

DÉTECTION DE TÊTES DANS UN NUAGE DE POINTS 3D À L'AIDE D'UN MODÈLE DE MÉLANGE SPHÉRIQUE

Denis Brazey ¹ & Bruno Portier ²

¹ *Société Pryncl, RD974 21190 Corpeau, France*
denis.brazey@insa-rouen.fr

² *Normandie Université, INSA de Rouen, LMI - EA 3226 685 Avenue de l'Université - BP 08 76801 St Etienne du Rouvray cedex, France*
bruno.portier@insa-rouen.fr

Résumé. L'objectif de ce travail est la détection de têtes dans un nuage de points 3D représentant un groupe de personnes. L'emploi d'un seul capteur conduit à une grande quantité de données manquantes puisque seule une face des objets est visible. Notre méthode est basée sur un nouveau modèle de mélange sphérique. La forme des objets est considérée de part la distribution statistique des points dans l'espace. La densité utilisée appartient à la famille des distributions elliptiques et modélise des points repartis autour d'une surface sphérique. Les paramètres inconnus sont estimés par un algorithme Espérance-Maximisation. Le nombre de composantes est déterminé par une approche ascendante et les têtes sont détectées par des règles heuristiques. L'application de la méthode sur des jeux de données réels permet de modéliser et de détecter correctement les têtes.

Mots-clés. Détection de tête, Modèle de mélange, Algorithme EM

Abstract. The objective of this work is the detection of heads in a 3D point cloud representing a group of people. The use of a single sensor implies a lot of missing data since only one side of objects is visible. Our method is based on a new spherical mixture model. The shape of objects is considered through the statistical distribution of points in 3D space. The probability density function models points spread around a spherical surface and is elliptically distributed. Unknown parameters are estimated with the Expectation-Maximization algorithm. The number of components is determined by a step by step ascending strategy and heads are detected with heuristic rules. The use of the method on real data sets enables to correctly detect and model heads.

Keywords. Head detection, Mixture model, EM algorithm

1 Introduction

La détection de têtes est une problématique centrale pour de nombreuses applications de vision par ordinateur comme par exemple la biométrie, les interfaces homme-machine, la réalité augmentée et la vidéo conférence.

Les capteurs 3D fournissent une information de distance pour chacun des pixels de l'image. Cette information peut être convertie en un nuage de points 3D représentant la scène. Récemment, le capteur Kinect commercialisé avec la plateforme de jeu Microsoft Xbox a permis de démocratiser l'utilisation de tels capteurs. Le système fournit une information géométrique en plus de l'information colorimétrique. Dans ce travail, seule la composante de distance est considérée lors des traitements.

L'acquisition de données 3D permet d'analyser la forme des objets. Le problème d'ajustement d'ellipsoïdes sur des têtes est présenté par Grammalidis et Strintzis (2000). Un algorithme d'extraction des maxima locaux a été proposé par Zhang, Yan et al (2012) pour segmenter les têtes. Les travaux de Zhang, He et al (2012) se basent sur un ensemble de zones positives et négatives afin de déterminer la position des têtes. Notons que dans la plupart des approches, la forme ellipsoïdale de la tête a été exploitée.

Les modèles de mélange, généralement gaussiens, ont été beaucoup utilisés en statistiques pour modéliser et classifier des données (voir par exemple le livre de McLachlan et Peel (2000)). Ils ont été appliqués à des problèmes variés dans le domaine du traitement d'images comme la soustraction de fond (Stauffer, 1999), la segmentation d'images (Zhang, He et al, 2012) ou la modélisation d'environnements 3D (Núñez et al, 2009).

Dans ce travail, nous considérons le problème de détection de têtes dans un nuage de points 3D représentant un groupe de personnes. Pour cela nous nous appuyons sur un modèle de mélange sphérique introduit par Brazey et Portier (2014) dont la densité de probabilité permet de modéliser des points répartis autour d'une surface sphérique. Les images étant acquises avec un seul capteur 3D, une grande partie des données sont manquantes et les observations sont non-uniformément réparties sur la surface des objets.

2 Le modèle de mélange sphérique

Un modèle de mélange fini est un modèle probabiliste utilisé pour estimer paramétriquement la distribution d'une variable aléatoire en la modélisant par une somme pondérée de K distributions. La densité de probabilité h du mélange est une combinaison convexe de K densités. Généralement, ces densités appartiennent à une famille de densités paramétriques et chaque composante se différencie par la valeur de ses paramètres. La densité du mélange est de la forme

$$h(x|\Theta) = \sum_{k=1}^K \pi_k f(x|\theta_k) \quad (1)$$

où les (π_k) sont les proportions du mélange satisfaisant $\pi_k \geq 0$ et $\sum_{k=1}^K \pi_k = 1$ et $\Theta = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$ est le vecteur des paramètres inconnus du modèle. La particularité de ce type de modélisation est qu'elle permet d'expliquer les données par un ensemble de composantes pouvant se chevaucher. Dans notre application, nous choisissons d'utiliser un

tel modèle dans lequel chaque composante est une primitive géométrique. Les observations retournées par le capteur étant localisées sur la surface des objets, il est clair que la distribution gaussienne n'est pas adaptée. Puisque la forme de la tête peut s'apparenter à celle d'une sphère, nous proposons d'utiliser la densité introduite par Brazey et Portier (2014) modélisant des points répartis autour d'une surface sphérique de centre $\mu \in \mathbb{R}^3$, de rayon $r > 0$ et de variance $\sigma^2 > 0$. La densité f décrivant la $k^{\text{ème}}$ composante du mélange s'écrit

$$f(x, \theta_k) = C \exp \left(-\frac{1}{2\sigma_k^2} (\|x - \mu_k\| - r_k)^2 \right) \quad (2)$$

où C est une constante de normalisation et $\theta_k = (\mu_k, r_k, \sigma_k)$ le vecteur des paramètres de la composante k . Sous l'hypothèse que $r_k \gg \sigma_k$, la constante C peut être approchée par

$$C = \frac{1}{2(2\pi)^{3/2} \sigma_k (r_k^2 + \sigma_k^2)} \quad (3)$$

L'hypothèse introduite ci-dessus est raisonnable dans notre cadre d'application puisque le rayon d'une tête est grand comparé à la variance des points autour de la surface. Notons enfin que cette distribution appartient à la famille des distributions elliptiques. On trouvera en Figure 1 un échantillon de points simulés selon la densité f .

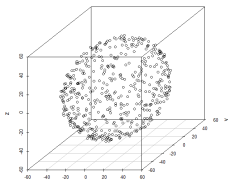


Figure 1: Un échantillon de 1000 réalisations de densité f de paramètres $((0, 0, 0)^T, 50, 1)$.

3 Estimation des paramètres

On s'intéresse dans cette partie à l'estimation du vecteur de paramètres Θ du modèle de mélange sphérique (1), à partir de n observations $\mathbf{x} = (x_1, \dots, x_n)$. On suppose que chaque observation $x_i \in \mathbb{R}^3$ provient de l'une des K composantes du modèle. Pour estimer Θ , nous choisissons l'estimateur du maximum de vraisemblance $\hat{\Theta}$ défini par

$$\hat{\Theta} = \arg \max_{\Theta} L(\mathbf{x}; \Theta) \quad (4)$$

où L désigne la fonction de log-vraisemblance donnée par

$$L(\mathbf{x}; \Theta) = \sum_{i=1}^n \log \left(\sum_{k=1}^K \pi_k f(x_i | \theta_k) \right) \quad (5)$$

L'algorithme EM (voir par exemple McLachlan et Thriyambakam (2008)) est la méthode la plus utilisée pour maximiser la log-vraisemblance dans le cas des modèles de mélange. L'algorithme consiste en deux étapes E et M itérées jusqu'à convergence.

Étape E : Calcul des probabilités d'appartenance $t_{i\ell}$ des points $\{x_i\}_{i=1}^n$ aux classes $\{C_l\}_{l=1}^K$ avec l'estimation courante des paramètres $\Theta^{(j)} = (\pi_1^{(j)}, \dots, \pi_K^{(j)}, \theta_1^{(j)}, \dots, \theta_K^{(j)})$

$$t_{i\ell} = \frac{\pi_\ell^{(j)} f(x_i | \theta_\ell^{(j)})}{\sum_{k=1}^K \pi_k^{(j)} f(x_i | \theta_k^{(j)})} \quad (6)$$

Étape M : Déterminer $\Theta^{(j+1)}$ maximisant l'espérance conditionnelle de la log-vraisemblance complétée étant donné l'observation \mathbf{x} et l'estimation courante $\Theta^{(j)}$. Pour $\ell = 1, \dots, K$,

$$\left\{ \begin{array}{l} \hat{\mu}_\ell = \frac{1}{\sum_{i=1}^n t_{i\ell}} \left(\sum_{i=1}^n t_{i\ell} x_i - \hat{r}_\ell \sum_{i=1}^n t_{i\ell} \frac{(x_i - \hat{\mu}_\ell)}{\|x_i - \hat{\mu}_\ell\|} \right) \\ \hat{r}_\ell = \frac{1}{\sum_{i=1}^n t_{i\ell}} \sum_{i=1}^n t_{i\ell} \|x_i - \hat{\mu}_\ell\| \\ \hat{\sigma}_\ell^2 = \frac{1}{\sum_{i=1}^n t_{i\ell}} \sum_{i=1}^n (\|x_i - \hat{\mu}_\ell\| - \hat{r}_\ell)^2 \\ \hat{\pi}_\ell = \frac{1}{n} \sum_{i=1}^n t_{i\ell} \end{array} \right. \quad (7)$$

Le processus est initialisé par un vecteur de paramètres $\Theta^{(0)}$ obtenu par un algorithme K-Means. Les estimateurs (7) ont été obtenus par dérivation de la log-vraisemblance complétée par rapport à chacun des paramètres. Les estimations du centre μ_ℓ et du rayon r_ℓ ne sont pas directement calculables puisqu'elles dépendent l'une de l'autre. Nous proposons donc d'employer une méthode de backfitting. Le centre est d'abord estimé par $\hat{\mu}_\ell^{(0)} = n^{-1} \sum_{i=1}^n x_i$ et le rayon par $\hat{r}_\ell^{(0)} = n^{-1} \sum_{i=1}^n \|x_i - \hat{\mu}_\ell^{(0)}\|$. Ces estimations sont ensuite améliorées en réévaluant récursivement les équations (7). Les équations des estimations sont donc évaluées plusieurs fois à chaque étape M. Ce processus itératif est stoppé lorsque la différence entre deux estimations successives est suffisamment faible (seuils ϵ_μ et ϵ_r).

Notons enfin que l'estimateur $\mu_\ell^{(1)}$ est particulièrement bien adapté à l'estimation de μ_ℓ lorsque les données ne sont pas uniformément réparties sur la surface (demi-sphère par exemple). La performance des estimateurs sur des données simulées et l'étude de leur convergence sont discutées par Brazey et Portier (2014).

4 Application à la détection de têtes

Dans cette partie, nous utilisons le modèle de mélange sphérique décrit en section 2 pour détecter et modéliser des têtes dans un nuage de points 3D représentant un groupe de personnes. Les images proviennent d'un seul capteur 3D dont la hauteur et l'orientation par rapport au sol sont supposées connues. Les images ont été segmentées par un seuillage des valeurs de profondeur. Les nuages de points sont non-organisés et la structure 2D des images n'est pas prise en compte.

Les expérimentations ont montré que les valeurs initiales des variances ont un impact important sur les résultats. Si la variance initiale est élevée, des observations appartenant à d'autres composantes participeront à l'estimation des paramètres. Par conséquent, nous préférons fixer les variances plutôt que de les ré-estimer à chaque itération de EM.

Le nombre de composantes K du modèle n'est pas connu a priori et doit être déterminé. S'il est choisi trop faible, le modèle ne sera pas assez complexe pour modéliser correctement les données. Si au contraire il est trop élevé, les données seront sur-segmentées. Nous adoptons donc une approche ascendante. Le paramètre K est initialisé à K_{min} et incrémenté jusqu'à K_{max} . Une incrémentation est validée si la log-vraisemblance a été améliorée d'un coefficient au moins égal à β . Sinon la valeur de K est considérée comme optimale.

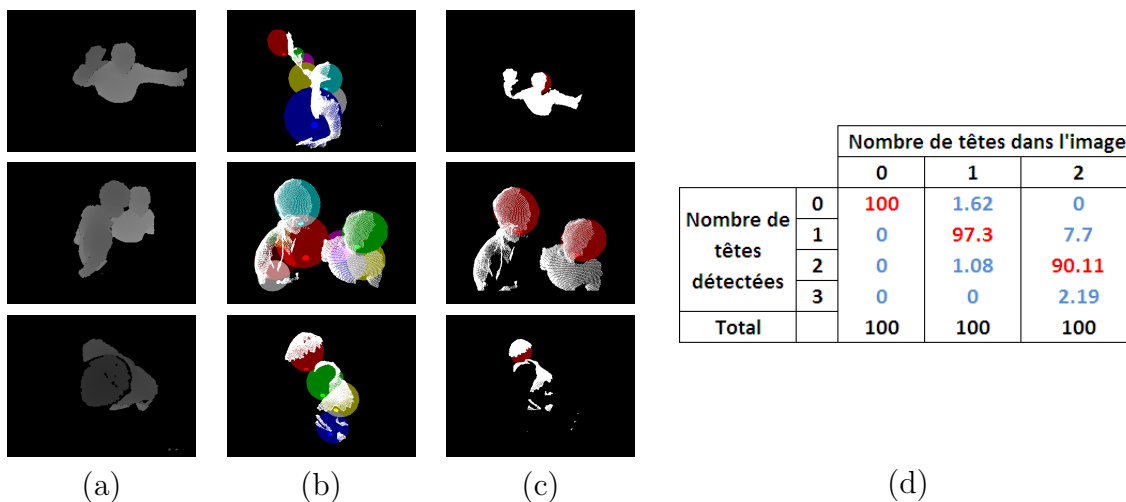


Figure 2: (a) images de profondeur, (b) nuages de points et modèles, (c) têtes détectées et (d) pourcentages de détection pour chaque nombre de têtes dans l'image.

Après l'estimation du modèle, nous analysons les sphères obtenues pour décider lesquelles représentent des têtes. Nous définissons pour cela des règles heuristiques. Dans un premier temps, une sphère est considérée comme une *tête potentielle* si son rayon appartient à l'intervalle $[r_{min}, r_{max}]$ et sa distance au sol supérieure à z_{min} . Nous prenons également

en compte la classification obtenue via EM. Chaque observation x_i est associée à la composante la plus probable C_k définie par $k = \arg \max_l \{t_{il} \geq 0.5\}$. La distance au carré moyenne entre la surface d'une sphère et les points y étant associés doit être inférieure à ϵ_{erreur} . Dans un second temps, une tête potentielle est classée comme *tête* si aucune autre tête n'est située au dessus et si au moins une composante représentant le corps est située en dessous.

La méthode est appliquée sur un ensemble de 1430 images. Les résultats obtenus sont représentés Figure 2. Le processus de backfitting est limité par $\epsilon_\mu = \epsilon_r = 0.5$ et un maximum de 50 itérations. L'algorithme EM est limité à 10 itérations. Les paramètres K_{min} et K_{max} ont été respectivement fixés à 2 et 15, le seuil β à 0.01 et les σ_k à 5. Les critères de détection ont été fixés à $r_{min} = 80$, $r_{max} = 120$, $z_{min} = 1000$ mm et $\epsilon_{erreur} = 300$. Les surfaces des composantes optimisées sont proches des observations malgré le grand nombre de données manquantes et la répartition inégale des points. On constate que les sphères convergent vers les zones du nuage présentant une forme sphérique comme la tête et les épaules. Les règles de détection proposées permettent de détecter et modéliser correctement les têtes dans plus de 96 % des cas. Le temps de traitement est de l'ordre plusieurs secondes par images. Une amélioration de la méthode consisterait à mettre en place une accélération de la méthode d'estimation (McLachlan et Thriyambakam, 2008) et à complexifier les règles de détection.

Bibliographie

- [1] D. Brazey et B. Portier (2014), A new spherical mixture model for head detection in depth images, Submitted.
- [2] N. Grammalidis et M. G. Strintzis (2000), Head Detection and Tracking by 2-D and 3-D Ellipsoid Fitting, Proc. CGI, 221-226.
- [3] G. J. McLachlan et D. Peel (2000), Finite Mixture Models, Wiley Series in Probability and Statistics.
- [4] G. J. McLachlan et K. Thriyambakam (2008), The EM Algorithm and Extensions, WILEY-Interscience.
- [5] P. Núñez, P. Drews Jr, R. Rocha, M. Campos et J. Dias (2009), Novelty detection and 3D shape retrieval based on gaussian mixture models for autonomous surveillance robotics, IROS, 4724-4730.
- [6] T. van Oosterhout, S. Bakkes et B. Kröse (2011), Head Detection in Stereo Data for People Counting and Segmentation, VISAPP, 620-625.
- [7] C. Stauffer (1999), Adaptive background mixture models for real-time tracking, CVPR.
- [8] X. Zhang, R. He, M. Yao et F. Zhu (2012), The Elliptical Contoured Mixture Model for Image Segmentation, Journal of Computational Information Systems, 7847-7855.
- [9] X. Zhang, J. Yan, S. Feng, Z. Lei, D. Yi et S.Z. Li (2012), Water Filling : Unsupervised People Counting via Vertical Kinect Sensor, AVSS.