De l'intérêt de la texture pour la segmentation du visage.

On the interest of texture for face segmentation.

P. Julian[1,2] V. Charvillat[1] C. Dehais[2] F. Lauze[3]

[1] IRIT
[2] FittingBox
[3] DIKU

La pyrénéenne, Labège, France
pauline.julian@fittingbox.com

Résumé

Dans cet article, nous évaluons l'intérêt d'utiliser la texture pour segmenter une image de visage en zones Peau, Fond et Cheveux. Tout d'abord, nous présentons des techniques existantes de détection de peau, nous justifions le choix a priori de la texture, ensuite nous construisons des indices caractéristiques locaux à partir de données apprises sur l'image. Nous définissons un modèle de classification s'appuyant conjointement sur les données couleur et texture. Nous montrons l'apport de cette approche en présentant quelques résultats prometteurs obtenus sur des images naturelles complexes.

Mots Clef

Segmentation, classification bayésienne, couleur, texture.

Abstract

In this paper, we assess the value of using texture to segment an image into skin, background and hair areas. First, we present existing techniques for skin detection, we justify our a priori choice of integrating texture, and then we build local descriptors learned directly from the image data. We define a classification model based jointly on both color data and texture. We demonstrate the viability of this approach by presenting some promising results obtained on complex natural images.

Keywords

Segmentation, bayesian classification, color, texture.

1 Introduction

Nous évaluons des méthodes qui permettent de segmenter une image de type portrait de face. Plus précisément, nous cherchons à classifier le plus correctement possible les pixels correspondant à des éléments de Peau, de Fond et de Cheveux, pour obtenir une segmentation similaire à celle représentée sur la figure 1. Nous utilisons cette segmentation comme un des outils d'analyse du visage de l'utilisateur dans nos logiciels de réalité augmentée [http ://www.enjoysunglasses.com]. Par exemple la zone Peau, une région sur laquelle nous pourrons incruster des lunettes virtuelles. Le remplacement de la zone Fond permet d'immerger la personne dans un nouvel environnement.

On peut facilement admettre que la segmentation est beaucoup plus aisée à réaliser lorsqu'il existe un fort contraste entre les zones Peau, Fond et Cheveux, et se révèle beaucoup plus délicate dans certaines configurations, en particulier lorsque des cheveux blonds se mêlent à une peau de couleur claire, ou bien dans des cas où l'arrière plan, qui n'est pas maîtrisé, est d'une couleur similaire aux Cheveux ou à la Peau. Notre évaluation est volontairement focalisée sur ces cas difficiles.

La littérature concernant la détection de Peau dans une image est abondante [1] [2]. Classiquement, elle est abordée par des approches statistiques. Un pixel de couleur c est classé comme appartenant ou non à la Peau selon la probabilité *a posteriori* $p(peau/c)$ déduite d'une distribution $p(c/peau)$ de couleur caractéristique. Il est connu, et nos résultats [3] confirment cette affirmation, que dans les cas délicats décrits précédemment, la couleur ne suffit plus. En revanche, il semble que la Peau et les Cheveux contiennent des textures répétitives et peu semblables. Notre étude a pour but de valider cette intuition. Nous constatons que la Peau présente des caractéristiques basses fréquences tandis que la région Cheveux a plutôt des caractéristiques hautes fréquences. La région Fond est plus délicate à caractériser *a priori* de part la diversité des images traitées. Nous faisons alors deux hypothèses. Premièrement, nous pouvons supposer que le fond présente des caractéristiques de couleur et de texture discriminante vis à vis des autres classes (au sens du travail de [4]) . Deuxièmement, on peut supposer que le fond peut être segmenté par une composition [5]. Dans cette approche, nous supposons que la région

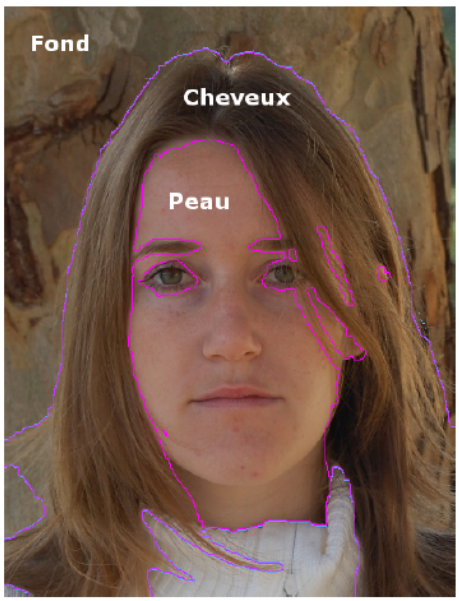


FIG. 1 – Classification des pixels de l'image dans les classes Peau, Fond et Cheveux.

Fond peut être (re)composée à l'aide d'un dictionnaire de quelques parties élémentaires classées *a priori* comme faisant partie du Fond. Les deux travaux précédemment cités sont des aiguillons pour notre travail, nous utiliserons leur concept pour notre étude. Notons que nous pouvons faire la même analyse pour les deux autres classes.
Pour construire les indices caractéristiques locaux spécifique à chaque région, nous introduisons une phase d'apprentissage supervisée dans notre algorithme. Nous collectons alors des échantillons sur des zones étiquetées *a priori*.
L'article est organisé comme suit. Dans la section 2, nous introduisons les notations du problème et nous définissons les paramètres à prendre en compte pour la phase d'apprentissage. Dans la section suivante, nous modélisons nos données. Dans la section 4, nous traitons la phase de décision de notre algorithme. Finalement, nous interprétons les résultats obtenus, avant de conclure.

2 Données de travail

Dans cette section, nous définissons les principes et posons les notations que nous allons employer. Nous travaillons dans l'espace de couleurs normalisé pour la caractérisation colorimétrique et en niveau de gris pour la caractérisation de texture. Notre travail repose sur trois étapes : supervision, classification, validation. La supervision introduit les connaissances *a priori* sur chaque région. Ces connaissances sont utilisées pour la classification, laquelle subit une validation grâce à une vérité terrain.

2.1 Notations

Nous notons $Ic(x) = (r, v, b) = (\frac{R}{R+V+B}, \frac{V}{R+V+B}, \frac{B}{R+V+B})$ la couleur du pixel de l'image dans l'espace RVB normalisé. Ce dernier apporte une indépendance linéaire au changement d'illumination. Notons $I_g(x)$ le niveau de gris du pixel x.
Nous notons R^0_{peau}, R^0_{fond} et $R^0_{cheveux}$ les régions d'apprentissage spécifiées *a priori* comme étant respectivement de la Peau, du Fond, des Cheveux (cf. figure 2.1).

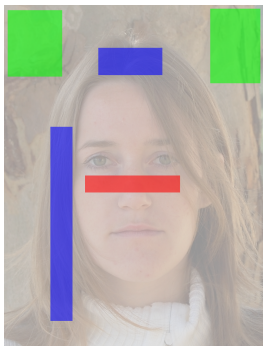


FIG. 2 – Affichage des régions R^0_{peau}, R^0_{fond} et $R^0_{cheveux}$ en rouge, vert et bleu.

Segmenter une image revient à obtenir une carte C caractérisant l'appartenance de chaque pixel à une classe. Ainsi, on a $C(x) \in \{w_1, w_2, w_3, w_4\} = \{\text{Peau, Fond, Cheveux, Indécis}\}$. Indécis est une classe de rejet que nous définirons par la suite.
La validation revient à confronter C avec une vérité terrain obtenue manuellement.

3 Notre modélisation

La sélection des régions Peau, Fond et Cheveux nous permet d'acquérir des connaissances *a priori*, desquelles nous extrayons des indices caractéristiques de couleur et de texture.

3.1 Mise en forme des données

Puisque $b = 1 - r - v$, nous travaillons uniquement sur les deux premières composantes de Ic : (r, v).

$$X_c = \begin{bmatrix} r \\ v \end{bmatrix} \quad (1)$$

Nous omettons la dépendance au pixel x pour des raisons de lisibilité. La mise en forme des données de texture de l'image, notées X_t, est plus complexe que pour la couleur. Nous définissons pour chacune des classes les caractéristiques de texture à l'échelle d'un échantillon de taille lxl, l valant typiquement 5. Pour apprendre la texture caractéristique de chacune des classes, nous construisons une base de vecteurs de taille l^2 permettant la reconstruction de chaque échantillon extrait de l'image en niveau de gris. Plus précisément, nous travaillons avec l'image Ig. Nous appliquons alors le principe de la segmentation par composition [5]. Un échantillon de Peau (respectivement Fond, Cheveux) est naturellement plus vite recomposé/reconstruit en utilisant la base apprise sur la Peau (respectivement Fond, Cheveux). L'indice de texture que nous retenons est déduit de

la "vitesse d'approximation" (cf. figure 4) d'un échantillon pour chacune des bases mises en concurrences. Les régions R^0_{peau}, R^0_{fond} et $R^0_{cheveux}$, nous permettent de définir , pour chacune des trois classes, des ensembles d'échantillons de taille l^2. Nous construisons pour chacun des ensembles, une base en utilisant la méthode d'Analyse en Composantes Principales (ACP) [6]. Nous obtenons alors trois bases Peau, Fond et Cheveux que nous notons respectivement B^p, B^f et B^c (cf. équation (3) et figure 3), dont les vecteurs sont ordonnés par ordre décroissant des valeurs propres.

$$\begin{aligned} B^p &= (\Phi^p_1, \Phi^p_2, ..., \Phi^p_{l^2}) \\ B^f &= (\Phi^f_1, \Phi^f_2, ..., \Phi^f_{l^2}) \\ B^c &= (\Phi^c_1, \Phi^c_2, ..., \Phi^c_{l^2}) \end{aligned} \quad (2)$$

De plus, on note $\bar{b}^p$, $\bar{b}^f$, $\bar{b}^c$ les moyennes des régions R^0_{peau}, R^0_{fond} et $R^0_{cheveux}$.

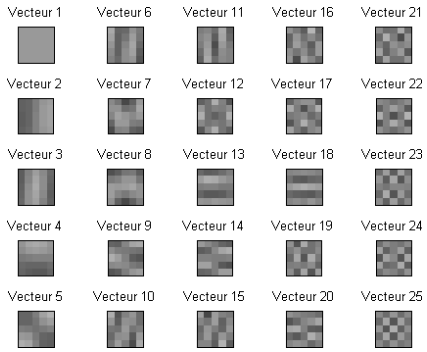


FIG. 3 – Vecteurs propres de la base Cheveux

Chacunes des trois bases permet de reconstruire la totalité des échantillons lxl issus de l'image, avec pour différence que les premiers vecteurs de la base B^p (respectivement B^f et B^c) favorisent la reconstruction des pixels Peau (respectivement Fond et Cheveux). Nous illustrons ces résultats en présentant le cas d'une reconstruction effectuée sur un échantillon de Cheveux (cf. figure 3).

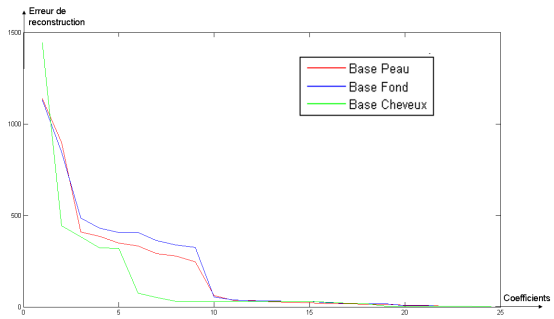


FIG. 4 – Vitesse d'approximation pour un échantillon de Cheveux pour les trois bases Peau Fond et Cheveux.

Sur la figure 4 nous avons en abscisse le nombre de coefficients utilisés et en ordonné l'erreur de reconstruction. Les courbes rouge, bleue et verte utilisent respectivement les bases B^p, B^f et B^c. Le positionnement des courbes montre que la reconstruction de la base Cheveux est la plus rapide pour l'échantillon de cheveux considéré.
Nous choisissons de caractériser la texture de l'image en calculant l'erreur de reconstruction [4] à partir des k vecteurs les plus représentatifs, au sens des valeurs propres obtenues par ACP, de chacune des bases Peau, Fond et Cheveux. Nous générons trois cartes d'erreur que nous notons e^p, e^f et e^c. Ainsi les échantillons "reconstruit jusqu'à k" pour chaque classe sont obtenus par :

$$\begin{aligned} \widehat{b}^p &= \bar{b}^p + \sum_{i=1}^{k<l^2} (b|\Phi^p_i)\ \Phi^p_i, \\ \widehat{b}^f &= \bar{b}^f + \sum_{i=1}^{k<l^2} (b|\Phi^f_i)\ \Phi^f_i, \\ \widehat{b}^c &= \bar{b}^c + \sum_{i=1}^{k<l^2} (b|\Phi^c_i)\ \Phi^c_i. \end{aligned} \quad (3)$$

Les erreurs de reconstructions d'un échantillon b selon chacunes des bases sont définies comme le carré à la distance euclidienne entre b et ses reconstructions :

$$\begin{aligned} e^p(b) &= ||b - \widehat{b}^p||^2, \\ e^f(b) &= ||b - \widehat{b}^f||^2, \\ e^c(b) &= ||b - \widehat{b}^c||^2. \end{aligned} \quad (4)$$

Le vecteur caractéristique de texture est alors défini par :

$$X_t = \frac{1}{\lambda^2} \begin{bmatrix} e^p(b) \\ e^f(b) \\ e^c(b) \end{bmatrix} \text{ Avec } \lambda = 255 \times l \text{ et } X_t \in [0,1]^3 . \quad (5)$$

3.2 Combinaison des données

Après avoir modélisé les données couleur et texture, nous combinons nos résultats en construisant une image vectorielle, notée X, contenant la concaténation de X_c et X_t :

$$X = \begin{bmatrix} X_c \\ X_t \end{bmatrix} \qquad X \in R^5. \quad (6)$$

Nous apprenons sur X, à partir des coordonnées des régions R^0_{peau}, R^0_{fond} et $R^0_{cheveux}$ les fonctions densité de probabilité gaussienne que nous notons $G^p(m_p, \Sigma_p)$, $G^f(m_f, \Sigma_f)$, $G^c(m_f, \Sigma_f)$ pour chaque classe Peau, Fond et Cheveux.
Nos expériences montrent que les distributions des vecteurs caractéristiques X_c et/ou X_t sont suffisamment unimodales pour les supposer gausiennes dans notre travail

d'évaluation. Des lois statistiques plus sophistiquées (Modélisation par mélange de gausiennes, χ^2, par exemple) pourraient aisément s'intégrer dans notre modélisation.
Nous avons désormais un vecteur caractéristique de dimension cinq et trois distributions (trois termes de vraisemblance : $p(X|w_i)$ avec w_i la classe Peau, Fond ou Cheveux).

4 Modélisation de la texture par le Tenseur de Strucure

Afin d'apprécier l'apport de notre modélisation de la texture, nous comparons à une modélisation reposant sur le tenseur de structure multi-échelles. Nous avons déjà utilisé une telle représentation dans un article précédent [3]. La texture est caractérisée par $log(||TS(\rho)||_F)$. Cette dernière modélisation de part son caractère multi-échelle est riche en information mais peu apparaitre couteuse en calcul. Le tenseur de structure à une échelle donnée ρ est définie par :

$$TS(\rho) = K_\rho * \left(\nabla I \, \nabla I^T\right) = K_\rho * \begin{pmatrix} Ig_x^2 & I_x Ig_y \\ Ig_x Ig_y & Ig_y^2 \end{pmatrix} \quad (7)$$

où K_ρ est un noyau gaussien d'écart-type ρ, Ig_k le grandient dans la direction k et $||\ ||_F$ la norme de Frobenius.
Nous vérifions sur le figure 5 que les distributions de $log(||TS(\rho)||_F)$ sont approchables par une gaussienne.

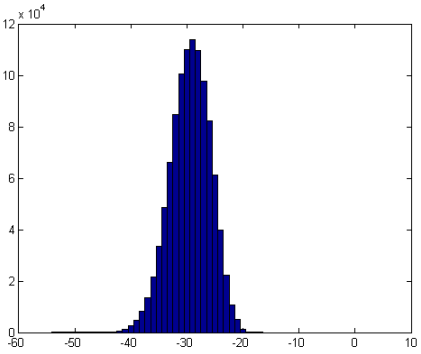


FIG. 5 – Distribution d'une composante $log(||TS(\mathrm{i})||_F)$

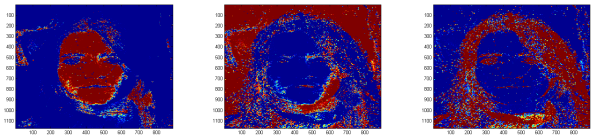
Le vecteur caractéristique en chaque pixel est obtenu en concaténant la couleur X_c et les caractéristiques de textures aux échelles $\rho_1, \rho_2 ... \rho_N$.

$$X = \begin{bmatrix} X_c \\ log(||TS(1)||_F) \\ log(||TS(2)||_F) \\ ... \\ log(||TS(10)||_F) \end{bmatrix}.$$

Plus de détails peuvent être trouvé dans [3].

5 Décision bayésienne

Dans ce travail, nous n'avons pas étudié l'influence des probabilités *a priori* qui sont choisies équiprobablement.



(a) Peau (b) Fond (c) Cheveux

FIG. 6 – Cartes de probabilité

Selon la décision bayésienne (cf. équation (8)), nous sommes en mesure de générer trois cartes (cf. figure 6), représentant la probabilité *a posteriori* de chaque pixel d'appartenir à de la Peau, du Fond et des Cheveux.

$$P(w_i|X) = \frac{p(X|w_i)\ P(w_i)}{\sum_i p(X|w_i) P(w_i)} \quad (8)$$

Avec $\begin{cases} p(X|w_i) = G^i(m_i, \Sigma_i) \\ P(w_i) \text{ la probabilité } \textit{a priori} \text{ et} \\ \mathrm{w}_i \in \{\text{Peau, Fond, Cheveux}\} \end{cases}$

$p(X|w_i)$ est la probabilité conditionnelle apprise comme expliqué dans la section 3.2. Les probabilité *a priori* $P(w_i)$ sont choisies équiprobables. Il serait intéressant d'étudier comment déterminer ces probabilité plus finement.
La classification repose sur le calcul des probabilité *a posteriori* d'appartenance à chacune des classes. Applicant la loi de Bayes, nous obtenons la figure 8. nous pouvons définie la carte C (cf. figure 8(c)) comme étant le maximum de vraissemblance $C(x) = \underset{w_i}{\text{argmax}} P(w_i/X)$.
Cependant, devant notre contexte applicatif, et puisque la méthode présentée ici sert d'étape d'initialisation aux méthodes variationnelles, nous cherchons à prendre des décisions plutôt conservatives et nous introduisons pour cela une classe de rejet. Pour évaluer ce compromis, nous utilisons une vérité terrain (cf. figure 1) et des courbes Rappel/Précision (cf. figure 7) calculées lorsque S varie, S étant le seuil de décision.
Notre règle de décision s'écrit alors :

$$w^* = \underset{w_1, w_2, w_3}{\text{argmax}} \left(P(w_i|x)\right)$$

$$\text{si } P(w^*|x) > S \text{ alors décider } w^*$$
$$\text{sinon décider } w_4 \text{ qui est la classe Indécis.}$$

A partir d'un ensemble vérité terrain , d'images segmentées manuellement, on définit la précision comme le nombre de pixels pertinents retrouvés rapporté au nombre de pixels total proposé pour la classe. Le rappel est défini par le nombre de pixels pertinents retrouvés au regard du nombre total de pixels pertinents. Ces courbes permettent d'obtenir une valeur pertinente pour le seuil S.

6 Résultats et interprétation

Dans cette partie, nous évaluons la qualité de la segmentation obtenue en utilisant la couleur et la texture locale, autrement dit des échantillons de texture. En comparant les résultats avec une approche uniquement couleur, ou pour une approche utilisant la couleur et le tenseur de structure multi-échelles (voir section 4). Nous présentons 3 de nos images de test. Nous rappelons que nous travaillons avec des images dont la couleur Peau, Fond et Cheveux n'est pas discriminante.

Sur les figures 7(b), 7(d), 7(f) nous représentons pour chaque configuration l'influence du seuil S sur le Rappel et la Précision. Au bilan, on choisit $S = 0.9$ car on favorise la qualité de détection à la quantité de détection.

Sur la figure 7(b), nous pouvons constater que la Peau est la classe la mieux détectée, puis vient le Fond et enfin les Cheveux. La texture améliore considérablement les résultats pour les classes Peau et Cheveux. En revanche, pour la Peau, le tenseur de structure est plus discriminant tandis que pour les cheveux l'approche par échantillons donne de meilleuts résultats.

Sur la figure 7(d), d'un point de vue général, le Fond est mieux détecté suivi de la Peau et enfin des Cheveux. La couleur donne de meilleurs résultats pour la Peau tandis que pour les autres classes la texture est plus adaptée. Ici, le tenseur de structure et les échantillons de texture donnent des résultats similaires.

Ainsi ces résultats confirment qu'une approche intégrant couleur et texture est plus efficace qu'une approche uniquement couleur pour la segmentation.

La figure 8 représente les classifications obtenues par les différentes méthodes. La première colonne correspond à l'image à traiter ; la deuxième colonne montre la classification obtenue à l'aide des données couleurs ; la troisième colonne permet de voir la classification obtenue à l'aide des données couleurs et des échantillons de texture ; enfin la quatrième colonne présente la classification obtenue à l'aide des données couleurs et du tenseur de structure multi-échelles.

Nous étiquetons en bleu foncé les pixels classés comme étant de la Peau, en jaune ceux classés Cheveux, en bleu clair ceux classés Fond et enfin en rouge les pixels Indécis. Nous constatons que pour la première image le résultat le plus concluant est obtenu par la méthode utilisant la couleur et les échantillons tandis que pour les deux dernières images, la méthode employant la couleur et le tenseur de structure est meilleure. Dans tous les cas, l'utilisation de la texture est bénéfique. On peut voir que l'approche par échantillons et l'approche par tenseur de structure sont similaires. On peut envisager d'améliorer les modèles de texture en passant avec une étude multi-échelle pour les échantillons et en sélectionnant les échelles du tenseur de structure en fonction des données image [7], en utilisant également plusieurs échelles pour les échantillons.

Les résultats obtenus par notre modélisation peuvent être utilisés pour initialiser avec robustesse une méthode de segmentation plus précise, par exemple exploitant une formulation variationnelle [3].

La segmentation est obtenue en appliquant sur la carte C la méthode de Chan et Vese [8] qui introduit la notion de cohérence de région, en l'occurence les régions Peau, Fond et Cheveux. On obtient alors le résultat présenté sur la figure 9(a).

Pour affiner encore la segmentation obtenue, qui par construction est une segmentation "dure", on peut ensuite utiliser une technique de calcul automatique de masque d'opacité (*matting* dans la littérature [9], [10]). Ce dernier permet d'obtenir un résultat plus réaliste que le résultat de la figure 9(b), en faisant passer les lunettes sous les cheveux par exemple lors de l'essayage virtuel de lunettes. Nous pouvons voir sur la figure 9(c) que les montures des lunettes mélées avec les cheveux ont été légèrement effacées. Ce qui rend la composition plus réaliste.

7 Conclusion et perspectives

En conclusion, l'utilisation de la texture pour segmenter un visage donne de bons résultats dans le sens où nous ne classons pas tous les pixels mais ceux que nous classons nous ne nous trompons pas. La segmentation obtenue est plutôt grossière. Elle sert d'initialisation aux méthodes variationnelles. A travers cette étude nous validons l'idée que la texture est utile pour segmenter des images de visage difficiles ; c'est à dire dont la couleur n'est pas discrimante au sens de la segmentation Peau, Fond, Cheveux. Nous avons étudier le construction de bases par ACP, aux vues des résultats utilisant le tenseur de structure. Nous pouvons envisager de comparer d'autre analyse de texture plus fine et peut-être plus approprié à la caractérisation de Cheveux, par exemple les ondelettes de Gabor ou bien les curvelets.

Enfin la dernière partie d'intégration réaliste à partir de la carte de Peau fera l'objet d'une étude plus approfondie dans le futur.

Références

[1] Vladimir Vezhnevets, Vassili Sazonov, et Alla Andreeva. A survey on pixel-based skin color detection techniques. Dans *Graphicon*, pages 85–92, 2003.

[2] Michael J. Jones et James M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision (IJCV)*, pages 274–280, 1999.

[3] Pauline Julian, Fraçois Lauze, Christophe Dehais, et Vincent Charvillat. Evaluation de méthodes variationnelles pour la segmentation du visage. Dans *COmpression et Representation des Signaux Audiovisuels (Coresa)*, 2009.

[4] Julien Mairal, Marius Leordeanu, Francis Bach, Martial Hebert, et Jean Ponce. Discriminative sparse image models for class-specific edge detection and image interpretation. *European Conference on Computer Vision (ECCV)*, pages 43–56, 2008.

[5] Shai Bagon, Oren Boiman, et Michal Irani. What is a good image segment ? a unified approach to segment extraction. Dans *European Conference on Computer Vision (ECCV)*, volume 5305 de *LNCS*, pages 30–44. Springer, 2008.

[6] R. O. Duda, P. E. Hart, et D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

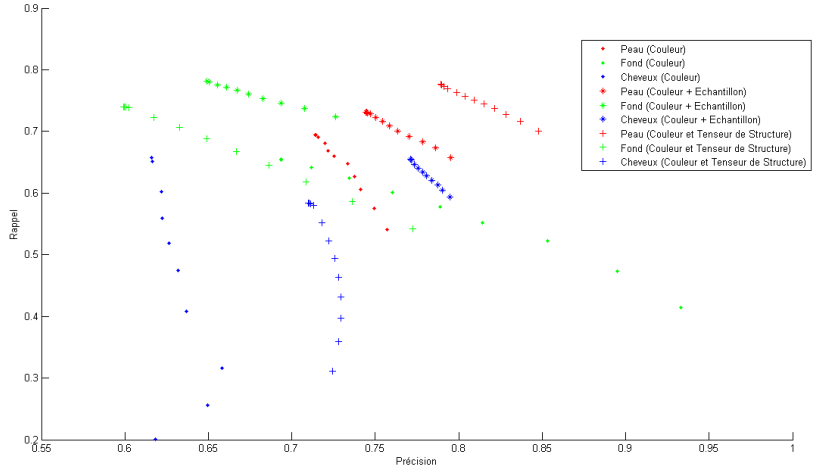
[7] T. Brox et J. Weickert. A tv flow based local scale measure for texture discrimination. Dans T. Pajdla et J. Matas, éditeurs, *European Conference on Computer Vision (ECCV)*, volume 3022, pages 578–590. Springer, 2004.

[8] Tony F. Chan et Luminita A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10 :266–277, 2001.

[9] Yung-Yu Chuang, Brian Curless, David H. Salesin, et Richard Szeliski. A bayesian approach to digital matting. Dans *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 264–271. IEEE Computer Society, 2001.

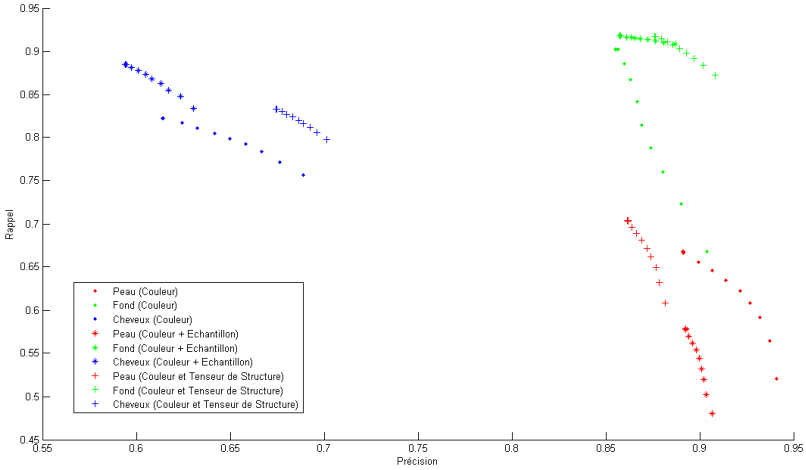
[10] Olivier Juan et Renaud Keriven. Unsupervised segmentation for digital matting. Rapport technique, Centre Enseignement Recherche Traitement Information Systèmes (CERTIS), 2004.

(a)



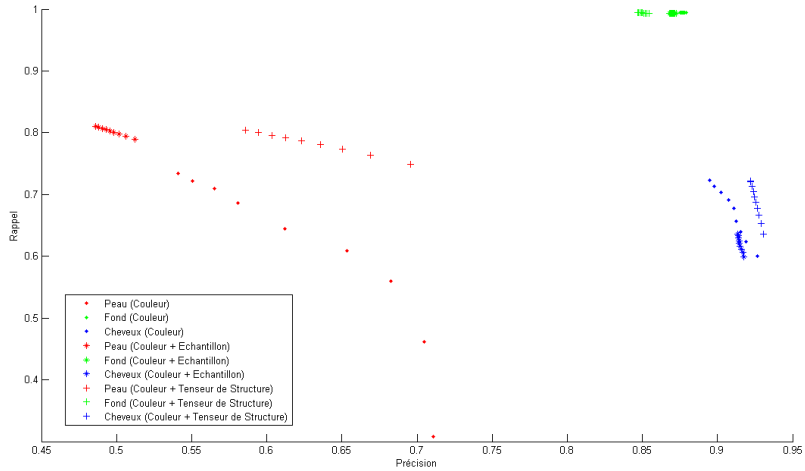
(b)

(c)



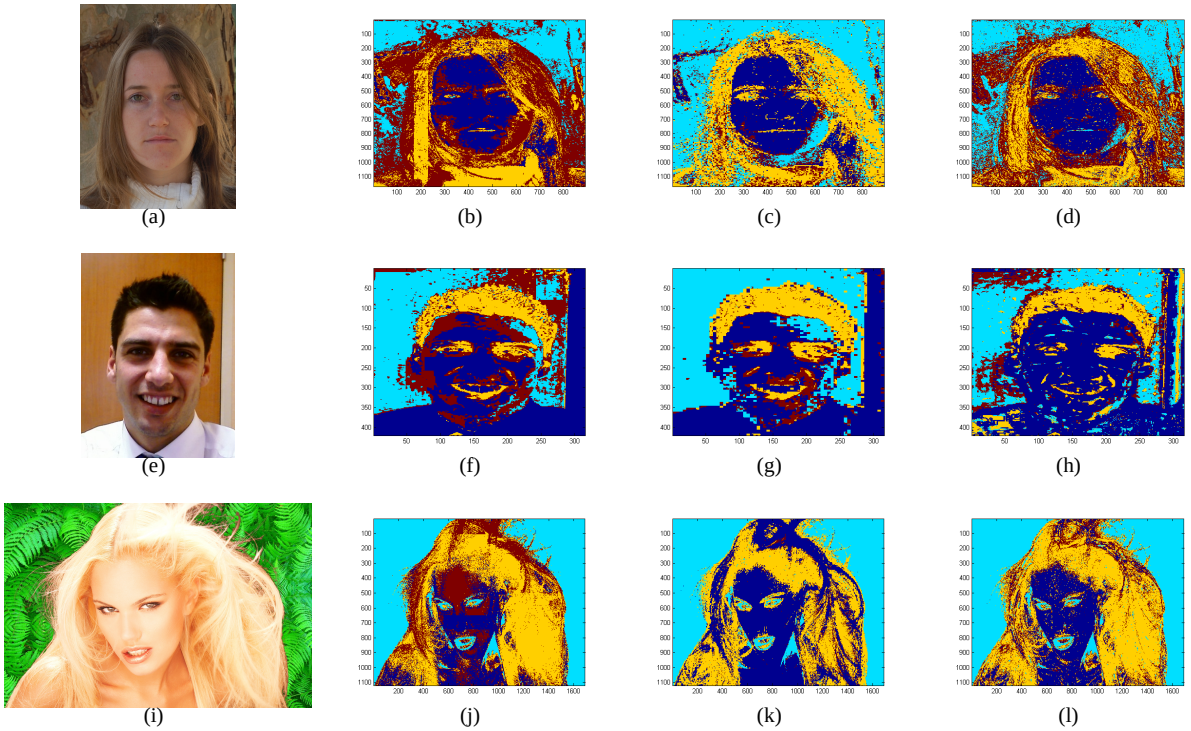
(d)

(e)



(f)

FIG. 7 – Rappel en fonction de la Précision pour les classes Peau, Fond et Cheveux avec les méthodes Couleur, Couleur et Echantillon de Structure, Couleur et Tenseur de Structure



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

FIG. 8 – Résultats des différentes méthodes de classification. Colonne 1 : Image originale ; Colonne 2 : Etude Couleur, Colonne 3 : Etude (Couleur + Echantillon) ; Colonne 4 : Etude (Couleur + Tenseur de Structure)

(a) (b) (c)

FIG. 9 – Résultat de segmentation et d'intégration de lunettes