



**HAL**  
open science

## Détection radiale de visages sur images omnidirectionnelles

Y. Dupuis, Xavier Savatier, Jean-Yves Ertaud, Pascal Vasseur

► **To cite this version:**

Y. Dupuis, Xavier Savatier, Jean-Yves Ertaud, Pascal Vasseur. Détection radiale de visages sur images omnidirectionnelles. Orasis, Congrès des jeunes chercheurs en vision par ordinateur, Jun 2013, Cluny, France. hal-00829350

**HAL Id: hal-00829350**

**<https://hal.science/hal-00829350>**

Submitted on 3 Jun 2013

**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.

# Détection radiale de visages sur images omnidirectionnelles

Y. Dupuis<sup>1,3</sup>

X. Savatier<sup>1</sup>

J.Y. Ertaud<sup>1</sup>

P. Vasseur<sup>2</sup>

<sup>1</sup> IRSEEM, ESIGELEC

<sup>2</sup> LITIS Lab, Université de Rouen

<sup>3</sup> CETE Normandie-Centre, Ministère de l'Écologie, du Développement Durable et de l'Énergie

Rouen, France

yohan.dupuis@developpement-durable.gouv.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 toute particulière doit être portée sur le choix des descripteurs lors de l'adaptation d'algorithmes développés pour la vision perspective à la vision omnidirectionnelle.*

## Mots Clef

Boosting, Détection de visages, 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.*

## Keywords

Boosting, Face Detection, Omnidirectional Vision.

## 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. La maturité et les performances atteintes par les algorithmes de traitement ont rendu possible leurs transferts 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 [25]. Un ordinateur ou une machine équipé d'une caméra est maintenant capable de détecter un visage en temps réel. Viola et Jones n'étaient pas des pionniers en la matière [21, 19]. Cependant, leur approche 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 toute 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 part leur champs 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. La solution proposée implique bien souvent l'utilisation d'un réseau de caméras. Les problèmes que la mise en œuvre de tels systèmes induisent peuvent être résolus.

Les caméras omnidirectionnelles sont aujourd'hui principalement utilisées en robotique mobile. Elles permettent de répondre aux problématiques de navigation, d'estimation de mouvements [5, 17] ou bien de reconstruction 3D de scènes [13]. 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

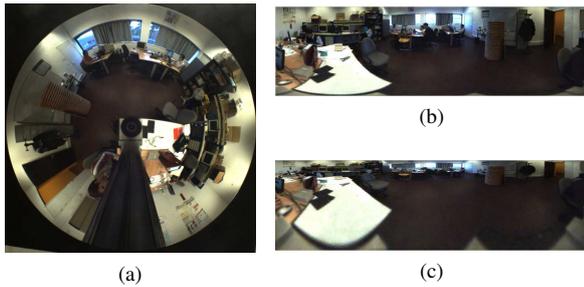


FIGURE 1 – Exemples de dépliements. (a) Image originale. (b) Projection sphérique. (c) Projection cylindrique.

points et des petites régions à proximité du pixel considéré [23]. 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 [1, 6, 15]. 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 [22] (c.f. Figure 1). Une fois la représentation intermédiaire de l'image omnidirectionnelle obtenue, le détecteur entraîné pour des images perspectives est appliqué.

## 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 en 2007 [7]. 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 d'un filtre à base de teinte de peau qui se révèle bruité et sensible aux conditions d'éclairage.

En 2009, Barczak *et al.* utilisent un capteur de vision omnidirectionnelle pour effectuer le suivi de visages [3]. 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. De plus, ces travaux ont confirmé une faisabilité sans toute fois apporter une étude qualitative de la performance du détecteur de visages sur des images dépliées.

En 2010, nous proposons de répondre à ce point [9]. 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. Ainsi, nous avons montré que le dépliement sphérique permet d'obtenir de meilleures performances.

Dans cet article, nous proposons de montrer que la détection de visages est envisageable sur une image omnidirectionnelle. Nous faisons notamment apparaître 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.

## 2 Les conséquences du dépliement

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. Ainsi, un détecteur entraîné sur les images perspectives peut être directement appliqué. Cependant, de nombreux problèmes découlent de cette approche.

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 notamment 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 2). La solution proposée est 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.

Troisièmement, même si l'anti-anamorphose vise à retrouver des propriétés géométriques proches des images pers-

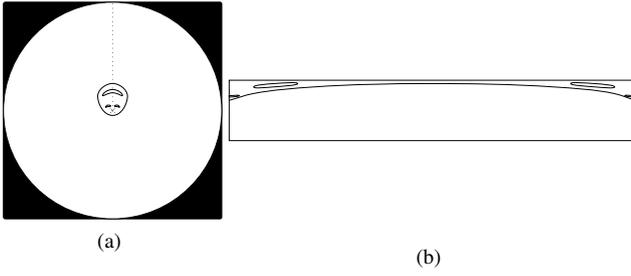


FIGURE 2 – Illustration de la segmentation de visages. (a) Dessin omnidirectionnel. (b) Image dépliée résultante.

pectives, 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 :

$$\mathcal{N} = \sum_{20 \leq i \leq n} (M - i) \times (N - i) \quad (1)$$

pour une sous-région de taille minimale  $20 \times 20$  pixels, une fenêtre de recherche qui augmente par pas de 1 pixel à chaque passe 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 1 devient donc :

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

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  $\pi n$ . L'équation 1 devient alors :

$$\mathcal{N}_{panoramic\ image} = \sum_{20 \leq i \leq \frac{n}{2}} \left( \frac{\pi}{2} n^2 - \frac{2\pi + 1}{2} ni + i^2 \right) (3)$$

La figure 3 donne un aperçu de l'évolution des équations 2 et 3 lorsque  $n$  varie de 200 à 2000. L'équation 2 est théorique si un seul détecteur est considéré. Cependant en adoptant l'approche utilisée par [24], plusieurs détecteurs entraînés pour une pose spécifique peuvent être fusionnés

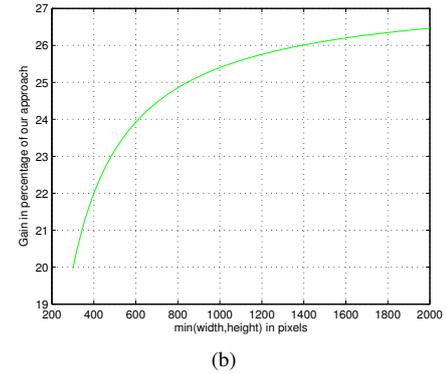
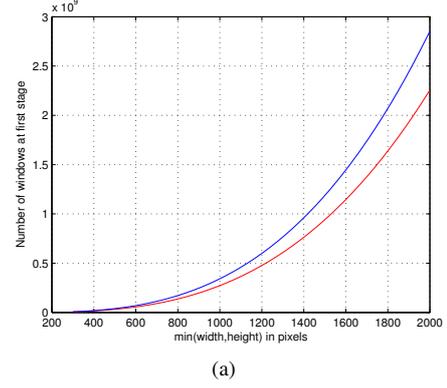


FIGURE 3 – (a) Nombre de fenêtres évaluées au premier étage sans dépliement (rouge) et avec dépliement (bleu). (b) Gain en % d'une approche directe comparée aux méthodes existantes.

pour permettre une détection robuste à la rotation dans le plan. En conséquence, l'équation 2 est totalement applicable. Comme indiqué, la réduction du nombre de sous-régions à évaluer atteint 25% pour une image  $1024 \times 1024$  pixels.

Dans cette première partie, nous avons montré que considérer une image omnidirectionnelle dépliée comme perspective est, à la fois, erroné mais aussi source de calculs supplémentaires. De plus, l'ajout de bordures supplémentaires imposées par la dés-anamorphose 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 partie à l'exposition de notre approche pour traiter ces problèmes.

## 3 Méthodologie

### 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

Paramètre	Valeur
$\xi$	1.00609
$\gamma_u$	391
$\gamma_v$	389
$u_0$	512
$v_0$	512

TABLE 1 – Paramètres du capteur de vision omnidirectionnelle

perspectives, les chercheurs peuvent cependant utiliser les bases de données des visages qui existent 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. Ainsi, ils obtiennent une base comprenant une diversité importante de teintes de peaux, de conditions d’illuminations qui auront pour conséquences d’obtenir un détecteur plus robuste aux conditions réelles [12]. De plus, les amers des visages ont tendance à être très 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.

Dans ces travaux, nous nous intéresserons 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 au modèle unifié proposé par Mei *et al.* [16] lui-même basé sur l’approche de Barreto et Geyer [11, 4]. 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 patch 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’entraî-

---

### Algorithm 1 Algorithme de l’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: for  $t = 1, \dots, T$  do
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:   if  $\epsilon_t \geq \frac{1}{2}$  then
6:     BREAK
7:   else
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:     $D_{t+1}(i) = \frac{D_t(i) \exp(-y_i c_{j,t}(x_i))}{Z_y}$  {où  $Z_t$  est le facteur de
      normalisation}
11:   end if
12: end for
13:  $H(x) = \text{sign}(\sum_{t=1}^T c_{j,t}(x))$ 

```

---

nement, nous considérons toutes les transformations géométriques possibles qui amèneraient le plan visage 3D  $\Pi_f$  à voir la projection de son point principal  $x_{pp}$  le long du rayon  $r$ . Nous définissons  $\pi_f$  la région d’intérêt correspondante et  $\mathbb{F}$  l’ensemble des  $\pi_f$  pour un rayon donné. Pour une image 1024 x 1024,  $\mathbb{F}$  comprend environ 100 000 éléments. Pour chaque  $\pi_f$ , grâce au modèle de la sphère unitaire, nous sommes capables de définir la fonction bijective suivante :

$$\Pi_f \xrightarrow{\sim} \pi_f \quad (4)$$

Par conséquent, nous sommes en mesure de synthétiser tout les  $\pi_f$  à partir d’une image perspective  $\Pi_f$ . 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.

## 3.2 Méthode d’apprentissage

La méthode d’apprentissage la plus utilisée pour la détection de visages est l’Adaboost. 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 :

$$H(X) = \sum_{k=1}^n \alpha_k h_k(x) \quad (5)$$

L’Adaboost implique deux pondérations. Les classifieurs faibles, comme indiqué dans l’équation 5, sont pondérés tout comme les exemples de la base d’entraînement. Les deux pondération sont intrinsèquement liées puisque le poids de chaque classifieur faible est proportionnel à son erreur sur la base d’entraînement étant donné la distribution des poids des observations. Plusieurs variantes d’Adaboost existent. Dans ces travaux, nous nous intéresserons

à l'Adaboost réel [20]. Généralisation de l'Adaboost discret [10], l'Adaboost réel a montré qu'il surpassait les autres variantes d'Adaboost dans la détection de visages. Comme montré dans [20], 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 réel est présenté dans l'algorithme 1.

### 3.3 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 [25], 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 [14, 2, 8]. 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 [22]. Ainsi, un noyau de filtrage rectangulaire devient un tétragone dont deux de ces 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. table 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 [18]. Le nombre d'opérations dépend uniquement du nombre de sommets du polygone. Ces travaux ont aussi montré que les attributs polygonaux surpassent les attributs rectangulaires sur les images perspectives. Les attributs polygonaux utilisés dans ces travaux se trouvent être 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 7).

## 4 Expérimentations

### 4.1 Protocole expérimental

La première étape est la construction d'une base d'apprentissage. Pour ce faire, nous avons synthétisé 2500 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 4). 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.

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

TABLE 2 – Transformation des attributs de Haar

Ainsi, les sujets pouvaient être assis ou debouts et ceci à plusieurs distances du capteur.

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 2500 échantillons de visages et d'arrière plans. Chaque couche a pour objectif un taux de vrais positifs égale à 0.999 et de faux positifs égale 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.

### 4.2 Invariance radiale

Dans cette partie, nous nous intéressons au taux de détection pour une rotation dans le plan nulle. Les variabilités sont dues uniquement à la position du visage sur le rayon. Comme nous pouvons le voir sur la figure 5b, 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% [18] 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 5a 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.



FIGURE 4 – Exemples d’images d’arrières plans utilisées pour l’apprentissage

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

TABLE 3 – Résumé des Expérimentations

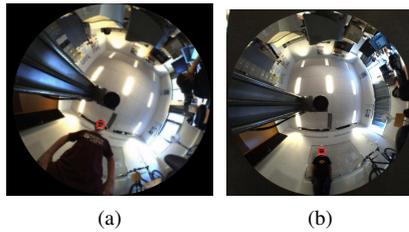


FIGURE 6 – Détection sur la base de test. (a) Angle :  $-25^\circ$ . (b) Angle :  $0^\circ$ .

### 4.3 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 détecteur, 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 5b, 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 6.

### 4.4 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 7). 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 7a). Le second attribut correspond au contraste qui existe entre le coin de l’oeil et le nez (c.f. figure 7b). Le cinquième attribut permet de trouver le contraste qui existe entre l’oeil et la région peri-orbitale (c.f. figure 7c). Ainsi, les attributs

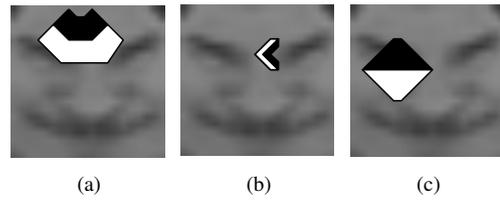


FIGURE 7 – Premiers attributs sélectionnés (a) Premier attribut. (b) Deuxième attribut. (c) Cinquième attribut.

polygonaux cherchent bien à localiser des régions que nous utilisons nous aussi pour caractériser un visage.

### 4.5 Performance de détection et dépliement

Dans cette dernière partie, nous nous focalisons sur la comparaison de la performance de EXP2 avec l’approche qui était utilisée auparavant, c’est-à-dire un dépliement d’image omnidirectionnelle sur laquelle est appliqué le détecteur de Viola et Jones. Nous avons effectué un dépliement sphérique tels que nous le préconisons dans nos travaux précédents [9]. Les performances de classifications obtenues sur notre base de test sont données dans le tableau 4. 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 semble obtenir des meilleurs résultats. Cependant, lorsque nous regardons la figure 5a, 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%, soit supérieur. 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, bien que nous ne traitons pas les rotations dans le plan pour le moment, notre méthode surpasse l’approche par dépliement sur plus de  $25^\circ$  (c.f. figure 5c). 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.

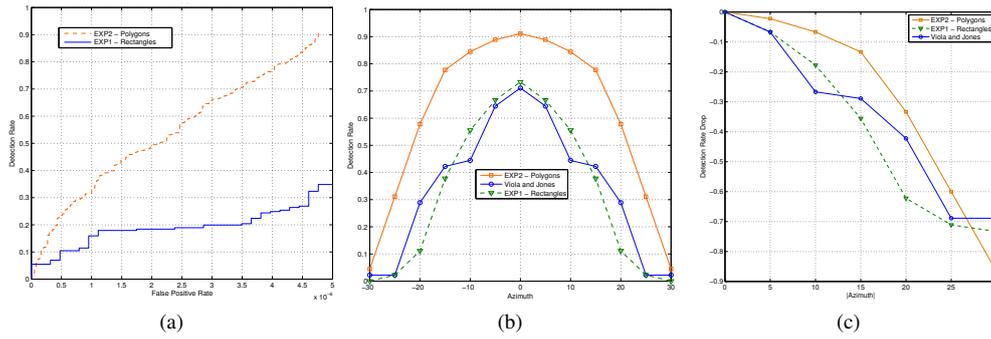


FIGURE 5 – Taux de détection sur la base de test. (a) Courbe ROC. (b) Performance absolue en fonction de la rotation dans le plan (en degrés). (c) Performance relative.

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

TABLE 4 – Performances au point de fonctionnement choisi

## 5 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 la 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.

Cette étude a principalement mis en avant le fait 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 seconde contribution est d’avoir montré qu’un détecteur d’objets, ici le visage, ne demande pas une base importante d’images omnidirectionnelle. La synthèse, à partir d’images perspectives, donnent 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.

Nos futurs travaux s’attacheront à la prise en compte des rotations dans le plan par leurs 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 financés en partie par la Région Haute-Normandie et par le Fond Européen de Développement Ré-

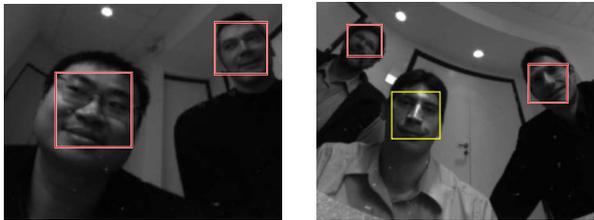
gional dans le cadre du projet INTERREG IV A NOBA qui donne lieu à une collaboration avec l’Université du Kent, Royaume-Uni.

## Références

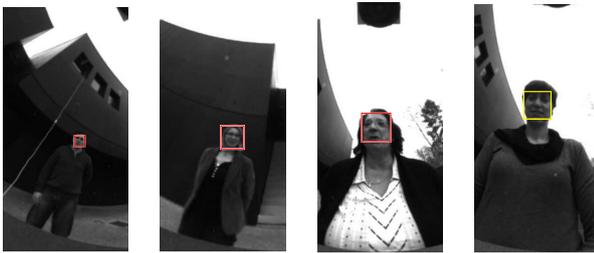
- [1] Zafer Arican and Pascal Frossard. OmniSIFT : Scale invariant features in omnidirectional images. In *IEEE International Conference on Image Processing*, pages 3505–3508, 2010.
- [2] A. L. C. Barczak, M. J. Johnson, and C. H. Messom. Real-time computation of haar-like features at generic angles for detection algorithms. *Res. Lett. Inf. Math. Sci.*, pages 1175–2777, 2005.
- [3] A. L. C. Barczak, J. J. Okamoto, and V. J. Grassi. Face Tracking Using a Hyperbolic Catadioptric Omnidirectional System. *Res. Lett. Inf. Math. Sci.*, 13 :55–67, 2009.
- [4] J.P. Barreto. A unifying geometric representation for central projection systems. *Computer Vision and Image Understanding*, 103(3) :208–217, 2006.
- [5] J.C. Bazin, C. Demonceaux, P. Vasseur, and I. Kweon. Motion Estimation by Decoupling Rotation and Translation in Catadioptric Vision. *Computer Vision and Image Understanding*, 114 :254–273, 2009.
- [6] Cédric Demonceaux, Pascal Vasseur, and Yohan D. Fougerolle. Central catadioptric image processing with geodesic metric. *Image Vision Computing*, 29(12) :840–849, 2011.
- [7] D. Douxchamps and N. Campbell. Robust Real Time Face Tracking for the Analysis of Human Behavior. In *International Conference on Machine Learning for Multimodal Interaction*, pages 1–10, 2007.
- [8] S. Du, N. Zheng, Q. You, Y. Wu, M. Yuan, and J. Wu. Rotated haar-like features for face detection with in-plane rotation. *Interactive Technologies and Sociotechnical Systems*, pages 128–137, 2006.
- [9] Y. Dupuis, X. Savatier, Jean-Yves Ertaud, and G. Hoblos. A Framework for Face Detection on Central Catadioptric Systems. In *IEEE International Workshop on Robotic and Sensors Environments*, 2010.
- [10] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences.*, 55 :119–139, 1997.



(a)



(b)



(c)



(d)

FIGURE 8 – Exemples supplémentaires. (a) Personne seule en intérieur. (b) Groupe en intérieur. (c) Personne seule en extérieur. (d) Groupe en extérieur. Jaune : détecté par Viola et Jones et EXP2. Rouge : détecté uniquement par EXP2

- [11] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *IEEE International Journal of Computer Vision*, 45(3) :223–243, 2001.
- [12] Vidit Jain and Erik Learned-Miller. FDDB : A Benchmark for Face Detection in Unconstrained Settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.
- [13] Maxime Lhuillier. Toward automatic 3D modeling of scenes using a generic camera model. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [14] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE International Conference on Image Processing*, volume 1, pages 900–903. IEEE, 2002.
- [15] Miguel Lourenço, João P. Barreto, and Abed Malti. Feature detection and matching in images with radial distortion. In *IEEE International Conference on Robotics and Automation*, pages 1028–1034, 2010.
- [16] C. Mei and P. Rives. Single View Point Camera Calibration from Planar Grids. In *IEEE International Conference on Robotics and Automation*, pages 3945–3950, 2007.
- [17] Christopher Mei, Eric Sommerlade, Gabe Sibley, Paul Newman, and Ian Reid. Hidden View Synthesis using Real-Time Visual SLAM for Simplifying Video Surveillance Analysis. In *IEEE International Conference on Robotics and Automation*, pages 4240–4245, May 2011.
- [18] M.T. Pham, Y. Gao, V.D.D. Hoang, and T.J. Cham. Fast polygonal integration and its application in extending haar-like features to improve object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 942–949. IEEE, 2010.
- [19] H. A. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 :23–38, 1996.
- [20] Robert E. Schapire and Yoram Singer. Improved boosting algorithms using confidence-rated predictions. In *Machine Learning*, pages 80–91, 1999.
- [21] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *IEEE International Conference on Computer Vision*, 2000.
- [22] O. Strauss and F. Comby. Variable structuring element based fuzzy morphological operations for single viewpoint omnidirectional images. *Pattern Recognition*, 40 :3578 – 3596, 2007.
- [23] P. Sturm, S. Ramalingam, J.P. Tardif, S. Gasparini, and J. Barreto. Camera Models and Fundamental Concepts Used in Geometric Computer Vision. *Foundations and Trends® in Computer Graphics and Vision*, 6(1–2) :1–183, 2011.
- [24] M. Viola, Michael J. Jones, and Paul Viola. Fast multi-view face detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2003.
- [25] P. Viola and M. J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2) :137–154, 2004.