \frac{1}{m}$
 - 3: **pour** $t = 1, \dots, T$ **faire**
 - 4: $h_t = \arg \min_{h_j \in \mathcal{H}} \epsilon = \sum_{i=1}^m D_t(i) \mathcal{I}(y_i \neq h_j(x_i))$ {où \mathcal{I} est la fonction caractéristique}
 - 5: **si** $\epsilon_t \geq \frac{1}{2}$ **alors**
 - 6: **BREAK**
 - 7: **sinon**
 - 8: $W_b^j = \sum_{i: x \in X_j, y_i = b} D(i)$
 - 9: $c_j = \frac{1}{2} \ln \left(\frac{W_+^j}{W_-^j} \right)$
 - 10: $D_{t+1}(i) = \frac{D_t(i) \exp(-y_i c_{j,t}(x_i))}{Z_t}$ {où Z_t est le facteur de normalisation}
 - 11: **fin si**
 - 12: **fin pour**
 - 13: $H(x) = \text{sign}(\sum_{t=1}^T c_{j,t}(x))$
-

distribution des poids des observations. Plusieurs variantes d'Adaboost existent. Dans ces travaux, nous nous intéressons à l'Adaboost réel (Schapire, Singer, 1999). Généralisation de l'Adaboost discret (Freund, Schapire., 1997), l'Adaboost réel a montré qu'il surpassait les autres variantes d'Adaboost dans la détection de visages. Comme montré dans (Schapire, Singer, 1999), l'Adaboost réel surclasse l'Adaboost discret puisque qu'il requiert moins d'itérations pour minimiser les erreurs d'apprentissage et de test. L'algorithme de l'Adaboost discret est présenté dans l'algorithme 1.

Dans notre cas, $h_t(x_i)$ a deux valeurs possibles $\{-1, +1\}$. Ainsi, si on pose $u_i = y_i h_t(x_i)$, u_i est lui aussi inclus dans l'ensemble $\{-1, +1\}$. L'algorithme d'Adaboost réel utilise une approche à base de partition d'ensembles (c.f. algorithme 2). Considérons que nous utilisons un arbre de décision à un nœud et deux branches comme classifieur faible et l'Adaboost réel comme classifieur fort. L'arbre de décision va déterminer le meilleur attribut \mathcal{X}^t et partitionner l'espace résultant en deux sous-ensembles disjoints: X_1 et X_2 . Nous définissons $c_j = h(x)$ si $x \in X_j$ et W_b^j la fraction de poids des observations appartenant au sous-ensemble j et de classe b . W_b^j est défini comme suit :

$$W_b^j = \sum_{i: x \in X_j, y_i = b} D(i); \quad (13)$$

où $D(i)$ est la pondération de l'observation i .

Ainsi, c_j devient :

$$c_j = \frac{1}{2} \ln \left(\frac{W_+^j}{W_-^j} \right) \quad (14)$$

La classe correspondant à la majorité pondérée dans le sous-ensemble j donne le signe de c_j . Dans le cas où la somme pondérée des deux classes, dans le sous-ensemble j , est égale, c_j devient zéro. c_j représente donc le taux de confiance de notre classifieur faible.

4.2. Descripteurs de Haar et images omnidirectionnelles

Parmi les descripteurs utilisés dans la détection de visages, les attributs de Haar sont les plus couramment usités. Grâce à l'image intégrale (Viola, Jones, 2004), ils demandent peu d'opérations pour être calculés. L'idée de l'image intégrale est de décomposer une intégrale polygonale en une somme pondérée d'intégrales polygonales élémentaires calculées en amont. A l'origine, ces polygones étaient limités à deux pentes, soit des rectangles. Depuis, de nombreuses variantes ont été introduites (Lienhart, Maydt, 2002 ; Barczak *et al.*, 2005 ; Du *et al.*, 2006). Dans ces travaux, nous nous sommes intéressés à la projection de régions rectangulaires sur le plan image omnidirectionnelle. Des travaux similaires ont été menés par Strauss *et al.* pour adapter les noyaux de filtrages aux images omnidirectionnelles (Strauss, Comby, 2007). Ainsi, un noyau de filtrage rectangulaire devient un tétragone dont deux de ses côtés sont des cercles concentriques. Les deux côtés restants correspondent à des segments de rayon. L'utilisation d'arc de cercle ne permet plus de tirer partie de l'image intégrale. En effet, la variation de la tangente est continue. Pour ce faire, nous nous sommes donc intéressés à la discrétisation de ces tétragones (c.f. tableau 2). Cette discrétisation nous permet d'obtenir un nombre fini et faible de pentes, ici trois. Récemment, Pham *et al.* ont proposé une méthode d'intégration compatible avec tout type de polygones (Pham *et al.*, 2010). Le nombre d'opérations dépend uniquement du nombre de sommets du polygone. Ces travaux ont aussi montré que les attributs polygonaux surpassent les



Figure 10. Exemples d'images d'arrière plans utilisées pour l'apprentissage

attributs rectangulaires sur les images perspectives. Les attributs polygonaux utilisés dans ces travaux sont très proches des attributs rectangulaires traditionnellement utilisés et projetés sur le plan image omnidirectionnelle. Nous avons donc pris le parti de réutiliser ces attributs dans nos travaux (c.f. figure 14).

Tableau 2. Transformation des attributs de Haar

| Formes | Rectangle | Projection sur l'image omnidirectionnelle | |
|--------------------------------|---|---|---|
| | | | Polygone |
| |  → |  | →  |
| Image intégrale | $I(i,j)$ → | $I(\theta,\varphi)$ | → $I(i,j)$ |
| Nature de l'image intégrale | Discrète → | Continue | → Discrète |

5. Expérimentations

5.1. Protocole expérimental

La première étape est la construction d'une base d'apprentissage. Pour ce faire, nous avons synthétisé 2 500 visages à partir de visages issus de la base Label Face into the Wild. Pour la base d'arrière-plans, nous avons pris des images naturelles issues de capteurs catadioptriques (c.f. figure 10). Plusieurs millions de sous-régions ont été extraites de ces images pour l'entraînement. Notre base de test comprend 45 images catadioptriques comprenant chacune un visage. La base de test a été construite pour prendre en compte un maximum de variations géométriques et colorimétriques. Ainsi, les sujets pouvaient être assis ou debout et ceci à plusieurs distances du capteur (c.f. figure 13). Enfin nous avons utilisé des images supplémentaires pour évaluer la robustesse de notre détecteur vis à vis de scène en intérieur, extérieur ou du genre de la personne (c.f. figures 15 et 16).

L'apprentissage du classifieur fort a été effectué en utilisant une approche symétrique. En conséquence, nous avons privilégié une approche à base de cascade attentionnelle plutôt que monolithique. Chaque couche est entraînée avec 2 500 échantillons de visages et d'arrière plans. Chaque couche a pour objectif un taux de vrais positifs égal à 0,999 et de faux positifs égal à 0,5.

Dans l'évaluation, nous comparons deux approches que nous proposons contre le détecteur de Viola et Jones. Le résumé complet est présenté dans le tableau 3. Nos deux approches impliquent un apprentissage sur notre base d'apprentissage. Notre première approche, EXP1, consiste à utiliser des descripteurs de Haar rectangulaires avec l'Adaboost réel comme classifieur fort.

Dans la seconde approche, EXP2, nous utilisons les descripteurs de Haar polygonaux introduits par Pham *et al.* et, là aussi, l'Adaboost réel. Nos deux détecteurs sont entraînés pour atteindre un taux de vrais positifs de 0,98 et un taux de faux positifs de 2×10^{-6} .

Tableau 3. Résumé des expérimentations

| Nom | Viola and Jones | Dupuis <i>et. al</i> EXP1 | Dupuis <i>et. al</i> EXP2 |
|--------------------|-----------------------------|-----------------------------|-----------------------------|
| Pré-traitement | Variance normalisée | Variance normalisée | Centrée réduite |
| Attribut de Haar | Rectangulaire | Rectangulaire | Polygonal |
| Classifieur faible | Arbre de décision à un nœud | Arbre de décision à un nœud | Arbre de décision à un nœud |
| Classifieur fort | Adaboost discret | Adaboost réel | Adaboost réel |
| Nombre de couches | 22 | 19 | 19 |

5.2. Invariance radiale

Dans cette section, nous nous intéressons au taux de détection de visages le long d'un rayon. Ainsi, la rotation dans le plan est nulle. Les variabilités sont dues uniquement à la position du visage sur le rayon. Comme nous pouvons le voir sur la figure 12a, le taux de détection de Viola et Jones chute de 23 % comparé à la performance annoncée sur la base MIT+CMU. Cette diminution s'explique par le fait que le détecteur ait été entraîné pour des images perspectives alors que les images de test sont des images omnidirectionnelles. Malgré un entraînement sur des images omnidirectionnelles synthétisées, les performances de EXP1 sont relativement basses. Nous nous attendions à un écart plus important avec le détecteur de Viola et Jones. L'écart avec EXP2 est vraiment significatif puisque, rappelons le, la seule différence est la nature des descripteurs utilisés. Dans le cas des images perspectives, celui-ci est de 5 % (Pham *et al.*, 2010) alors qu'il est ici de 20 % environ. Les attributs polygonaux sont donc plus descriptifs que les attributs rectangulaires sur les images omnidirectionnelles.

La figure 11 donne un aperçu de la courbe ROC pour ces deux approches. Comme nous pouvons le constater, les attributs polygonaux surpassent grandement les attributs rectangulaires; le taux de faux positifs étant beaucoup moins important pour un taux de vrais positifs donné. De fait, les attributs polygonaux sont donc aussi plus discriminants que les attributs rectangulaires dans le cadre des images omnidirectionnelles.

5.3. Invariance rotationnelle

Les images omnidirectionnelles impliquent une rotation dans le plan de par leur processus de formation. Ainsi, nous souhaitons évaluer la robustesse de notre dé-

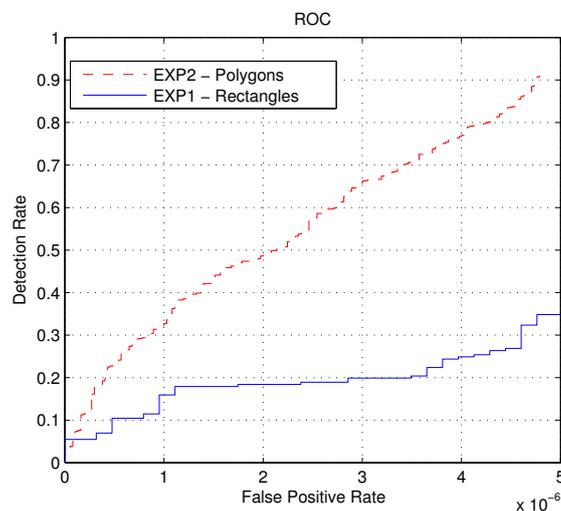
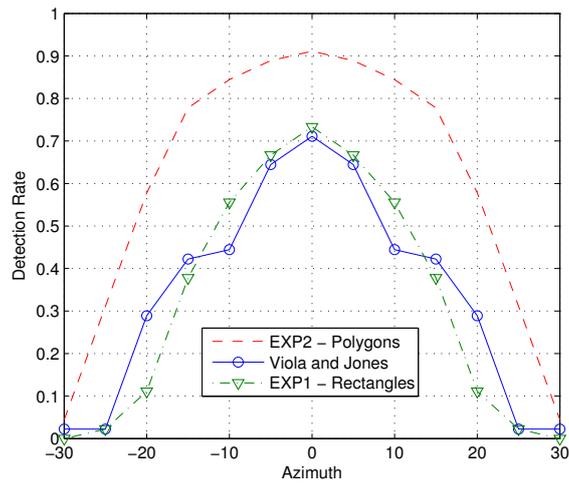


Figure 11. Taux de détection sur la base de test - Courbe ROC

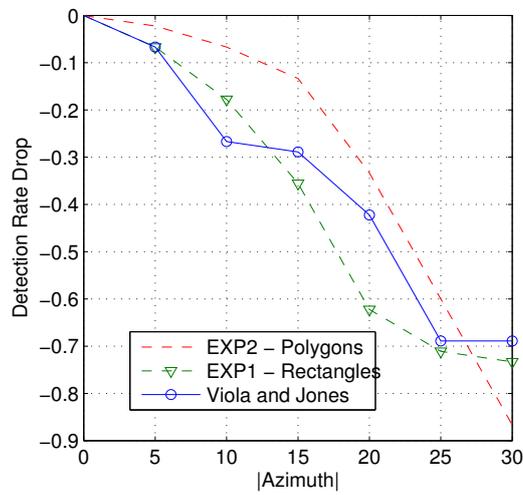
tecteur, entraîné spécifiquement pour un rayon donné, à cette rotation dans le plan. Pour évaluer cet impact uniquement, nous avons effectué une rotation des images omnidirectionnelles par rapport à leurs centres. Comme nous pouvons le voir dans la figure 12a, la courbe est symétrique par rapport à zéro. De plus, les attributs polygonaux surpassent là encore les attributs rectangulaires et le détecteur de Viola et Jones sur tous les angles évalués. De surcroît, alors que EXP1 et Viola et Jones chutent significativement entre $|\theta| = 15^\circ$ et $\theta = 0^\circ$, EXP2 ne perd que 13% de taux de détection. Cela démontre que les attributs polygonaux sont plus robustes à la rotation dans le plan que les attributs rectangulaires. Cela nous conforte dans l'idée que les attributs polygonaux sont mieux adaptés que les attributs rectangulaires. Deux exemples de détection peuvent être vus sur la figure 13.

5.4. Pertinence des attributs sélectionnés

Nous avons montré jusque là que les attributs étaient statistiquement descriptifs et discriminants. Nous nous intéressons maintenant à la pertinence morphologique des premiers attributs sélectionnés (c.f. figure 14). Les premiers attributs sont en fait des régions morphologiques reconnaissables. Le premier attribut localise le fort contraste qui existe entre la région des yeux et du front (c.f. figure 14a). Le second attribut correspond au contraste qui existe entre le coin de l'œil et le nez (c.f. figure 14b). Le cinquième attribut permet de trouver le contraste qui existe entre l'œil et la région péri-orbitale (c.f. figure 14c). Ainsi, les attributs polygonaux cherchent bien à localiser des régions que nous utilisons nous aussi pour caractériser un visage.



(a)



(b)

Figure 12. Performance en fonction de la rotation dans le plan (en degrés)
 (a) Performance absolue (b) Performance relative

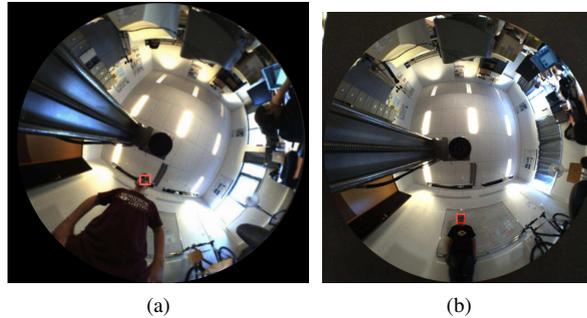


Figure 13. Détection sur la base de test (a) Angle: -25° (b) Angle: 0°

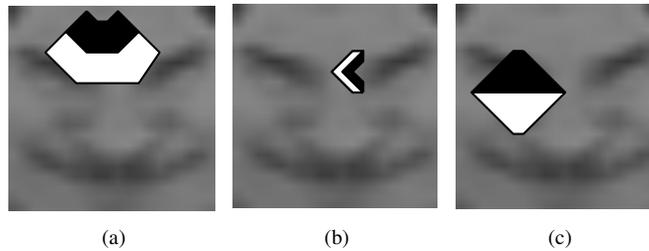


Figure 14. Premiers attributs sélectionnés (a) Premier attribut (b) Deuxième attribut (c) Cinquième attribut

Du point de vue de l'algorithme d'apprentissage, les attributs apparaissent aussi pertinents. En effet, comme présenté dans le tableau 4, le taux de confiance associé aux attributs de la figure 14 apparaissent comme corrects pour un classifieur faible et rentre en adéquation avec le philosophie de l'Adaboost. En effet, un classifieur faible qui serait trop fort entraîne une chute dans la performance globale de la procédure de boosting (Wickramaratna *et al.*, 2001).

Tableau 4. Pouvoir discriminant des attributs

| | Attribut 1 | | Attribut 2 | | Attribut 5 | |
|-------------------|------------|--------|------------|--------|------------|--------|
| $\text{sgn}(C_j)$ | -1 | 1 | -1 | 1 | -1 | 1 |
| $ C_j $ | 0.8427 | 1.1164 | 0.5419 | 0.8912 | 0.3981 | 0.5406 |
| Bin Error Rate | 0.179 | 0.14 | 0.253 | 0.171 | 0.315 | 0.259 |

5.5. Performance de détection et dépliement

Dans cette dernière section, nous nous focalisons sur la comparaison de la performance de EXP2 avec l'approche qui était utilisée auparavant, c'est-à-dire un dé-

plissement d'image omnidirectionnelle sur laquelle est appliqué le détecteur de Viola et Jones. Nous avons effectué un dépliement sphérique tel que nous le préconisons dans nos travaux précédents (Dupuis *et al.*, 2010). Les performances de classifications obtenues sur notre base de test sont données dans le tableau 5. Nous trouvons aussi dans ce tableau la performance obtenue pour EXP2 au point de fonctionnement choisi sur la courbe ROC. Comme nous pouvons le constater, notre approche directe surpasse l'approche par dépliement. Le taux de faux positifs reste bas pour les deux méthodes. De ce point de vue, l'approche par dépliement obtient des meilleurs résultats. Cependant, lorsque nous regardons la figure 11, nous nous rendons compte que pour le taux de faux positifs atteints par l'approche de dépliement, notre approche directe atteint un taux de vrais positifs de 87 %. Il est aussi intéressant de mentionner que pour le taux de vrais positifs de l'approche par dépliement, notre approche directe donne un taux de faux positifs de $4,384 \times 10^{-5} \%$. Ce dernier résultat est très prometteur puisque notre méthode surpasse l'approche par dépliement sur plus de 25° bien que nous ne traitons pas les rotations dans le plan pour le moment (c.f. figure 12b). Ainsi, quinze détecteurs spécifiquement entraînés pour une plage de rotation dans le plan permettraient de surpasser l'approche par dépliement que ce soit du point de vue de la performance de détection mais aussi de calcul.

Tableau 5. Performances au point de fonctionnement choisi

| | Projection sphérique | Approche directe |
|-----------------------|----------------------------|--------------------------|
| Taux de détection | 81.63 % | 91.11% |
| Taux de faux positifs | $4.6679 \times 10^{-5} \%$ | $4.79 \times 10^{-5} \%$ |

6. Conclusion

Nous avons présenté une méthode de détection de visages sur images omnidirectionnelles qui s'adapte à la nature intrinsèque de ces images. Notre approche utilise la méthodologie usuelle de détection d'objets. Dans un premier temps, nous avons présenté la manière dont les images omnidirectionnelles étaient actuellement traitées. Ensuite, nous nous sommes intéressés aux problèmes conceptuels et pratiques induits par le dépliement des images. Nous avons notamment montré que le temps de traitement augmente significativement lorsque l'image est dépliée. Ce temps de traitement étant un facteur important, nous avons basé notre approche directe sur des techniques qui ont montré une efficacité calculatoire notable.

La première contribution est une étude qualitative des déformations subies par les images omnidirectionnelles dans le cadre de la détection de visages. Les deux outils mathématiques utilisés montrent que la variabilité intra-classe est plus forte que dans le cas des images perspectives. La variabilité est à la fois géométrique et colorimétrique. Par conséquent, les deux classes étudiées seront moins disjointes et donc plus difficiles à distinguer.

La seconde contribution est la mise en exergue qu'une attention toute particulière doit être portée au choix des descripteurs et non pas à une éventuelle représentation intermédiaire de l'image omnidirectionnelle.

La troisième contribution est d'avoir montré qu'un détecteur d'objets, ici le visage, ne demande pas une base importante d'images omnidirectionnelles. La synthèse, à partir d'images perspectives, donne des performances convenables. Nous avons notamment démontré que cela était le cas pour les attributs de Haar. Nous pensons que cela pourrait aussi être le cas pour d'autres descripteurs.

Les figures 15 et 16 illustrent les performances de notre détecteur sur des images non incluses dans notre base de test. Les deux couleurs permettent de distinguer les cas où le détecteur de Viola et Jones fonctionne du cas où seul notre détecteur fonctionne.

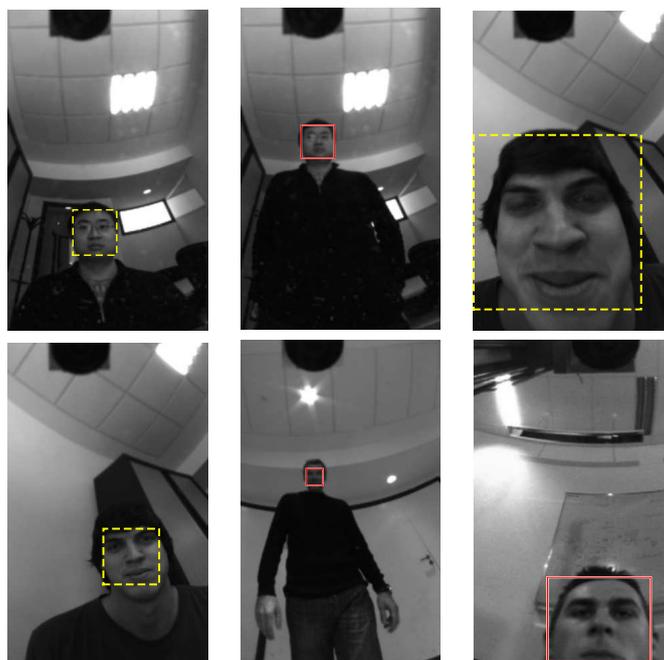
Nos futurs travaux s'attacheront à la prise en compte des rotations dans le plan par leur introduction dans le jeu de d'entraînement. L'objectif sera alors de minimiser le nombre supposé de détecteurs pour couvrir la rotation totale dans le plan.

Remerciements

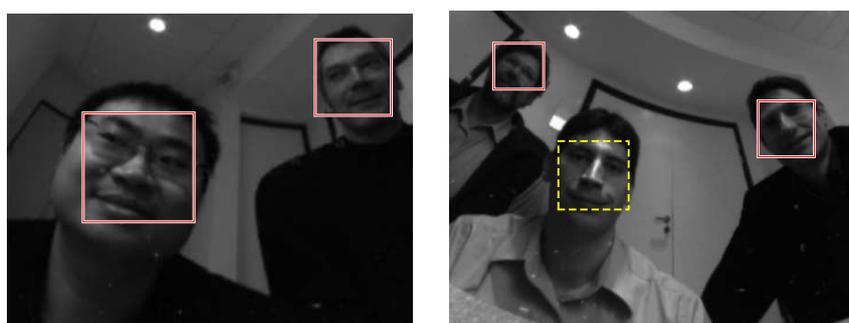
Ces travaux ont été menés dans le cadre du projet NOBA, qui a été sélectionné dans le cadre du programme européen de coopération transfrontalière INTERREG IVA France (Manche) - Angleterre, cofinancé par le FEDER. Le projet NOBA a également bénéficié du soutien de la région Haute-Normandie.

Bibliographie

- Arıcan Z., Frossard P. (2010). OmniSIFT: Scale Invariant Features in Omnidirectional Images. In *IEEE International Conference on Image Processing*, p. 3505-3508.
- Baker S., Nayar S. (1999). A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, vol. 35, n° 2, p. 175-196.
- Barczak A. L. C., Johnson M. J., Messom C. H. (2005). Real-time Computation of Haar-like features at generic angles for detection algorithms. *Res. Lett. Inf. Math. Sci.*, p. 1175-2777.
- Barczak A. L. C., Okamoto J. J., Grassi V. J. (2009). Face Tracking Using a Hyperbolic Catadioptric Omnidirectional System. *Res. Lett. Inf. Math. Sci.*, vol. 13, p. 55-67.
- Barreto J. (2003). *General Central Projection Systems : Modeling, Calibration and Visual Servoing*. Thèse de doctorat non publiée, University of Coimbra.
- Barreto J. (2006). A unifying geometric representation for central projection systems. *Computer Vision and Image Understanding*, vol. 103, n° 3, p. 208-217.
- Bazin J., Demonceaux C., Vasseur P., Kweon I. (2009). Motion Estimation by Decoupling Rotation and Translation in Catadioptric Vision. *Computer Vision and Image Understanding*, vol. 114, p. 254-273.
- Bradski G. R., Kaehler A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly.



(a) Individuels



(b) Groupes

Figure 15. Intérieur.

Ligne pointillée : Visages détectés par VJ & EXP2 - Ligne continue: EXP2 uniquement



(a) Individuels



(b) Groupes

Figure 16. Extérieur

Ligne pointillée : Visages détectés par VJ & EXP2 - Ligne continue: EXP2 uniquement

Bunschoten R. (2003). *Mapping and Localization from a Panoramic Vision Sensor*. Thèse de doctorat non publiée, University of Amsterdam.

Demonceaux C., Vasseur P., Fougerolle Y. D. (2011). Central catadioptric image processing with geodesic metric. *Image Vision Computing*, vol. 29, n° 12, p. 840-849.

Douchamps D., Campbell N. (2007). Robust Real Time Face Tracking for the Analysis of Human Behavior. In *International Conference on Machine Learning for Multimodal Interaction*, p. 1-10.

- Du S., Zheng N., You Q., Wu Y., Yuan M., Wu J. (2006). Rotated haar-like features for face detection with in-plane rotation. *Interactive Technologies and Sociotechnical Systems*, p. 128–137.
- Dupuis Y., Savatier X., Ertaud J.-Y., Hoblos G. (2010). A Framework for Face Detection on Central Catadioptric Systems. In *IEEE International Workshop on Robotic and Sensors Environments*.
- Freund Y., Schapire R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, vol. 55, p. 119–139.
- Geyer C. (2003). *Catadioptric Projective Geometry: Theory and Applications*. Thèse de doctorat non publiée, University of Pennsylvania.
- Geyer C., Daniilidis K. (2001). Catadioptric projective geometry. *IEEE International Journal of Computer Vision*, vol. 45, n° 3, p. 223–243.
- Jain V., Learned-Miller E. (2010). *Fddb: A Benchmark for Face Detection in Unconstrained Settings*. Rapport technique n° UM-CS-2010-009. University of Massachusetts, Amherst.
- Lhuillier M. (2008). Toward automatic 3D modeling of scenes using a generic camera model. In *IEEE International Conference on Computer Vision and Pattern Recognition*, p. 1–8.
- Lienhart R., Maydt J. (2002). An Extended Set of Haar-like Features for Rapid Object Detection. In *IEEE International Conference on Image Processing*, vol. 1, p. 900–903.
- Lin S., Gu J., Yamazaki S., Shum H. (2004). Radiometric calibration from a single image. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, p. II–938.
- Lourenço M., Barreto J. P., Malti A. (2010). Feature detection and matching in images with radial distortion. In *IEEE International Conference on Robotics and Automation*, p. 1028–1034.
- Mei C., Rives P. (2007). Single View Point Camera Calibration from Planar Grids. In *IEEE International Conference on Robotics and Automation*, p. 3945–3950.
- Mei C., Sommerlade E., Sibley G., Newman P., Reid I. (2011, May). Hidden View Synthesis using Real-Time Visual SLAM for Simplifying Video Surveillance Analysis. In *IEEE International Conference on Robotics and Automation*, p. 4240–4245.
- Pham M., Gao Y., Hoang V., Cham T. (2010). Fast polygonal integration and its application in extending haar-like features to improve object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 942–949.
- Rowley H. A., Baluja S., Kanade T. (1996). Neural Network-Based Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, p. 23–38.
- Scaramuzza D., Martinelli A., Siegwart R. (2006). A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *ICVS*, p. 45–45.
- Schapire R. E., Singer Y. (1999). Improved Boosting Algorithms Using Confidence-rated Predictions. In *Machine Learning*, p. 80–91.
- Schneiderman H., Kanade T. (2000). A statistical method for 3D object detection applied to faces and cars. In *IEEE International Conference on Computer Vision*.

- Strauss O., Comby F. (2007). Variable structuring element based fuzzy morphological operations for single viewpoint omnidirectional images. *Pattern Recognition*, vol. 40, p. 3578 - 3596.
- Sturm P., Ramalingam S., Tardif J., Gasparini S., Barreto J. (2011). Camera Models and Fundamental Concepts Used in Geometric Computer Vision. *Foundations and Trends® in Computer Graphics and Vision*, vol. 6, n° 1-2, p. 1-183.
- Svoboda T. (1999). *Central panoramic cameras design, geometry, egomotion*. Thèse de doctorat non publiée, Center of Machine Perception, Czech Technical University in Prague.
- Viola M., Jones M. J. (2003). Fast Multi-view Face Detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*.
- Viola M., Jones M. J. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision*, vol. 57, n° 2, p. 137-154.
- Wickramaratna J., Holden S., Buxton B. (2001). Performance Degradation in Boosting. In J. Kittler, F. Roli (Eds.), *Multiple Classifier Systems*, vol. 2096, p. 11-21. Springer Berlin / Heidelberg.

Article reçu le 4/10/2013

Accepté le 14/04/2014

