



**HAL**  
open science

## Une nouvelle approche de classification de personnes à partir d'une plate-forme multi-kinect

Kyis Essmaeel, Cyrille Migniot, Albert Dipanda

### ► To cite this version:

Kyis Essmaeel, Cyrille Migniot, Albert Dipanda. Une nouvelle approche de classification de personnes à partir d'une plate-forme multi-kinect. Journées francophones des jeunes chercheurs en vision par ordinateur, Jun 2015, Amiens, France. hal-01161844

**HAL Id: hal-01161844**

**<https://hal.science/hal-01161844>**

Submitted on 9 Jun 2015

**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.

# Une nouvelle approche de classification de personnes à partir d'une plate-forme multi-kinect

Kyis Essmaeel

Cyrille Migniot

Albert Dipanda

LE2I- UMR CNRS 6306, Université de Bourgogne  
9 avenue Alain Savary, 21078 Dijon Cedex  
{kyis.essmaeel, cyrille.migniot, albert.dipanda}@u-bourgogne.fr

## Résumé

*Dans ce papier nous présentons un descripteur 3D original pour la classification de personnes. Nous avons réalisé une plate-forme multi-kinect qui fait l'acquisition de nuages de points complets (c'est à dire selon une vue à 360°). Un post-traitement (suppression d'arrière-plan) isole les éléments qui sont candidats à représenter une personne. Notre descripteur exploite ces données et permet, au travers d'un classifieur SVM (Support Vector Machine), de classifier les personnes et de déterminer l'orientation du plan frontal de la personne. Il utilise la répartition spatiale et l'orientation des normales aux surfaces dans un procédé dérivé de celui des HOG (Histogram of Oriented Gradient).*

*Nous avons construit une base de données originale pour l'apprentissage du classifieur et l'évaluation de notre descripteur. Le descripteur permet de reconnaître une personne d'un objet de manière très satisfaisante et fournit également l'orientation du plan frontal de la personne avec une grande précision.*

## Mots Clef

Classification de personnes, descripteur 3D, système multi-kinect.

## Abstract

*In this paper we present a new 3D descriptor for human classification. We realize a multi-kinect platform that acquires the complete point cloud (i.e 360° view) of a subject. Post-processing (background subtraction) isolates a set of candidates. Our new descriptor exploits this information to classify person and determines the frontal direction of the person using the Support Vector Machine (SVM) classifier. It uses the spatial repartition of surface normal in a process derived from the HOG (Histogram of Oriented Gradient). We have created a new database to learn the classifier and to evaluate the efficiency of our descriptor. Our evaluation of the 3D descriptor shows promising results. The descriptor can discriminate human from non-human objects and provides the frontal face direction of the human with high precision.*

## Keywords

Human classification, 3D descriptor, Multi-kinect.

## 1 Introduction

Les travaux présentés ici s'inscrivent dans le développement du projet RAHA (Reconstruction et Analyse du mouvement Humain par stéréovision Active) initié par le laboratoire LE2I. L'objectif de ce projet est la mise en place d'une plate-forme d'acquisition d'images permettant la reconstruction 3D et l'analyse du mouvement temps réel d'un sujet déformable, sans marqueur. Les applications visées sont principalement l'analyse de performances des sportifs, la rééducation fonctionnelle et la surveillance des malades. Nous nous concentrerons ici sur le problème de la détection des personnes sur les données de reconstitution 3D produite par la plate-forme en question. Cette détection servira de phase d'initialisation d'un processus de suivi qui lui même servira à l'analyse du comportement (par exemple pour la reconnaissance de mouvements à risque ou l'analyse de progression de la rééducation).

La détection de personnes est un sujet qui suscite de nombreuses recherches. Elle est utilisée pour une grande variété d'applications telles que la surveillance médicale, l'assistance à la conduite ou bien l'analyse du comportement. Ce problème est particulièrement complexe du fait de la large variation de poses, de couleurs et de textures dans la classe des personnes ainsi que de la complexité de l'environnement dans lequel la personne évolue. Bien que la plupart des méthodes existantes de détection de personnes utilisent les images couleur, l'essor ces dernières années des technologies de capture de la profondeur a ouvert de multiples perspectives. L'introduction d'une caméra 3D précise mais abordable en terme de coûts, la kinect de Microsoft, a décuplé l'intérêt pour ces technologies. De nos jours de nombreuses méthodes de détection de personnes utilisent les capteurs de profondeur. De plus les données 3D sont de plus en plus utilisées dans les applications médicales de surveillance des malades [16]. Néanmoins les caractéristiques de l'enveloppe 3D d'une personne sont très peu ex-

exploitées. En effet l'information de profondeur est la plupart du temps utilisée uniquement pour réduire le temps des traitements.

On recense deux grandes catégories de méthodes de détection de personnes : d'une part la combinaison d'un descripteur avec un classifieur et d'autre part la comparaison à un gabarit. Dans la première catégorie, les HOG (Histogram of Oriented Gradients) [4] sont considérés comme le descripteur 2D le plus efficace pour la détection de personnes. Les HOD (Histogram of Oriented Depths) [13, 2] sont l'adaptation la plus connue des HOG appliquée aux images de profondeur. Les HOD calculent localement la direction de la variation de profondeur. En fait ils utilisent les données de profondeur en tant qu'image 2D pour appliquer le calcul du HOG. Il est alors difficile d'appliquer ce processus quand plusieurs sources d'information sont combinées comme c'est le cas avec un système multi-capteurs. Le problème est le même avec les RDSF [10] dont le descripteur rassemble les comparaisons entre régions de l'image de profondeur 2D. Dans la seconde catégorie de méthodes, les données 2D (images) ou 3D (nuage de points) sont comparées à un ou plusieurs gabarits de parties du corps. La forme distinctive en forme de  $\Omega$  formée par la tête et les épaules est un exemple de gabarit discriminant [14]. Pour la comparer aux données, Xia [15] utilise une distance de chanfrein et Choi [3] une distance de Hamming.

Dans cet article nous proposons une méthode de classification de personnes qui fonctionne à partir d'un nuage de points et qui exploite les seuls caractéristiques 3D des personnes et non les données couleur. Le descripteur 3D proposé ici peut être considéré comme une généralisation du descripteur HOG. Dans un premier temps le nuage de points 3D est séparé en blocs 3D. Le descripteur 3D est ensuite obtenu par le calcul de l'histogramme des orientations des normales des points pour chaque bloc. Enfin nous utilisons un classifieur SVM pour déterminer si le descripteur représente ou non une personne. De plus, le descripteur fournit une indication supplémentaire à propos de l'orientation du plan frontal de la personne. Une telle information est primordiale pour de nombreuses applications comme l'initialisation du suivi, les interactions homme-machine ou l'analyse du comportement.

Le descripteur proposé a été conçu pour manipuler des nuages de points complets (comme celui de la Figure 1). Nous avons conçu une plate-forme multi-kinect. Construire un tel système implique de répondre à plusieurs défis comme la calibration, les interférences et la suppression du bruit [6]. La plate-forme couvre la totalité de l'environnement de travail et donc la forme 3D du sujet dans son intégralité est reconstruite. Cela a de multiples avantages : aucune partie caractéristique de la personne n'est cachée, l'orientation du plan frontal de la personne peut être estimée, l'acquisition est robuste et la description de l'objet est plus riche (couleur, profondeur, normale aux surfaces, courbure). Ces données sont isotropes ; ce qui serait impossible à obtenir avec une unique caméra. Le descripteur

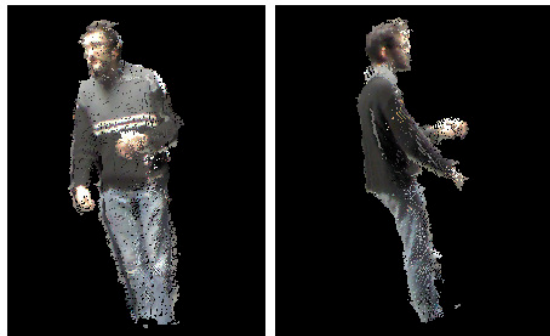


FIGURE 1 – Exemple de nuage de points complet selon deux points de vue différents. La surface de la personne est intégralement reconstruite. Pour une meilleure compréhension, le nuage de points est ici affiché avec les informations couleur qui ne sont pas utilisées dans nos traitements.

obtenu permet alors d'estimer l'orientation du plan frontal de la personne.

Dans la suite du papier, nous détaillerons tout d'abord la mise en place de la plate-forme d'acquisition, la construction du descripteur puis le processus de classification dans la section 2. Ensuite nous présenterons les résultats de nos évaluations dans la section 3 avant de formuler nos conclusions dans la section 4.

## 2 La classification de personnes

La méthode proposée pour la classification d'une personne et de son orientation suit l'approche descripteur/classifieur. Elle manipule un nuage de points complet. Un descripteur est tout d'abord calculé pour convertir l'information donnée par le nuage de points dans un espace plus discriminant. Puis le modèle du classifieur est construit en utilisant l'algorithme d'apprentissage du SVM.

### 2.1 Le système d'acquisition

Le descripteur 3D est calculé à partir d'un nuage de points complet, c'est à dire selon une vue à 360°. Afin d'obtenir cette couverture complète de la scène, une plate-forme multi-kinect a été mise en place. Cette plate-forme consiste en trois kinects positionnées de telle sorte que deux kinects successives partagent une partie de leur champ de vision.

Le système multi-kinect est ensuite calibré. L'estimation des paramètres intrinsèques permet de transformer l'image 3D en sortie de la kinect en nuage de points dans le repère monde. L'estimation des paramètres extrinsèques permet de réunir les nuages de points des différentes caméras dans un même référentiel. Il existe des méthodes efficaces reconnues d'estimation de ces paramètres [5, 9, 12]. Pour calculer la profondeur avec la kinect, une caméra infra-rouge capture la projection du motif issu d'un projecteur infra-rouge. Ensuite la disparité  $d$ , entre le motif pré-enregistré et celui capturé, est calculée. La profondeur est estimée pour

chaque pixel de l'image de profondeur par :

$$z = \frac{1}{c_v \times d + c_u} \quad (1)$$

où  $c_u$  and  $c_v$  sont les coordonnées du point central de l'image. Le capteur de profondeur de la kinect peut être décrit comme un modèle de sténopé. Le point de l'espace 3D de coordonnées  $(x,y,z)$  est projeté sur l'image au niveau du pixel de coordonnées  $(u,v)$  selon l'équation :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{avec } K = \begin{pmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

où  $K$  est le paramètre intrinsèque de la caméra,  $f_u$  et  $f_v$  sont les distances focales. On a alors :

$$x = z \frac{u - c_u}{f_u} \quad \text{et} \quad y = z \frac{v - c_v}{f_v} \quad (3)$$

Pour la seconde calibration, les paramètres extrinsèques sont les matrices de rotation  $R_{i,o}$  et de translation  $T_{i,o}$  entre le repère de la kinect et le repère de référence (qui est en général confondu avec celui d'une des kinects). Le nuage de points  $pc_i$  capturé par une kinect est converti dans le repère de référence en utilisant ces matrices. Finalement le nuage de points complet  $PC$  est obtenu par :

$$PC = \bigcup_{i \in [1,N]} R_{i,o} \cdot pc_i + T_{i,o} \quad (4)$$

où  $N$  est le nombre de kinects dans la plate-forme. Les applications visées ainsi que le fonctionnement des kinects impliquent des enregistrements dans un environnement fermé qui contiennent de nombreuses surfaces planes comme les sols ou les murs. Un post-traitement est donc réalisé sur les enregistrements pour supprimer ces surfaces et ainsi isoler les éléments candidats à représenter une personne. Un algorithme de RANSAC amélioré est appliqué au nuage de points complet. Les points ayant une normale de même orientation sont regroupés. Puis un algorithme traditionnel de RANSAC [7] est exécuté sur les groupes les plus importants. Si une surface bien plane contenant un nombre suffisant de points est reconnue, elle est supprimée du nuage de points. Ainsi, le nuage de points complet résultant contient uniquement l'ensemble des candidats à classifier.

## 2.2 La construction du descripteur

Le descripteur 3D proposé transpose le principe des HOG au nuage de points 3D. Pour les HOG, une fenêtre est divisée en une grille uniforme de blocs. Dans chaque bloc, les orientations des gradients de l'ensemble des pixels sont calculés et collectés dans un histogramme 1D. Calculer le gradient d'un nuage de points 3D n'a pas de sens. Le gradient est donc remplacé par la normale aux surfaces de chaque point (voir Figure 2). Comme une normale est un vecteur 3D, elle ne peut pas être associée à un histogramme

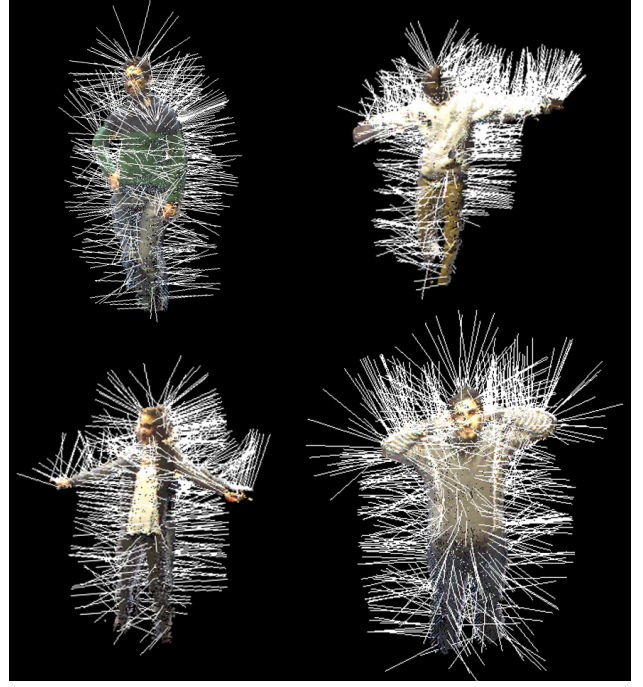


FIGURE 2 – Exemple de normales aux surfaces représentées pour un ensemble aléatoire de points pour plusieurs personnes.

1D. Pour résoudre ce problème, nous avons utilisé la quantification de l'orientation 3D proposée par Kläser [11]. Le procédé est illustré sur la Figure 3. Le vecteur normal est placé au centre d'un polyèdre régulier et est projeté sur les différentes faces du polyèdre.

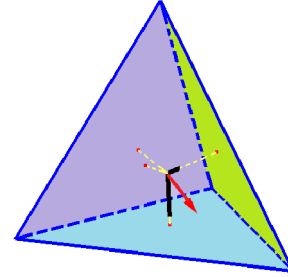


FIGURE 3 – Illustration de la quantification de l'orientation par projection sur un tétraèdre (en bleu). Le vecteur de la normale (en rouge) est projeté sur les vecteurs menant aux centres de chaque face (en jaune) selon l'équation 5. L'histogramme est calculé à partir de l'amplitude de la projection de chaque points (en noir) par les équations 6 et 7.

Chaque face du polyèdre correspond à une classe de l'histogramme. La projection du vecteur normal sur une face est définie par :

$$p(\vec{n}, f) = \begin{cases} \vec{n} \cdot \vec{n}_f & \text{si } \vec{n} \cdot \vec{n}_f > 0 \\ 0 & \text{sinon} \end{cases} \quad (5)$$

où  $\vec{n}$  est le vecteur normal et  $\vec{n}_f$  est le vecteur allant du centre du polyèdre jusqu'au centre de la face  $f$ . Ensuite l'histogramme associé au bloc  $b$  est calculé par :

$$H_b(f) = \frac{h_b(f)}{\sum_{f'} h_b(f')} \quad (6)$$

avec :

$$h_b(f) = \sum_{\vec{n} \in \mathcal{C}_b} p(\vec{n}, f) \quad (7)$$

où  $\mathcal{C}_b$  est l'ensemble des vecteurs normaux des points du bloc  $b$ . Il existe 5 polyèdres réguliers convexes (ou solides de Platon) : le tétraèdre (4 faces), le cube (ou hexaèdre - 6 faces), l'octaèdre (8 faces), le dodécaèdre (12 faces) et l'icosaèdre (20 faces). Lors de la phase d'évaluation nous avons réalisé des tests avec chacun de ces solides.

L'espace 3D doit être divisé en sous-régions (blocs). Nous utilisons une coupe cylindrique similaire à celle proposée par Gond [8]. Le cylindre qui circonscrit le nuage de points est divisé comme suit :

- tout d'abord une coupe radiale divise le cylindre (voir Figure 4-a)
- puis une coupe azimutale divise le cylindre en secteurs (voir Figure 4-b)
- enfin une coupe axiale divise le cylindre en secteurs (voir Figure 4-c)

Le bloc résultant de ce processus est un secteur en forme de coquillage comme présenté sur la Figure 4-d.

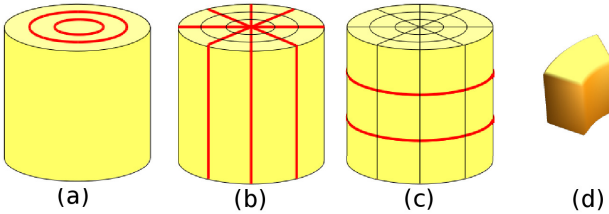


FIGURE 4 – Partage du cylindre en blocs : une coupe radiale (a), une coupe azimutale (b) et une coupe axiale (c). Le volume qui est résulte (d) est un bloc.

La figure 5 illustre ce procédé sur un nuage de points. Un bloc contient un certain nombre de points 3D. Puis l'histogramme des normales orientées est calculé dans chaque bloc selon l'équation 6. Les histogrammes de chaque bloc sont regroupés pour former le descripteur comme suit :

$$D = \{H_1 \cdot H_2 \cdot \dots \cdot H_{N_b}\} \quad (8)$$

où  $N_b$  est le nombre de blocs.

### 2.3 La classification

La méthode proposée fonctionne à partir d'un nuage de points complet enregistré dans un environnement fermé comme il a été mentionné précédemment. À notre connaissance, il n'y a pas de base de données existante qui fournisse de telles données. Nous avons alors décidé

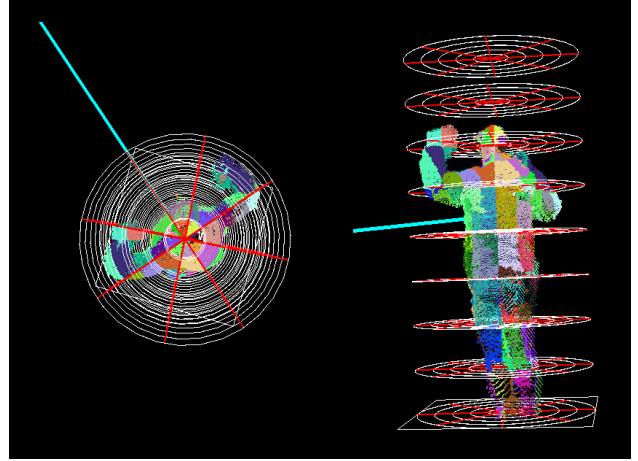


FIGURE 5 – Partage du nuage de points dans les blocs. Chaque bloc est représenté par une couleur différente.

de construire une base de données originale de nuages de points complets. Cette dernière est composée de deux types d'éléments : les positifs (représentant une personne) et les négatifs (représentant des objets). Des exemples d'éléments de cette base de données sont visibles sur la Figure 6. L'ensemble des éléments positifs de la base de données contient 600 nuages de points. Il a été réalisé à partir de 17 personnes différentes avec une variation importante de poses, de vêtements et de morphologies. L'ensemble des éléments négatifs représentant des non-humains contient lui aussi approximativement 600 exemples. Il s'agit d'objets pouvant se trouver dans une scène en intérieur : meuble, empilement de cartons, matériel informatique,... Le sujet (personne ou objet) est placé au centre de la scène lors de la construction de la base de données. L'indication de l'orientation du plan frontal de chaque personne est sauvegardée à chaque enregistrement d'une nouvelle position. Cette information sera utilisée à la fois pour l'apprentissage du classifieur (comme l'orientation est recherchée il faut que tous les exemples positifs de l'apprentissage soient orientés de la même manière) mais aussi en tant que vérité-terrain pour l'étape d'évaluation.

Un classifieur SVM (Support Vector Machine) [1] a été choisi pour construire le modèle de classification. Les SVM sont couramment associés aux descripteurs HOG. L'apprentissage se réalise à partir d'un ensemble d'exemples positifs et négatifs.

Le classifieur peut également déterminer l'orientation du plan frontal de la personne. Pour cela il a recours aux vecteurs d'orientation. Pour chaque nuage de points lors de la phase de test, un vecteur de direction est dans un premier temps désigné arbitrairement. Celui-ci subit ensuite une suite de rotations pour simuler les différentes orientations possibles. Chaque rotation du vecteur de direction induit donc un nouveau descripteur comme présenté sur la Figure





FIGURE 6 – Exemples d’éléments de la base de données de nuages de points complets pour des personnes (en haut) ou des objets (en bas).

7. L’orientation des normales est calculée par rapport au vecteur de direction. L’histogramme de chaque bloc varie donc à chaque rotation de ce vecteur. Les descripteurs calculés pour chaque nuage de points sont alors testés par le classifieur. L’orientation du plan frontal de la personne est finalement donnée par le vecteur de direction du descripteur produisant le score de détection le plus élevé (voir Figure 8).

### 3 Expérimentations et résultats

Dans cette section, nous présentons l’évaluation que nous avons faite de notre méthode. Étant donné qu’il n’y a pas dans la littérature de base de données similaire à celle requise ici, aucune comparaison avec d’autres méthodes ne sera présentée. Par conséquent, nous avons créé 3 ensembles ( $\mathcal{S}_1$ ,  $\mathcal{S}_2$  et  $\mathcal{S}_3$ ) d’exemples afin d’évaluer la classification proposée et d’optimiser les différents paramètres utilisés. Chacun de ces ensembles contient 64 exemples positifs et 64 exemples négatifs. Ces 3 ensembles ont été enre-

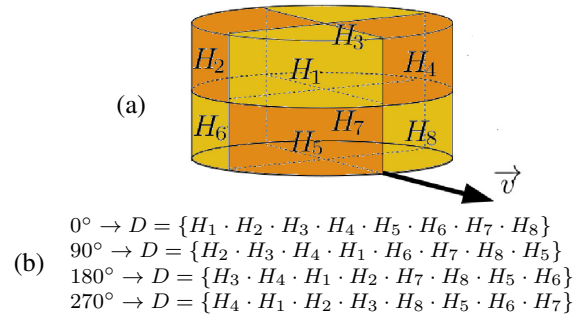


FIGURE 7 – Illustration de la rotation du vecteur de direction. Dans cet exemple le cylindre est découpé en 8 blocs (a). Un histogramme est calculé pour chaque bloc. On désigne un vecteur de direction initiale  $\vec{v}$ , puis on réalise plusieurs rotation de  $\vec{v}$ . Chaque position de  $\vec{v}$  créé un nouveau descripteur (b).

gistrés à partir de la plate-forme multi-kinect comme pour les données d’apprentissage. La classification proposée est appliquée à tous les éléments de chaque ensemble. Le classifieur retourne pour chaque objet un score correspondant à la probabilité qu’il représente une personne.

#### 3.1 Performances de la classification

Il y a plusieurs paramètres utilisés pour le calcul du descripteur. Nous avons réalisé les tests de classification en utilisant diverses combinaisons des valeurs de ces paramètres. La Table 1 liste les valeurs de chacun de ces paramètres donnant le résultat de classification optimal.

Paramètre	Valeur optimale
Hauteur du cylindre	2 m
Rayon du cylindre	0,5 m
Polyèdre	octaèdre
Coupe radiale du cylindre	5 cercles
Coupe azimutale du cylindre	8 secteurs
Coupe axiale du cylindre	8 sections

TABLE 1 – Paramètres du descripteur donnant la meilleure classification.

La figure 9 montre les résultats obtenus en utilisant différents polyèdres et différentes valeurs du rayon du cylindre. À la différence des ensembles  $\mathcal{S}_1$  et  $\mathcal{S}_2$ , l’ensemble  $\mathcal{S}_3$  est constitué d’objets assez simples à analyser. Les résultats obtenus dans cet ensemble sont parfaits (tout les exemples positifs et négatifs sont correctement désignés). Nous ne l’avons donc pas représenté dans la figure 9. Au niveau du rayon du cylindre, on voit qu’une valeur trop faible (0,25m) ou trop grande (1m) rend le système moins robuste. Pour garantir une précision optimale nous choisissons pour la suite un rayon de 0,5m qui donne les meilleurs résultats. Au niveau du polyèdre choisi, les résultats sont moins nets. Le tétraèdre donne des résultats parfaits pour les deux ensembles  $\mathcal{S}_2$  et  $\mathcal{S}_3$ . L’octaèdre donne de très bons résultats

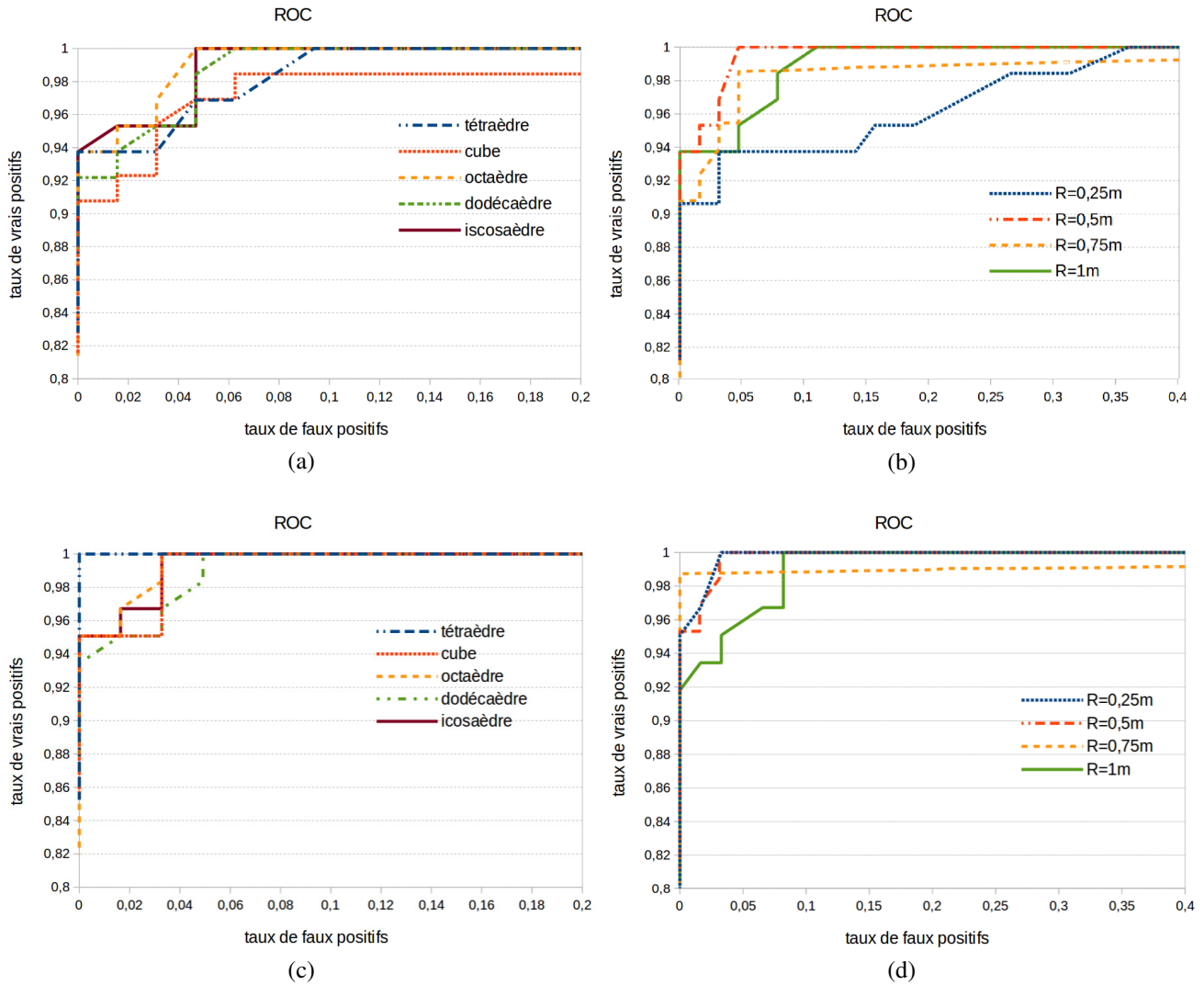


FIGURE 9 – Résultats de la classification sur les ensembles  $\mathcal{S}_1$  ((a) et (b)) et  $\mathcal{S}_2$  ((c) et (d)). Les courbes (a) et (c) sont obtenues en projetant les vecteurs normaux sur différents polyèdres. Les courbes (b) et (d) sont obtenues en faisant varier le rayon du cylindre.

sur les trois ensembles. C’est lui qui donne les meilleurs résultats sur l’ensemble le plus complexe ( $\mathcal{S}_1$ ). Nous le choisissons donc pour la suite.

Pour quantifier notre évaluation nous utilisons les critères de précision, de rappel et la  $F_{measure}$  qui combine les deux critères précédents. Ces critères sont définis par :

$$\left\{ \begin{array}{l} \text{rappel} = \frac{\text{nombre de personnes bien détectées}}{\text{nombre de personnes}} \\ \text{précision} = \frac{\text{nombre de personnes bien détectées}}{\text{nombre de détections}} \\ F_{measure} = \frac{2 \times \text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}} \end{array} \right. \quad (9)$$

Nous obtenons alors les résultats de la Table 2. Ces excellents résultats valident l’efficacité de notre méthode. Le temps de traitement moyen nécessaire pour le calcul du

descripteur à partir d’un code C++ non optimisé tournant sur un processeur 3GHz est d’environ 30ms. C’est un temps assez court en adéquation avec les contraintes des applications temps-réels de surveillance médicale que nous visons.

polyèdre	critère	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$
tétraèdre	précision	0,97	1	1
	rappel	0,95	1	1
	$F_{measure}$	0,96	1	1
octaèdre	précision	0,97	1	1
	rappel	0,97	0,97	1
	$F_{measure}$	0,97	0,98	1

TABLE 2 – Valeurs des critères quantitatifs.

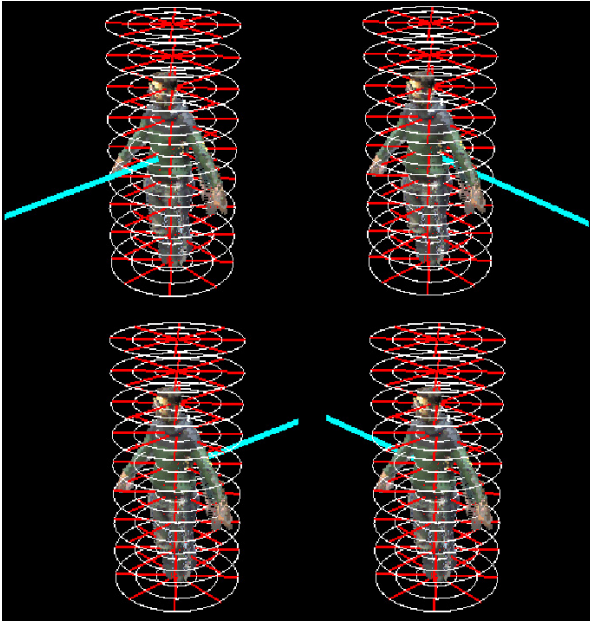


FIGURE 8 – Le vecteur d’orientation (ici en cyan) simule l’estimation de l’orientation du plan frontal de la personne. On réalise plusieurs rotations successives de ce vecteur pour chercher celui dont le score de classification sera le plus élevé.

### 3.2 Estimation de l’orientation

Afin d’évaluer nos résultats sur l’estimation de l’orientation du plan frontal de la personne nous avons réalisé des tests sur les exemples positifs avec différents vecteurs de direction. En pratique nous avons réalisé quatre rotations du vecteur de direction. En effet les rotations intermédiaires ne sont pas significatives en terme d’étude du comportement. À chaque rotation nous calculons le descripteur correspondant. Pour chaque exemple positif de notre base de test nous comparons l’orientation qui donne le plus haut score de classification avec les vérités-terrain. La répartition des valeurs de  $\Delta$ , l’écart entre l’orientation estimée et celle de la vérité terrain, sur nos trois ensembles est résumée dans la Table 3. Dans la majorité des cas l’estimation de l’orientation est correcte. Le plan de dos est parfois reconnu. En effet lorsque la personne a les bras le long du corps, ses formes vues de face et de dos sont très proches.

$\Delta$	$S_1$	$S_2$	$S_3$
0°	90%	74%	86%
90°	2%	10%	7%
180°	8%	16%	7%

TABLE 3 – Résultats de l’estimation de l’orientation du plan frontal. La bonne orientation est bien estimée dans la majorité des cas.

## 4 Conclusion

Dans ce papier nous proposons un nouveau descripteur 3D pour la classification de personnes. Ce descripteur permet également l’estimation de l’orientation du plan frontal de la personne. Nous travaillons avec un nuage de points complet obtenu par un système multi-kinect. Notre descripteur adapte le principe des HOG aux données 3D en remplaçant le gradient par la normale aux surfaces. Afin de pouvoir le valider, nous avons construit une base de données originale. La détection de personnes est exécutée avec une excellente précision et l’orientation du plan frontal est estimée de manière très satisfaisante.

En définitive, nos principales contributions dans ce travail sont :

- la création d’une plate-forme permettant l’acquisition de nuages de points complets et la création d’une base de données liée à la classe des personnes à partir de cette plate-forme
- la définition d’un nouveau descripteur pour la classification des personnes qui modifie les HOG et exploite au mieux les caractéristiques des nuages de points complets
- l’évaluation du descripteur sur la base de données à partir de critères quantitatifs pertinents.

Dans ce papier nous avons restreint notre étude à la classification d’objets isolés. Nos futurs travaux réaliseront la détection complète. C’est à dire qu’au lieu de chercher si une personne est présente dans la scène, il faudra trouver le nombre de personnes présentes et leur localisation dans la scène. L’idée étudiée actuellement consiste à effectuer un balayage spatial de la scène. Une méthode conventionnelle de suppression des non-maxima sera exécutée sur les sorties du classifieur pour obtenir une détection par personne présente.

## Références

- [1] C.C. Chang et C.J. Lin, LIBSVM : a Library for Support Vector Machines, *ACM Transactions on Intelligent Systems and Technology*, Vol. 27, pp. 1-27, 2011.
- [2] B. Choi, C. Mericli, J. Biswas et M. Veloso, Fast Human Detection for Indoor Mobile Robots Using Depth Images, *IEEE International Conference on Robotics and Automation*, pp. 1108-1113, 2013.
- [3] B. Choi, C. Pantofaru et S. Savarese, Detecting and Tracking People using an RGB-D Camera via Multiple Detector Fusion, *IEEE Conference on Computer Vision Workshops*, pp. 6-13, 2011.
- [4] N. Dalal et B. Triggs, Histograms of Oriented Gradients for Human Detection, *IEEE Computer Vision and Pattern Recognition*, Vol. 1, pp. 886-893, 2005.
- [5] J.C. Devaux, H. Hadj-Abdelkader et E. Colle, A Multi-sensor Calibration Toolbox for Kinect : Application to Kinect and Laser Range Finder Fusion, *International Conference on Advanced Robotics*, 2013.



- [6] K. Essmaeel, L. Gallo, E. Damiani, G. De Pietro et A. Dipanda, Multiple Structured Light-Based Depth Sensors for Human Motion Analysis : A Review, *Lecture Notes in Computer Science*, Vol. 7657, pp. 381-395, 2012.
- [7] M.A. Fischler et R.C. Bolles, Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, Vol. 24, pp. 381–395, 1981.
- [8] L. Gond, P. Sayd, T. Chateau et M. Dhome, A 3D Shape Descriptor for Human Pose Recovery, *Lecture Notes in Computer Science*, Vol. 5098, pp. 370-379, 2008.
- [9] D. Herrera, J. Kannala et J. Heikkilä, Accurate and Practical Calibration of a Depth and Color Camera Pair, *Lecture Notes in Computer Science*, Vol. 6855, pp. 437-445, 2011.
- [10] S. Ikemura et H. Fujiyoshi, Real-Time Human Detection Using Relational Depth Similarity Features, *Asian Conference on Computer Vision*, pp. 25-38, 2011.
- [11] A. Kläser, M. Marszalek et C. Schmid, A Spatio-Temporal Descriptor Based on 3D-Gradients, *British Machine Vision Conference*, pp. 1-10, 2008.
- [12] C. Raposo, J.P. Barreto et U. Nunes, Fast and Accurate Calibration of a Kinect Sensor, *IEEE International Conference on 3D Vision*, pp. 342-349, 2013.
- [13] L. Spinello et K.O. Arras, People Detection in RGB-D Data, *IEEE International Conference on Intelligent Robots and Systems*, pp. 3838-3843, 2011.
- [14] Q. Tian, B. Zhou, W. Zhao, Y. Wei et W. Fei, Human Detection using HOG Features of Head and Shoulder Based on Depth Map, *Journal of Software*, Vol. 8, pp. 2223-2230, 2013.
- [15] L. Xia, C. Chen et J.K. Aggarwal, Human Detection using Depth Information by Kinect, *IEEE Computer Vision and Pattern Recognition Workshops*, pp. 15-22, 2011.
- [16] C. Zong, X. Clady et M. Chetouani, An Embedded Human Motion Capture System for an Assistive Walking Robot, *IEEE International Conference on Rehabilitation Robotics*, pp. 1-6, 2011.