

Catégorisation des expressions faciales par marches aléatoires sur graphe

Youssef CHAHIR¹, Youssef ZINBI¹ et Kheir Eddine AZIZ¹

¹ 1 GREYC - URA CNRS 6072, Université de Caen
Campus II - BP 5186, 14032 Caen Cedex

{chahir,yzinbi,kaziz}@info.unicaen.fr

Résumé

La description des expressions faciales est un ensemble d'interprétations successives de composantes faciales. Un sens est construit à partir de distances de bas niveaux (présence ou non d'une composante, position éventuelle). Dans cet article, nous proposons une approche basée sur la diffusion géométrique par marches aléatoires sur graphe, pour la catégorisation des expressions faciales. L'idée de base est de considérer un graphe modélisant tous les visages d'une base de grande dimension à travers les descripteurs des différentes composantes faciales. Cette approche permet de caractériser au mieux l'information visuelle en se basant sur les graphes de similarités et l'exploitation de leurs propriétés spectrales. Pour cela, nous proposons un modèle unifié pour l'extraction des points caractéristiques d'une expression faciale, et nous exploitons un ensemble de distances caractéristiques auquel est associé un ensemble de règles logiques.

Mots clefs

Composante faciale, expression faciale, diffusion, graphe, marche aléatoire, règles logiques.

1 Introduction

Le visage est un moyen de communication important et complexe. Il émet en permanence des signes dont le décodage, non seulement renseigne sur l'état émotionnel de la personne, mais aussi éclaire sur ce qui est dit. Nous sommes très sensibles à son message visuel. Une expression faciale est une manifestation visible de l'état émotionnel, de l'activité cognitive, de l'intention, de la personnalité et de la psychopathologie d'une personne. Mehrabian [1] a mis en évidence le fait que 55% du message émotionnel soient communiqués par l'expression faciale alors que 7% seulement par le canal linguistique et 38% par le paralinguage. Ainsi, les expressions faciales jouent un rôle important dans la communication humaine mais aussi en interaction homme-machine.

Pour faire de la reconnaissance des expressions faciales, il faut résoudre trois tâches principalement: localiser le visage; extraire de ce visage les informations utiles à la reconnaissance de l'expression faciale; définir l'ensemble

des expressions à reconnaître ainsi que le processus de classification associé. La première étape à tout système d'analyse d'émotions concerne la localisation du visage dans l'image. De nombreux travaux ont été effectués à ce sujet. Le lecteur pourra se référer aux deux articles de synthèse [2,3] pour avoir une description détaillée des algorithmes existants. Une fois le visage localisé (sous la forme d'une boîte englobante), il s'agit d'en extraire les informations pertinentes qui conduiront à l'analyse des expressions. On distingue pour cela deux types d'approche:

- Les approches qui considèrent le visage comme une entité globale dont les caractéristiques et les déformations sont apprises. Ces méthodes requièrent des phases significatives d'apprentissage. On peut citer pour exemple les algorithmes développés dans [4,5] qui utilisent une représentation du visage par modèle statistique d'apparence.
- Les approches qui s'intéressent à des traits particuliers du visage afin d'étudier les déformations. Les traits ou points caractéristiques pris en compte diffèrent d'une méthode à l'autre même si les yeux, la bouche et les sourcils sont des caractéristiques prises en compte de manière dominante. La difficulté réside dans le développement d'algorithmes permettant l'extraction et/ou le suivi des traits sélectionnés et ce de manière automatique.

La dernière étape réside dans la phase de classification. Les classifieurs les plus couramment utilisés sont les réseaux de neurones, les classifieurs de type bayésien et les classifieurs à base de règles. Pour plus de détails à propos des différents systèmes de reconnaissance d'expressions, le lecteur pourra se référer aux articles de synthèse [6,7]. Ekman et Friesen [1] ont établi qu'il existe un nombre limité d'expressions reconnues par tous, indépendamment de la culture. Ces expressions innées correspondent aux sept émotions suivantes: la neutralité, la joie, la tristesse, la surprise, la peur, la colère et le dégoût. La terminologie utilisée généralement consiste à décrire les expressions faciales à partir des informations telles que la position et la géométrie des actions faciales (déformations faciales des yeux et de la bouche) [8]. Des expériences psychologiques ont montré que les contours des yeux, des sourcils et des lèvres, sont des informations qui doivent être prises en compte dans le processus de

categorisation des expressions faciales. En se basant sur ces observations, nous proposons un traitement pour l'extraction des points caractéristiques de ces traits. Par la suite, et dans le but de mesurer les déformations de ces traits, nous définirons un ensemble de distances caractéristiques. Nous leur associons un ensemble de règles logiques, issues de la description MPEG4, pour décrire chaque expression faciale.. Dans la section suivante, nous décrirons dans un premier temps l'algorithme utilisé pour la categorisation des expressions faciales et dans un second temps, l'application de cet algorithme sur l'ensemble des vecteurs d'informations issus des composantes faciales détectées par notre approche. Un processus de classification spectral [9], dans un espace de diffusion, a été proposé pour categoriser les expressions faciales. Enfin, la dernière partie évoquera la conclusion avec les perspectives de ce travail.

2 Extraction de composantes faciales

Plusieurs méthodes de segmentation de traits faciaux ont été abordées dans la littérature. Abdel et al [10] ont remarqué que ces traits sont caractérisés par les propriétés chrominances de l'image (EyeMapC) et de luminance (EyeMapL). D'autres travaux se basent sur les modèles d'apparence [11].

2.1 Marches aléatoires sur graphe :

Nous voulons représenter un ensemble finie d'objets $\Omega = \{u_1, u_2, \dots, u_n\}$ où $u_i \in \mathbb{R}^n$ est un élément de dimension n . Nous considérons un graphe $G = (V, E)$ qui consiste en un ensemble fini $V = \Omega$ possédant N sommets et un ensemble fini $E \subseteq V \times V$ de M arêtes. Deux sommets u et v sont adjacents si l'arête $(u, v) \in E$. Un graphe est considéré comme un graphe pondéré, si on peut lui associer une fonction de poids $w : V \times V \rightarrow \mathbb{R}^+$ avec $w(u, v) = w(v, u)$ pour chaque sommet $u, v \in V$. Cette fonction de poids reflète le degré de similarité entre deux sommets du graphe et décrit ainsi l'interaction du premier ordre entre les sommets du graphes. Le graphe peut être représenté par sa matrice de similarité W : $W(u, v) = w(u, v)$ si les sommets u et v sont adjacents et $W(u, v) = 0$ dans le cas contraire. Les sommets étant liés à eux-mêmes nous avons pour tout sommet u $W(u, u) = 1$. Le degré d'un sommet u est défini par :

$$d(u) = \sum_{v \in V} w(u, v).$$

Nous allons nous intéresser à un processus de marche aléatoire (ou de diffusion dans le graphe G). Le temps est discrétisé $t = (0, 1, 2, \dots)$. A chaque instant, un marcheur est localisé sur un sommet et se déplace à l'instant suivant vers un sommet choisi aléatoirement et uniformément parmi les sommets voisins. La suite des sommets visités est alors une marche aléatoire, et la

probabilité de transition du sommet u au sommet v est à chaque étape : $p(u, v) = w(u, v) / d(u)$

Ceci définit la matrice de transition P de la chaîne de Markov correspondant à la marche aléatoire. La matrice P est stochastique, en effet :

$$\forall u, \forall v, 0 \leq p(u, v) \leq 1 \text{ et } \sum_{v \in V} p(u, v) = 1$$

Nous pouvons aussi écrire $P = D^{-1}W$ en introduisant la matrice diagonale des degrés des sommets D : $D(u, u) = d(u)$ et $D(u, v) = 0$ pour $u \neq v$. Considérons $p_t(u, v)$ le noyau correspondant à la t ème puissance de P : P^t . $p_t(u, v)$ peut être interprété comme la probabilité pour un marcheur d'attendre le sommet v en partant du sommet u en t étapes.

L'intérêt d'introduire cette matrice de transition est que l'exploration du graphe par la marche aléatoire qu'elle engendre permet de déterminer des propriétés topologiques du graphe [12], reliées aux propriétés spectrales de P . La décomposition spectrale de la matrice P donne un ensemble de valeurs propres $|\lambda_1| = 1 \geq |\lambda_2| \geq \dots \geq |\lambda_{N^2}| \geq 0$ engendrant un ensemble de vecteurs propres $\{\phi_1, \phi_2, \dots, \phi_m\}$, solutions de: $P \phi_m = \lambda_m \phi_m$.

Dans le cadre de l'analyse et de la caractérisation des composantes faciales, nous avons proposé un nouveau descripteur de texture unique et simple à calculer, qui est basée sur la trace de la matrice de transition P représentant le graphe au voisinage local de chaque pixel [13]. Pour extraire les composantes faciales, une chaîne de traitement (fig. 1) a été mise au point :

- Trace des matrices de transition pour caractériser les composantes faciales.
- Classification et marquage pour labelliser les régions.
- Extraction des points caractéristiques de chaque composante faciale

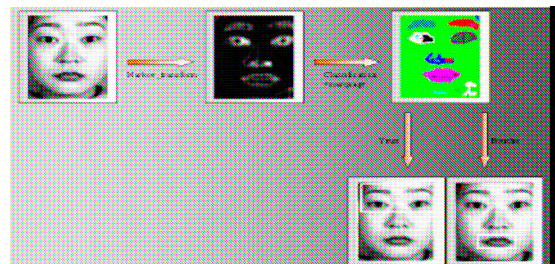


Figure 1 - Extraction des composantes faciales par la trace des matrices de transitions

2.2 Description par des distances MPEG4

La norme MPEG-4 fournit une description des transformations subies par chacun des traits du visage lors

de la production de chacune des six émotions universelles [14]. La Figure 2 décrit l'ensemble des distances mpeg4. D2 et D7 donnent une mesure de la distance entre les yeux et les sourcils. D6 mesure la distance entre les yeux et la bouche. D3 et D4 mesurent le degré d'ouverture de la bouche, D1 mesure le degré d'ouverture des yeux.

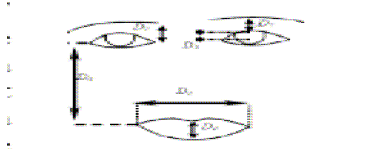


Figure 2 - Définition des distances D_i

3 Catégorisation des composantes faciales

Après avoir obtenu les composantes faciales sous forme de vecteurs d'intensité (trace des matrices de transition) et de distances (points caractéristiques), on s'intéresse dans ce paragraphe à trouver une projection de ces vecteurs dans un espace réduit.

Rappelons la transformation

$$\psi_t : \mathbb{R}^n \rightarrow \mathbb{R}^{m(t)}$$

$$u \rightarrow \psi_t(u) = (\lambda_1^t \phi_1(u), \lambda_2^t \phi_2(u), \dots, \lambda_{m(t)}^t \phi_{m(t)}(u))$$

Cette transformation est communément utilisée pour l'analyse et la réduction de donnée de grande dimension [15,16]. Elle permet donc de passer d'un espace de mesure de dimension n à un espace de représentation homogène de dimension $m(t)$ plus réduit représentant toutes les informations ainsi que les propriétés structurales du graphe. Soulignons que de nombreux travaux ont montré que les propriétés structurales du graphe sont capturées principalement par les vecteurs propres de P^t liés aux plus grandes valeurs propres [17] et leur visualisation fournissent une représentation des données dans un espace où il est bien séparé et peut facilement être groupé. On a généralement $m(t) \leq 3$. Le schéma suivant résume notre approche de catégorisation :

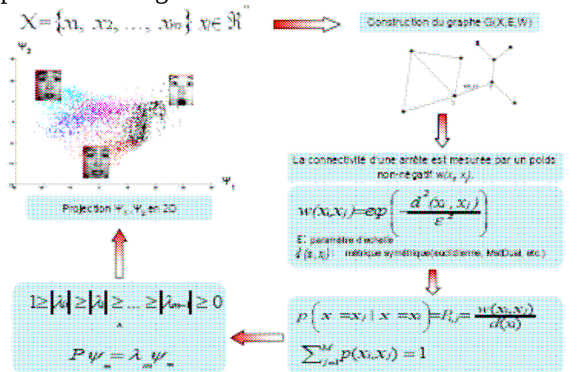


Figure 3 – Classification spectrale

4 Expérimentation sur un corpus de visages

Après avoir présenté, les propriétés spectrales des marches aléatoires sur graphe, nous présentons dans cette section deux démonstrations de catégorisation basées sur les vecteurs d'intensités et les vecteurs de distances. Nous avons testé notre approche sur la base de visage JAFFE (fig. 4) répartie en sept classes d'expressions faciales universelles.

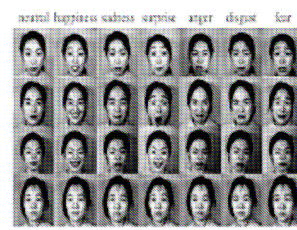


Figure 4 - Base de visage JAFFE

4.1 Similarité basée sur les vecteurs de caractérisation de la texture (intensités)

Le visage est une zone importante du corps humain qui possède une quarantaine de muscles. Il est à noter que les muscles de la zone supérieure du visage n'ont que peu d'influence sur les muscles de la zone inférieure et vice-versa [1]. Il est donc possible de découper l'analyse en deux zones : la zone des composantes faciales yeux et la zone de composante faciale bouche. Pour cela, nous avons testé le résultat de la classification sur chacune des composantes..

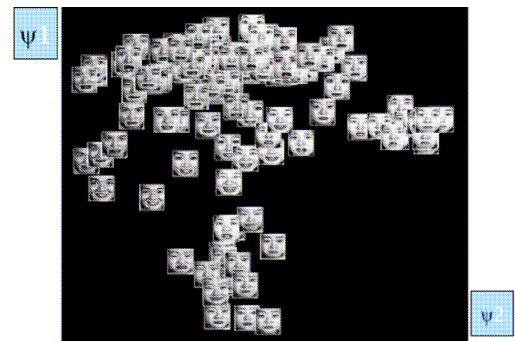


Figure 5 - La projection des deux premiers vecteurs propres dans l'espace 2D selon la forme de la bouche.

En se basant uniquement sur le vecteur de caractérisation (intensité) de la zone bouche, deux catégories d'expressions faciales se mettent en évidence : « Sourire » et « Surprise » (Figure 5). Ce résultat s'explique par la différence d'intensité qui existe entre les deux formes correspondant à la composante bouche ouverte. En revanche, le reste des images avec des bouches fermées est dispersé dans l'ensemble des classes selon la forme des contours de la bouche.

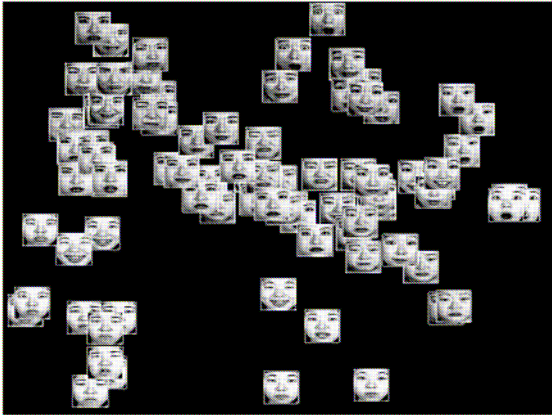


Figure 6 - La projection des deux premiers vecteurs propres dans l'espace 2D selon la forme des yeux.

De même, la projection de la composante faciale des yeux (figure 6), montre une organisation selon la forme, et cela est dû au fait que la déformation entre l'œil et le sourcil est ici estimée par une simple différence d'intensité. Il s'avère important de prendre en compte la variation de distance entre les deux mimiques faciales sourcils et yeux, et de lier cette distance à la déformation de la bouche. D'où la nécessité de prendre en compte des distances issues de la l'information des points caractéristiques de chaque composante faciale.

4.2 Mesure de similarité basée les distances

Dans leur travaux [18] Hammal et al. ont proposé de définir des distances caractéristiques et ont établi une base de règles pour la modélisation des expressions faciales. La table extraite de leur travaux est présentée ci-dessous (cf. tableau 2).

Sourire	C-	S/C-	C+	C+	C-
Surprise	C+	C+	C-	C+	C+
Dégoût	C-	C-	S/C+	C+	S/C-
Colère	C+	C-	S	S/C-	S
Tristesse	C-	C+	S	S	S
Peur	S/C+	S/C+	S/C-	S/C+	S/C+
Neutre	S	S	S	S	S

Tableau 2 - Etats symboliques associés à chaque expression

- Etat C+ pour lequel la distance D_i est plus grande que celle pour l'expression neutre.
- Etat S pour lequel la distance D_i est de même ordre de grandeur que celle pour l'expression neutre.
- Etat C- pour lequel la distance D_i est plus petite que celle pour l'expression neutre.

En considérant qu'une expression faciale est capturée par l'ensemble de ces distances, nous allons calculer la similarité entre une expression donnée et une expression

neutre. En se basant sur ces états symboliques, nous avons construit une matrice de similarité appropriée où les lignes représentent les expressions neutres de chaque visage et les colonnes représentent l'ensemble de visages qui vont vérifier les états décrits dans le tableau ci-dessus. Par conséquent, le nouveau calcul de la matrice de similarité se fera de la manière suivante :

$$w(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right), x \in \Omega_1, y \in \Omega_2$$

Ω_1 : ensemble des composantes faciales correspondantes aux visages d'expression neutre.

Ω_2 : ensemble des composantes faciales correspondantes aux visages de différentes expressions.

$$W' = W \cdot W^t$$

La figure 7 montre un résultat de catégorisation par simple projection sur les deux premiers axes, en utilisant les distances mpeg4. Le résultat obtenu n'est pas significatif. Par contre, dans la figure 8, où il s'agit d'utilisation combinant les deux distances, la répartition des visages donne un résultat plus au moins acceptable dans le cas d'expressions neutres, de joie et de surprise.

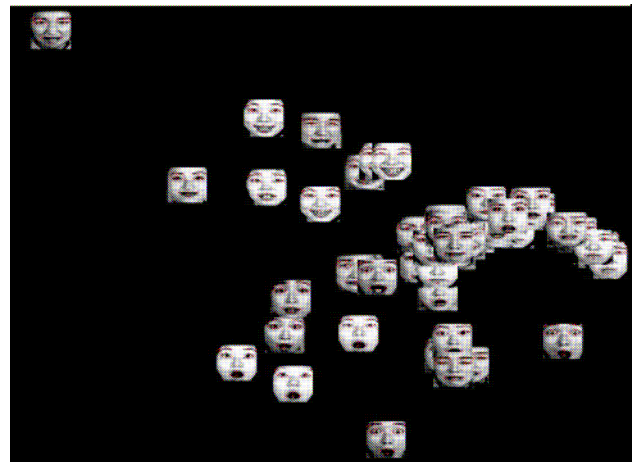


Figure 7 - Projection des deux premiers vecteurs propres dans l'espace 2D selon les distances entre les points caractéristiques.

En revanche, dans certains cas, des erreurs de confusion sont apparues entre les différents types d'expressions. Ces confusions sont dues principalement à l'imprécision lors de la détection automatique de points caractéristiques de chaque composante faciale ainsi que par des erreurs de segmentation des composantes faciales, notamment avec les composantes faciales des bouches ouvertes et les sourcils. Un autre point intéressant à notre sens (un avantage) est que l'expression neutre se trouve dans un groupe au milieu du nuage de points, ce qui confirme la validité de la proposition concernant le calcul de la nouvelle matrice de similarité décrite précédemment.

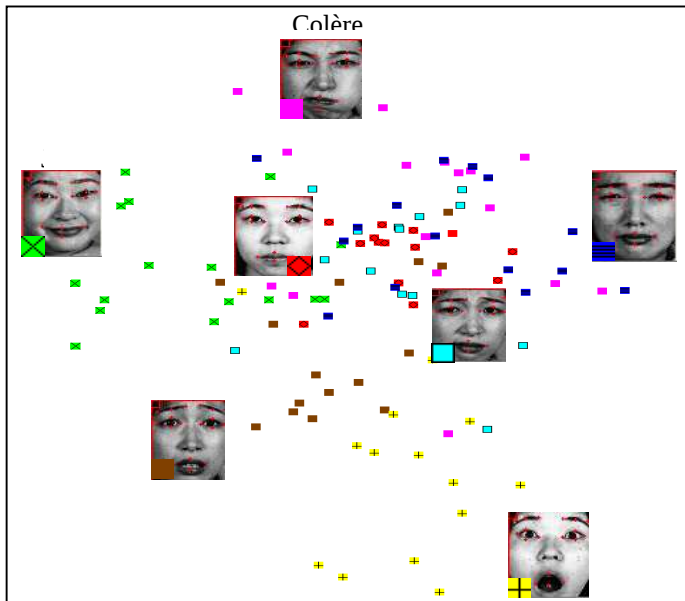


Figure 8 - Répartition des composantes faciales dans l'espace 2D par approche mixte

5 Conclusion et perspectives

Dans ce travail, nous avons présenté notre contribution en essayant de répondre aux deux questions liées à la catégorisation des expressions faciales. Dans la première partie de ce travail, nous avons présenté un algorithme de détection et de segmentation des composantes faciales des yeux et de la bouche. Ce modèle est basé sur la partition locale de graphe de grande dimension. Il a montré son efficacité du fait qu'il caractérise non seulement les composantes faciales mais aussi les déformations musculaires autour des composantes. Toute déformation est caractérisée par l'interaction des pixels constituant la zone musculaire autour de la composante faciale, plus la composante. Ce qui fait de ce modèle un outil pour caractériser les actions faciales notamment dans la vidéo. En se basant sur les composantes faciales issues de la phase de la segmentation, nous avons utilisé un algorithme de classification spectral a été choisi pour visualiser la répartition de ces composantes. Ainsi, pour réaliser cette tâche, deux informations essentielles ont été prises en compte : l'information d'intensité (niveau de gris), les mesures de distances trouvées à partir de la déformation des points extrême correspondant aux composantes faciales. La projection des deux vecteurs propres, calculés par la décomposition matricielle de la matrice de similarité, calculé d'une manière différente du mode classique, construite à partir des vecteurs d'information des composantes faciales, a montré de bons résultats de classification des expressions faciales 'Sourire', 'Surprise' et 'Neutre'. D'un autre côté, pour palier aux fausses détections, il apparaît nécessaire d'exploiter les contours caractéristiques du visage par des approches de contours actif.

Références

- [1] P. Ekman, -Facial Expression-, The Handbook of Cognition and Emotion, T. Dalgeleish and M. Power, John Wiley & Sons Ltd. 1999.
- [2] H. Hjelmäs, B. Low - Face detection: a survey - Computer Vision and Image Understanding, 83, pp. 236-274, 2001.
- [3] M.H. Yang, D. Kriegman, and N. Ahuja - Detecting face in images : a survey - IEEE Trans on PAMI, vol.24, n°1, pp. 34-58, January, 2002
- [4] G.J. Edwards, T.F Cootes, C.J. Taylor - Face recognition using active appearance models - Proc. Europ. Conf. Comp.Vision, vol.2, pp. 581-595, 1998.
- [5] B. Abboud, F. Davoine, Facial expression recognition and synthesis based on an appearance model, Signal Processing: Image Communication, Elsevier, Vol. 19, No. 8, pages 723-740, 2004.
- [6] B. Fasel, J. Luetttin, Automatic Facial Expression analysis: a survey, PR., vol.36, pp.259-275, 2003
- [7] M. Pantic, L.J.M Rothkrantz -Automatic analysis of facial expressions: the state of the art- IEEE trans. on PAMI, vol.22, N°12, pp. 1424-1445, 2000.
- [8] Z. Hammal, et al., Extraction des traits caractéristiques du visage à l'aide de modèles paramétriques adaptés - Traitement Du Signal, volume 22, N°1, pp.59-72, 2005.
- [9] A. Ng, M. Jordan, Y. Weiss. On Spectral Clustering Analysis and an algorithm. NIPS 2001: pp 849-856
- [10] M. Abdel-Mottaleb, R-L Hsuy, A.K. Jain . Face Detection in Color Images. 2002
- [11] F. Davoine, B. Abboud,, M. Dang. Analyse de visages et d'expressions faciales par modèle actif d'apparence, Traitement du Signal, Vol.1, n°3, 2004
- [12] S. Ingve et al., Diffusion on complex networks : a way to probe their large-scale topological structures. Physica A: Statistical Mechanics and its Applications, 336(1-2) : 163-173, May 2004.
- [13] Y. Chahir et al. , « Caractérisation de la texture par marches aléatoires locales sur graphe », gretsi 2007
- [14] Z. Hammal et al. Parametric models for facial features segmentation. 2004.
- [15] B. Gaveau, A. Lesne, L.S Schulman, Spectral signatures of hierarchical relaxation, Physical Letters A: 258, pp 222-228
- [16] R.R Coifman et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data Diffusion map. Proceeding of the National Academy of Science, 102: 7426-7431, May 2005.
- [17] Yair Weiss. Segmentation using eigenvectors: a unifying view. Proceedings IEEE International Conference on Computer Vision p. 975-982 ,1999.
- [18] Z. Hammal, A. Caplier, M. Rombaut, Classification des expressions faciales par la théorie de l'évidence, Rencontres Francophones sur la Logique Floue et ces Applications LFA 2004.