

Kheir Eddine Aziz. [kaziz@etu.info.unicaen.fr](mailto:kaziz@etu.info.unicaen.fr)

Master Professionnel Réseaux, Applications Documentaires, Image

Novembre - Mars 2007



UNIVERSITE DE CAEN  
BASSE-NORMANDIE



GROUPE DE RECHERCHE EN INFORMATIQUE,  
IMAGE, INSTRUMENTATION DE CAEN;  
GREYC-DoDoLa-IMAGE

Projet annuel

# Analyse de visage et catégorisation des expressions faciales.

Tuteurs du projet : Youssef Chahir. [Chahir@info.unicaen.fr](mailto:Chahir@info.unicaen.fr)

Abderrahim Elmoataz. [Abder.Elmoataz@greyc.ensicaen.fr](mailto:Abder.Elmoataz@greyc.ensicaen.fr)

# Remerciements

Je tiens à remercier également Youssef Chahir et Abderrahim Elmoatez avec qui j'ai travaillé en collaboration sur ce sujet, pour sa patience, son support, ses inestimables conseils et ses explications dans le domaine de vision.

Un merci spécial à Hichem et Anne Lise Alexandre qui m'ont aidé à perfectionner mon art grammatical.

*A mes frères, à mes soeurs*

**Table of Contents**

Introduction .....	3
Analyse.....	5
1.2 Psychophysiologie du visage.....	6
1.2.1 Physiologie du visage.....	6
1.2.2 Psychologie du visage.....	7
1.3 Extraction des composantes faciales.....	7
1.4 Formalismes de description.....	8
1.4.1 FACS.....	9
1.4.2 FACS+.....	10
1.4.3. MPEG4.....	10
Détection et segmentation des composantes faciales .....	13
2.1 La transformation markovienne:.....	13
2.1.1 Description de l'algorithme:.....	14
2.1.2 Détection et segmentation des composantes faciales.....	14
2.2. La carte des yeux et de la bouche EyeMap/MouthMap.....	15
2.2.1 Description de l'algorithme EyeMap/MouthMap.....	16
2.2.2 Détection et segmentation des composantes faciales.....	17
Catégorisation des composantes faciales.....	19
3.1 Les marches aléatoires sur graphe(Graphe de Laplacien)[CHU97]:.....	19
3.2 Tests et résultats.....	21
3.2.1 Catégorisation avec les vecteurs d'intensités:.....	22
3.2.1 Catégorisation avec les vecteurs de distances:.....	24
Conclusion.....	28
Bibliographie.....	29

# 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.

Les premiers travaux sur le domaine sont dûs principalement à Charles Darwin qui a été l'un des premiers scientifiques à reconnaître que l'expression faciale est le moyen immédiat le plus puissant chez les êtres humains pour communiquer leurs émotions, intentions, et opinions.

De plus 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 [MER06] a mis en évidence le fait que 55% d'un message émotionnel sont communiqués par l'expression faciale alors que 7% seulement par le canal linguistique et 38% par le paralanguage (comme l'intonation). Ceci implique que les expressions faciales jouent un rôle important dans la communication humaine. Donc l'interaction homme-machine tirera un véritable bénéfice d'un système automatique de reconnaissance d'expressions faciales.

Dans le travail présenté nous nous sommes intéressés à la catégorisation des expressions faciales à partir d'un ensemble d'images de visages. Comme tout autre comportement humain, reconnaître une expression faciale est une tâche complexe à accomplir par un système de vision par ordinateur à cause de la grande variabilité entre les individus. De nombreux travaux ont été proposés pour la reconnaissance et l'interprétation des expressions faciales. Ces travaux ont fait émerger deux questions importantes :

- *Quels sont les indices pertinents qui doivent être extraits d'un visage?*
- *Comment le comportement de ces indices peut être modélisé et traduit pour la catégorisation des expressions faciales?*

Dans ce travail nous avons essayé d'apporter notre contribution à la réponse à ces deux questions.

Notre but est la catégorisation des expressions faciales issues des images de visage de face. La première étape est donc d'identifier quelle est l'information présente dans un visage qui va être utilisé pour reconnaître une expression. Dans ce contexte, l'expression faciale est définie comme un changement dans le visage, perceptible visuellement, dû à l'activation (volontaire ou non) d'un ou de plusieurs muscles composant le visage (yeux, lèvres, etc). Ekman et Friesen [MER06] 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**. C'est plus spécialement sur ces expressions que notre travail va porter.

La terminologie utilisée pour décrire les expressions faciales est basée sur la position, l'intensité des actions faciales (déformations faciales des yeux et de la bouche). L'intensité des actions faciales peut être mesurée en déterminant la déformation géométrique des structures faciales ou la densité des rides apparaissant dans certaines régions du visage telle que les yeux et la bouche.

En outre, 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 catégorisation des expressions faciales. Notre système de catégorisation analysera ensuite le comportement de ces traits.

En se basant sur ces observations, nous proposons un traitement pour l'extraction des points caractéristiques de ces traits. L'extraction de ces données, contrairement aux travaux de [HAM06], est basé sur l'information de niveaux de gris.

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 pour décrire chaque expression faciale. Ces règles sont basées sur la description MPEG-4 des six émotions universelles (**neutralité**, **joie**, **tristesse**, **surprise**, **peur**, **colère** et **dégoût**).

Basé sur ces règles logiques et l'intensité des composantes faciales, un processus de clustering spectrales a été choisi pour finalement catégoriser les expressions faciales.

Récemment, l'intérêt pour le traitement de données s'est tourné vers les méthodes de clustering spectrales, en raison de nombreux succès [CHU97][TEN00][ROW00][GEN05]. Ces méthodes emploient le contenu spectral d'une matrice de similarité (distance entre chaque paire de données) pour réaliser la réduction et la partition d'un jeu de données. Plus spécifiquement, les vecteurs propres sont vus comme un outil fournissant une représentation (visualisation) des données dans un espace où il est bien séparé et peut facilement être groupé (classifier).

Dans la première partie, nous présenterons les principaux concepts liés à l'expression faciale ainsi que les principaux formalismes de leurs descriptions. Dans une seconde partie, nous évoquerons en détaille deux méthodes dédiées à la détection et la segmentation des structures faciales. Dans la troisième partie, nous décrirons dans un premier temps l'algorithme utilisé pour la catégorisation 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 dans la deuxième partie. Enfin, la dernière partie évoquera les perspectives de ce travail.

# Chapitre 1

# Analyse

On présente dans ce chapitre un rappel des connaissances sur le domaine de l'analyse du visage et des expressions du visage:

- ★ Les particularités du visage humain et du système humain d'analyse du visage.
- ★ Les principales approches employées pour la détection des structures faciales.
- ★ Les différents formalismes de description des expressions du visage.

## 1.1 Définitions

Expressions et émotions sont très liées et parfois confondues, c'est pour cette raison qu'on se tiendra aux définitions suivantes par la suite :

- **Mimique faciale** : une mimique faciale est un état du visage composé par un ensemble de configurations des muscles faciaux. Le sourire est par exemple une mimique faciale composé d'un certain nombre d'activation de muscles faciaux (mouvements des muscles zygomatiques).
- **Emotion** : l'émotion est un des générateurs des expressions faciales. L'émotion se traduit via de nombreux canaux comme la position du corps, la voix et les expressions faciales. Une émotion implique généralement une expression faciale correspondante (dont l'intensité peut être plus ou moins contrôlée selon les individus), mais l'inverse n'est pas vrai : il est possible de mimer une expression représentant une émotion sans pour autant ressentir cette émotion. Alors que les expressions dépendent des individus et des cultures, on distingue généralement un nombre limité d'émotions universellement reconnues.
- **Expression faciale** : une expression faciale est une mimique faciale chargée de sens. Le sens peut être l'expression d'une émotion, un indice sémantique ou une intonation dans la Langue des Signes.

L'interprétation d'un ensemble de mouvements musculaires en expression est dépendante du contexte d'application. Dans le cas d'une application en interaction Homme-Machine où l'on désire connaître une indication sur l'état émotionnel d'un individu, on cherchera à classifier les mesures en terme d'émotions.

Pour une application en Langue des Signes, les mesures seront combinées pour construire un sens, qui ne reflète pas forcément l'état émotionnel de l'individu.

## 1.2 Psychophysiologie du visage

Avant de s'intéresser à l'analyse automatique des expressions du visage, on doit

connaître les particularités du visage : son anatomie et le lien entre le visage et les émotions.

### 1.2.1 Physiologie du visage

Le visage est une zone importante du corps humain qui possède une quarantaine de muscles. L'activation musculaire, et en particulier l'activation des muscles du visage, peut généralement être découpée en trois phases :

- la phase d'attaque «attack» ou «onset», qui correspond à la période pendant laquelle l'activité du muscle passe de la valeur nulle à sa valeur maximale.

- la phase de soutien «sustain» ou «apex», qui correspond à la période pendant laquelle l'activité stagne à son maximum.

- la phase de relâchement «relaxation» ou «offset», qui correspond à la période pendant laquelle l'activité du muscle baisse jusqu'au niveau initial.

### Indépendance des 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[MER06]. Il est donc possible de découper l'analyse en deux zones.

### 1.2.2 Psychologie du visage

Les expressions faciales peuvent former une indication sur l'état émotionnel d'un individu : ce sont les expressions dites «spontanées» en contraste avec les expressions qui peuvent être «forcées». Les deux types d'expression sont générées par deux zones distinctes du cerveau.

On distingue par exemple deux types de sourire, dit «de Duchenne» ou non. Le sourire de Duchenne est un sourire sincère reflétant une émotion positive ; l'activation des muscles entourant les yeux accompagne en général ce sourire. Il semblerait aussi que les mesures temporelles (onset, apex et offset) de l'activation des muscles soient différentes entre les deux types de sourire.

La distinction entre les deux types peut être mise en évidence si les différentes mesures sont suffisamment précises.

### Aspect universel des émotions

Ekman et Friesen [MER06] 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**.

### 1.3 Extraction des composantes faciales

L'approche la plus utilisée pour l'analyse du visage consiste à considérer le visage



comme un ensemble de composantes ayant une certaine configuration spatiale. L'analyse par cette approche consiste alors à détecter les composantes en utilisant des détecteurs spécialisés basés sur l'intensité au cas d'une image en niveaux de gris (la *transformation markovienne*, *contour actif*, *Gradient*, *etc*) ou sur la luminance et la chrominance de l'image couleur (*EyeMap/MouthMap*). Ces composantes vont être réorganisées spatialement pour les faire correspondre à un modèle du visage. Les principales composantes qui composent le visage sont:

- **Sourcils**

Les sourcils, s'ils sont présents, sont généralement plus foncés que le reste du visage. Un détecteur de contours peut permettre la détection des sourcils.

- **Yeux**

La zone des yeux est composée de plusieurs composantes intéressantes pour l'analyse des expressions : les paupières, les yeux (iris, blanc), les muscles entourant les yeux qui entrent en jeu dans l'action faciale «plissement des yeux» et les rides sur le côté.

- **Paupières**

Les paupières ont la couleur de la peau. Détecter une zone ayant la couleur de la peau à l'endroit des yeux indique la présence des paupières.

Le clignement d'yeux est un processus naturel qui intervient relativement fréquemment. Il est possible de détecter les yeux en se basant sur le principe qu'ils clignent avec une certaine fréquence et tous les deux en même temps.

- **Iris, blanc des yeux**

Les yeux possèdent un profil horizontal de couleur très spécifique : clair, foncé et clair. Une analyse du gradient ou des histogrammes verticaux / horizontaux peut permettre de détecter les caractéristiques des yeux. La forme ronde de l'oeil peut aussi servir à la détection, par exemple par une transformée de Hough.

## 1.4 Formalismes de description

La description des expressions du visage est un ensemble d'interprétations successives d'indices visuels. Un sens (dépendant du domaine d'application) est construit à partir de mesures de bas niveau (présence ou non d'une composante, position éventuelle). Ces mesures sont combinées successivement spatialement, puis de manière temporelle pour former le sens attendu (émotion sous-jacente par exemple).

On introduit ici un vocabulaire nécessaire à la description.

- **Attribut facial** : un attribut facial est une propriété élémentaire «centrée objet» caractérisant un visage. La position des yeux est un attribut facial. La présence de barbe est un autre attribut. Les attributs faciaux directement visibles sont dit de premier ordre. Les attributs qui ne peuvent être mesurés qu'à partir d'autres attributs de premier ordre, sont des attributs de second ordre et ainsi de suite

[HAM06].

La mesure de distance entre les deux yeux est un attribut du second ordre, par exemple, puisqu'elle ne peut être calculée qu'à partir de la position respective des deux yeux qui est une mesure du premier ordre.

Certains attributs sont invariants par rapport au temps pour un même individu : la distance entre les deux yeux est par exemple un attribut statique, composé cependant de deux attributs dynamiques (position des yeux qui change au cours du temps).

- **Indice visuel** : un indice visuel est une propriété élémentaire «centrée observateur» du visage : c'est un attribut facial qui est observé et visible. Certains attributs ne sont pas visibles chez certaines personnes (barbe, moustache, sourcils) ; certains ne sont visibles qu'à certains moments (un oeil peut être caché lors d'une rotation de la tête par exemple).
- **Action faciale** : une action faciale est un ensemble d'indices visuels intégrés de manière temporelle. Le relèvement des sourcils est par exemple une action faciale composée d'un ensemble de positions successives des sourcils. Une action faciale est généralement décrite par sa dynamique : le relèvement des sourcils consiste en une position actuelle des sourcils plus haute que sa position précédente. Les actions faciales sont généralement caractérisées par leur «profil temporel»: durée d'attaque, durée de maintien et durée de relâchement.
- **Composante faciale** : une composante faciale est une partie du visage. Le découpage en composantes est celui du langage naturel : les yeux, le nez, la bouche, les joues, les sourcils, la barbe, etc. Bien que certaines puissent être entièrement caractérisées par un ensemble d'attributs faciaux (les yeux peuvent être caractérisés par leur forme, leur couleur, la présence et la longueur des cils, etc.), d'autres ne sont que des mesures floues et sont difficiles à caractériser à partir d'indices visuels, objectifs et élémentaires. C'est le cas par exemple des joues dont les limites sont difficiles à fixer, même pour un observateur humain. Cependant, ces mesures sont importantes puisqu'un humain a plus de facilités à manipuler des données floues que des données précises (les joues sont gonflées).

#### 1.4.1 FACS

En 1978, Ekman et Friesen présentent un système de codification *manuelle* des expressions du visage. Leurs travaux d'observation leur permettent de décomposer tous les mouvements *visibles* du visage en terme de 46 *Actions Unitaires* (*Action Unit*, qui correspond aux *actions faciales* définies plus haut) qui décrivent les mouvements *élémentaires* des muscles. N'importe quelle mimique observée peut donc être représentée sous le forme d'une combinaison d'*Actions Unitaires*.

Ce système de codage est connu sous le nom de «Facial Action Coding System»(FACS). FACS s'est imposé depuis comme un outil puissant de description des mimiques du visage, utilisé par de nombreux psychologues.

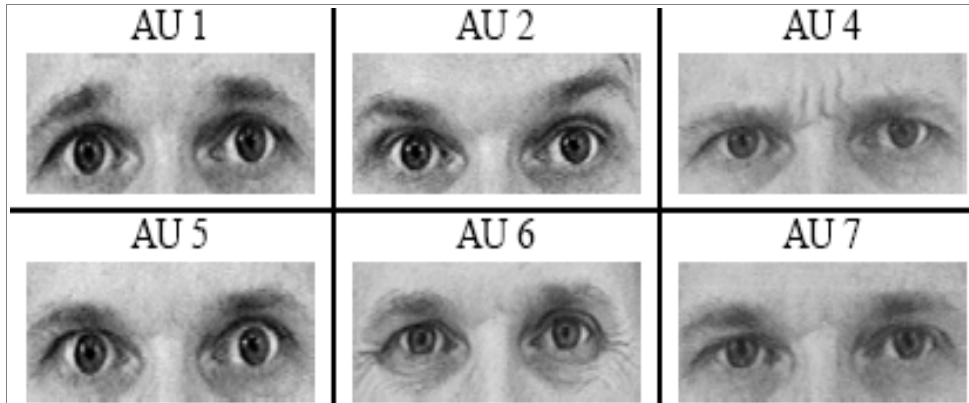


Figure 1.1: Exemples d'Action Units du haut du visage

#### 1.4.2 FACS+

Bien que FACS soit un système de description bénéficiant d'une grande maturité (environ vingt années de développement), il souffre cependant de quelques inconvénients :

- **Complexité** : on estime qu'il faut 100 heures d'apprentissage pour en maîtriser les principaux concepts,
- **Difficulté de manipulation par une machine** : FACS a d'abord été créé pour des psychologues, Certaines mesures restent floues et difficilement évaluables par une machine.
- **Manque de précision** : les transitions entre deux «états» d'un muscle sont représentées de manière linéaire, ce qui est une approximation de la réalité. En particulier les mesures temporelles de l'activation des muscles faciaux (onset, apex et offset) ne sont pas mises en évidence.



Figure 1.2: Exemples d'Action Units du bas du visage

### 1.4.3. 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. Cette description est la suivante[HAM06] :

- **Joie** : la bouche s'ouvre, les commissures se retirent en arrière en direction des oreilles, les sourcils sont décontractés ;
- **Tristesse** : les coins intérieurs des sourcils se courbent vers le haut, les yeux se ferment légèrement, la bouche est décontractée;
- **Colère** : les coins intérieurs des sourcils s'abaissent ensemble, les yeux s'ouvrent largement, les lèvres se serrent l'une contre l'autre ou bien elles s'ouvrent pour laisser apparaître les dents;
- **Peur** : les sourcils se lèvent ensemble et leur partie intérieure est courbée vers le haut, les yeux sont contractés et en état d'alerte;
- **Dégoût** : la lèvre supérieure se lève et se courbe souvent de manière asymétrique, les sourcils et les paupières sont décontractés;
- **Surprise**: les sourcils se lèvent, les paupières supérieures s'ouvrent, la bouche s'ouvre, les paupières inférieures sont relâchées;

Afin de traduire numériquement toutes ces descriptions, un ensemble de distances particulières sur chaque squelette a été définie : la Figure 6 décrit toutes les distances  $D_i$  considérée.

$D_2$  et  $D_7$  donnent une mesure de la distance entre les yeux et les sourcils.  $D_6$  mesure la distance entre les yeux et la bouche.  $D_3$  et  $D_4$  mesurent le degré d'ouverture de la bouche,  $D_1$  mesure le degré d'ouverture des yeux.

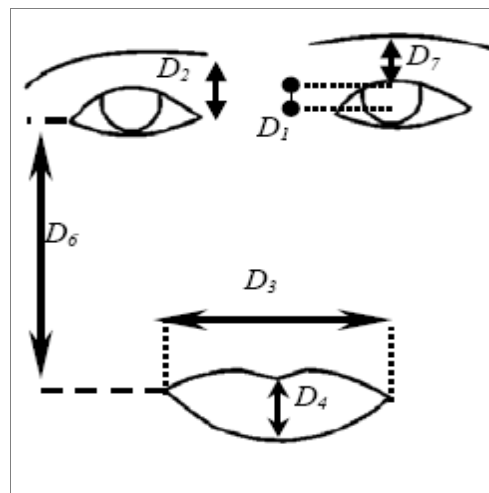




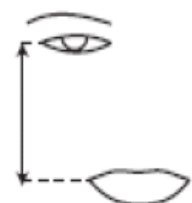


Figure 1.3: Définition des distances  $D_i$

D1		Distance entre la paupière supérieure et inférieure
D2		Distance entre le coin intérieur de l'œil et le celui du sourcil
D3		Largeur de bouche.
D4		Ouverture de bouche en hauteur
D5		Distance entre un coin de la bouche et celui du coin extérieur de l'œil

**Tableau 1.1:Tableau de définition de distance**

la liste des critères associées à chaque description d'émotions :

- **Joie:** {D<sub>4</sub> augmente}, {D<sub>3</sub> diminue ET D<sub>6</sub> diminue}, {les autres distances restent constantes}
- **Tristesse:** {D<sub>2</sub> augmente ET D<sub>7</sub> diminue}, {D<sub>1</sub> diminue}, {les autres distances restent constantes}
- **Colère :** {D<sub>2</sub> diminue}, {D<sub>1</sub> augmente}, {D<sub>4</sub> diminue ou D<sub>4</sub> augmente}
- **Peur :** {D<sub>2</sub> augmente ET D<sub>7</sub> augmente mais plus que D<sub>2</sub>}, {?}
- **Dégoût :** {D<sub>3</sub> augmente ET D<sub>4</sub> augmente}, {les autres distances restent constantes}
- **Surprise :** {D<sub>2</sub> augmente}, {D<sub>1</sub> augmente}, {D<sub>4</sub> augmente}, {les autres distances restent constantes} .

## Chapitre 2

# Détection et segmentation des composantes faciales

Les méthodes de détection et la segmentation des composantes faciales sont abordées dans plusieurs travaux. Nous évoquerons en détail dans ce chapitre deux méthodes de ses travaux, le *EyeMap/MouthMap*[ABD02] et la *transformation markovienne*.

Les images de tests sont empruntées à la base de données de visage JAFFE<sup>1</sup>. Ces images sont représentées en mémoire par 140\*115 pixels répartis en sept classes d'expressions faciales universelles.



Exemples de la base de visage JAFFE

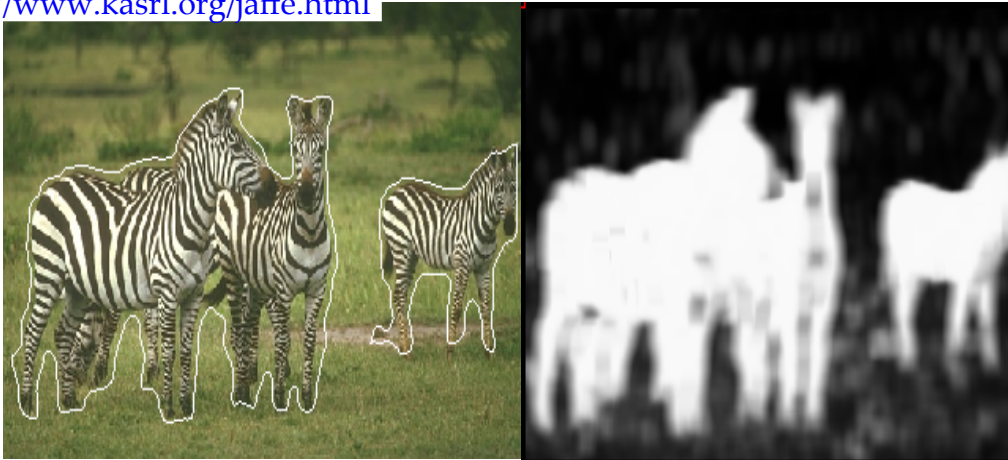
### 2.1 La transformation markovienne:

Cette méthode repose sur la caractérisation des interactions entre pixels liées à un voisinage local, basées sur les *marches aléatoires locales sur graphe*. Les principaux avantages de cette approche sont:

- a. Il capture bien les structures homogènes au sens de l'information visuelle de haute fréquence dans l'image.
- b. Un outil efficace pour l'exploitation locale du graphe correspondant à l'image, qui est globalement dense.

c. Les propriétés spectrales du graphe construit fournissent une mesure simple de la texture.

1 <http://www.kasrl.org/jaffe.html>



**Figure 2.1: Transformation markovienne**

Cette méthode peut aussi être utilisée pour la caractérisation des composantes faciales et les actions faciales sur les différents types d'image (intensité, couleur).

### **2.1.1 Description de l'algorithme:**

L'idée de base de ce descripteur est d'analyser les valeurs propres formées directement à partir de fenêtres de niveaux de gris ou de couleur locaux.

Brièvement, le descripteur est de nature unidimensionnel, formé de la somme des valeurs singulières/propres, et qui fournit l'information sur la dépendance linéaire entre des lignes et les colonnes de la matrice locale des niveaux de gris/couleur liés à chaque pixel. La procédure est répétée pour tous les pixels qui composent l'image.



Soit une image  $X = \{x_1, x_2, \dots, x_m\}$ ,

Pour chaque  $V_{w \times w}(x_i)$  :

■ Construction et décomposition de la matrice de transition de Markov de chaque fenêtre locale de  $x_i$ . Le résultat est les fonctions propres de la matrice de transition.

$$\left\{ \begin{array}{l} 1 = \lambda_0 \geq \lambda_1 \geq \dots \geq 0 \\ \text{et} \\ \{\psi_j\}_{j=1, w} \text{ telque : } P\psi_j = \lambda_j\psi_j. \end{array} \right.$$

■ Remplacer  $x_i$  par la somme des valeurs singulière.

$$\Gamma(x_i) = \sum_{h=r}^w \hat{\lambda}_h.$$

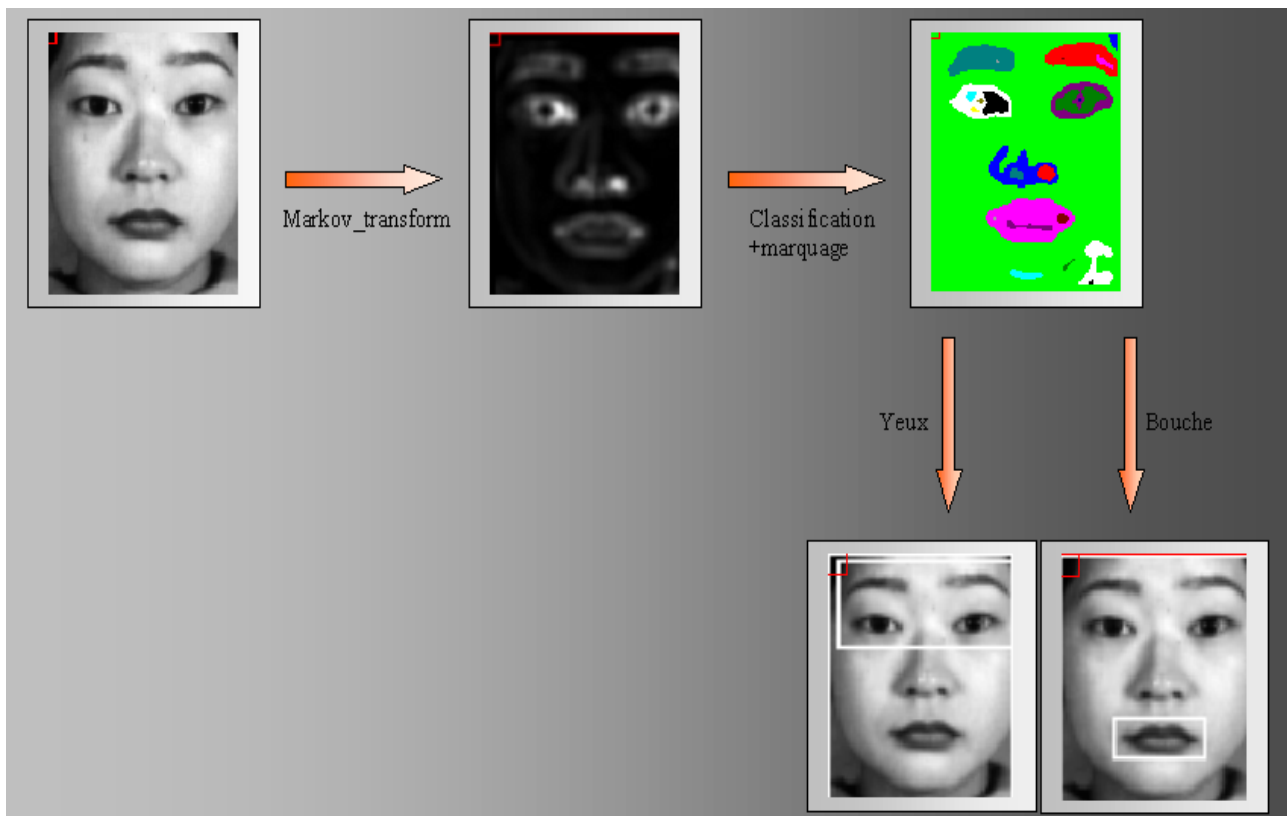
Tableau 2.1: L'algorithme de la transformation markovienne.

### 2.1.2 Détection et segmentation des composantes faciales

Afin de segmenter et d'extraire les composantes faciales (zones d'intérêts), une chaîne de traitement qui a été mise en point fondé en premier lieu sur la «classification des pixels». Par conséquent, les composantes faciales sont obtenues avec les étapes suivantes :

- Normalisation de l'image: la normalisation sert à atténuer les changements d'apparence du visage qui sont causés par la position de la tête ou les variations d'éclairage.
- Transformation markovienne: pour caractériser les composantes faciales.
- Classification et marquage: la classification sert à séparer les structures détectées du fond. Le marquage sert à étiqueter les structures détectées.
- Les composantes faciales en question seront les composantes de la taille la plus grande.





**Figure 2.2: La détection et la segmentation des composantes faciales par la transformation markovienne.**

## 2.2. La carte des yeux et de la bouche EyeMap/MouthMap

EyeMap/MouthMap est une méthode basée sur un modèle de couleur dans l'espace YCbCr, et qui consiste à établir des relations entre les différentes composantes de cet espace.

[Abdel & all 02] ont remarqués que les yeux sont caractérisés par les hautes valeurs de la composante bleus et et les valeur basses de la composante rouge, aussi bien que par beaucoup de pixels sombres et brillants (image dans l'espace YcbCr en dessous). Ces faits ont étaient utilisés pour concevoir une carte qui donne une information sur la position des yeux sur un visage.

Cette carte est construite par la multiplication de deux cartes élémentaires, l'une bassée sur les propriétés chrominances de l'image(EyeMapC) et l'autre sur les propriétés lumineuses(EyeMapL).

Une approche semblable est appliquée pour produire une carte de la bouche. La carte est construite en se basant sur les pixels de la bouche qui ont des valeurs Cr grandes et des valeurs Cb basse.

### 2.2.1 Description de l'algorithme EyeMap/MouthMap

Nous résumons le calcul de la carte de EyeMap et la carte de MouthMap par les formules suivantes:

$$EM_C = \frac{1}{3} \left( C_b^2 + \bar{C}_r^2 + \frac{C_r}{C_b} \right) \quad (1)$$

$$EM_L = \frac{Y_{dil}(p, q)}{Y_{er}(p, q) + 1} \quad (2)$$

$$EyeMap = EM_C \times EM_L \quad (3)$$

- $C_b$ ,  $C_r$  et  $Y$ : sont les composantes couleur de l'espace de couleur YCBCR.
- $\bar{C}_r$  : la négation du canal  $C_r$  ( $255 - C_r$ ).
- $Y_{dil}$  et  $Y_{er}$  : présente la composante  $Y$  dilatée et érodée respectivement.

$$MM = C_r^2 \cdot \left( C_r^2 - \eta \cdot \frac{C_r}{C_b} \right)^2 \quad (4)$$

$$\eta = 0.95 \cdot \text{avg}(C_r^2) / \text{avg}\left(\frac{C_r}{C_b}\right) \quad (5)$$

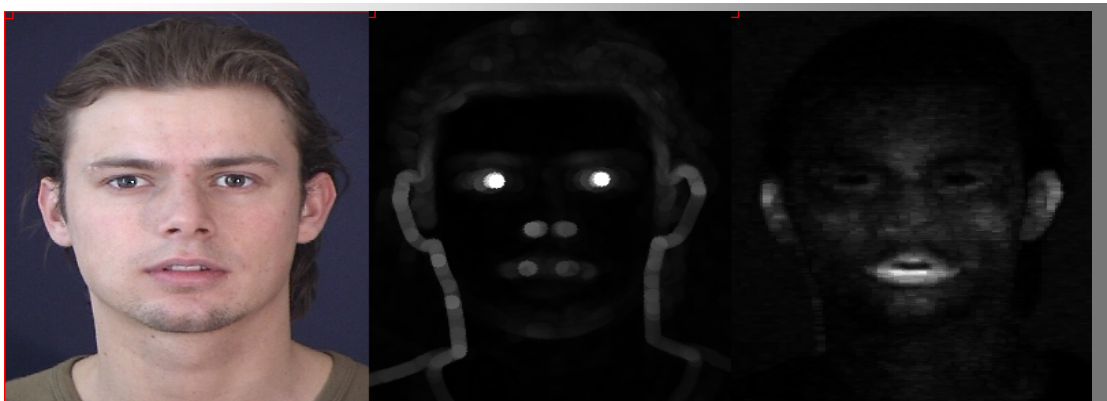


Image couleur

EyeMap

MouthMap

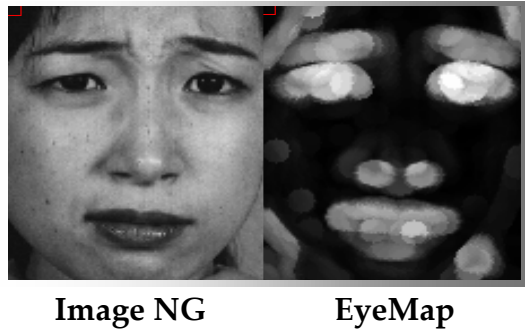
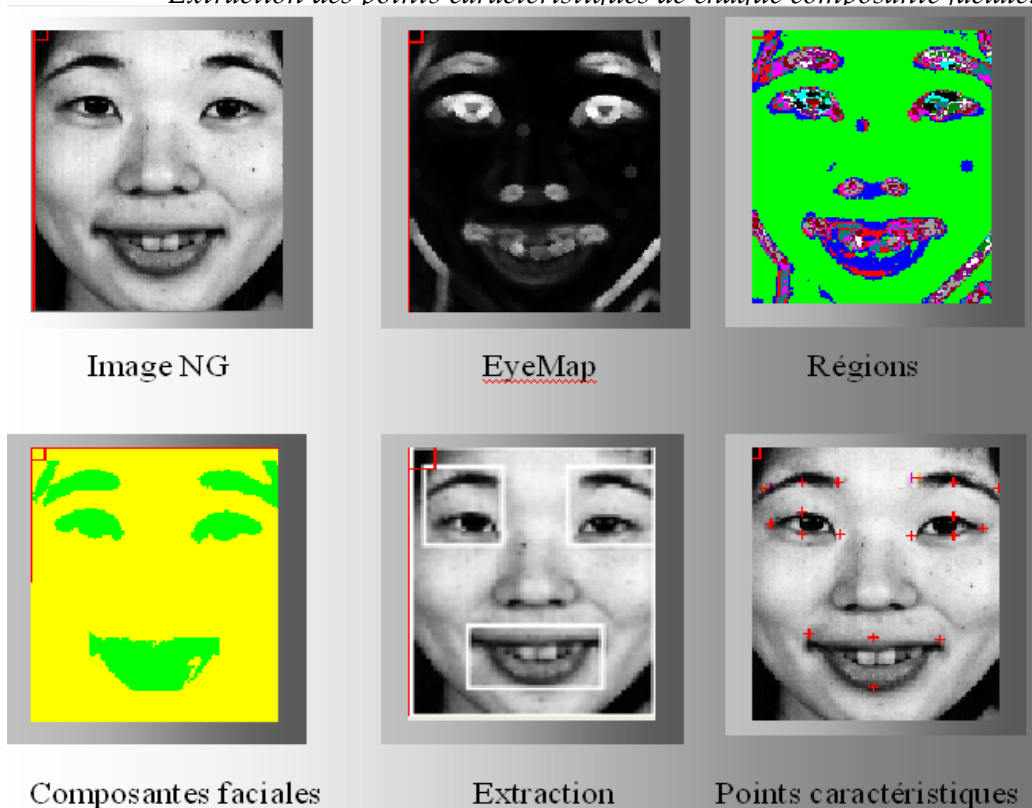


Figure 2.3: EyeMap/MouthMap sur une image de visage couleur et image de visage en niveau de gris

### 2.2.2 Détection et segmentation des composantes faciales

Les principales démarches de cet algorithme reposent essentiellement sur la segmentation par classification de pixels et l'extraction des composantes connexes (composantes faciales) par un balayage vertical et horizontal de l'image segmentée. En outre, les points caractéristiques correspondent aux extrémités de rectangle englobant chaque composante faciale. En conséquence, l'obtention des composantes faciales repose sur le traitement suivant:

- *Normalisation de l'image: la normalisation sert à atténuer les changements d'apparence du visage qui sont causés par la position de la tête ou les variations d'éclairage.*
- *EyeMap: pour localiser les composantes faciales.*
- *Segmentation par classification des pixels: la classification sert à séparer les structures les plus contrastées.*
- *Extraction des composantes faciales par un balayage vertical et horizontal de l'image segmentée.*
- *Extraction des points caractéristiques de chaque composante faciale.*



**Figure 2.4: La détection et la segmentation des composantes faciales par EyeMap**

## Chapitre 3

### **Catégorisation des composantes faciales**

Après avoir obtenues les composantes faciales sous forme de vecteurs d'intensités(transformation markovienne, EyeMap) et de distances(points caractéristiques), on s'intéresse dans ce chapitre à trouver une projection de ces vecteurs, qui sont distribués dans un espace de dimension grand, dans un espace 2D. Cette

projection doit être estimée en ne considérant qu'un nombre fini de vecteurs et en ne connaissant pas à l'avance leurs coordonnées dans l'espace d'arrivée. Il s'agit par conséquent d'un problème particulier du domaine informatique de l'*apprentissage non supervisé*.

Ces problèmes sont classiquement abordés par les techniques de sélection de variables et de réduction de dimension, qui visent à trouver des structures intrinsèques de dimension plus réduite cachée dans les observations dimensionnelles, dont l'analyse en composantes principales (ACP) est un exemple bien connu.

Récemment, une nouvelle famille de méthodes est apparue, fondée sur des approximations globales de données: citons les algorithmes *Isomap* [TEN00], *LLE* [ROW00], qui font l'objet de nombreuses recherches (voir par exemple [GEN05]), et les marches aléatoires sur graphe (graphe de Laplacien) [TEN00][ROW00] qui vise à approximer localement les données et qu'on va l'expérimenter en cours de cette partie.

### 3.1 Les marches aléatoires sur graphe (Graphe de Laplacien) [CHU97]:

Le premier point intervenant dans notre construction doit voir les vecteurs obtenus dans la phase de segmentation  $X = \{x_1, x_2, \dots, x_n\}$  comme étant les nœuds d'un graphe pondéré symétrique  $G = (V, E)$  où chaque point de l'ensemble  $X$  correspond à un nœud de ce graphe. La connectivité entre chaque deux nœud  $x$  et  $y$  de  $G$  est mesurée par un poids non-négatif  $w(x, y)$  qui reflète la similitude (interaction) entre  $x$  et  $y$ .

Dans notre cas, la similitude entre  $x$  et  $y$  peut être mesurée en termes d'intimité par le *noyau Gaussien*:

$$w(x, y) = \exp\left(-\frac{\|x - y\|^2}{\varepsilon}\right).$$

où  $\varepsilon$  est un paramètre d'échelle.

Ce choix correspond à la certitude que  $x$  et  $y$  seront numériquement connectés s'ils sont suffisamment proche tandis que d'autres mesure de la distance euclidienne peuvent être aussi employées telle que le *MstDual*. Cette matrice désormais appelée matrice d'affinité.

Etant donné la matrice d'affinité de notre ensemble de données, nous allons nous intéresser à un processus de marche aléatoire dans le graphe  $G$  construit. Un marcheur est localisé sur un sommet et se déplace 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  $i$  au sommet  $j$  est à chaque étape:

$$P_{ij} = \frac{W_{ij}}{d(i)}$$

(où  $W$  est la matrice d'affinité du graphe  $G$  et  $d(i)$  le degré du sommet  $i$ ).

Ceci définit la matrice de transition  $P$  de la chaîne de Markov correspondante. Nous pouvons aussi écrire  $P = D^{-1}A$  en introduisant la matrice diagonale des degrés des sommets  $D$  ( $D_{ii} = d(i)$  et  $D_{ij} = 0$  pour  $i \neq j$ ).

Après avoir construit le graphe Markovien, on s'attache maintenant aux propriétés spectrales liées à ce graphe, et plus précisément aux intérêts mathématiques des vecteurs propres.

La décomposition spectrale de la matrice  $P$  donne un ensemble de valeurs propres

$1 = |\lambda_0| > |\lambda_1| > |\lambda_2|, \dots, |\lambda_k| = 0$ , engendrent elles même un ensemble de vecteurs propres, soient :

$$\psi_j \in L_2(V)$$

solutions de:

$$P \psi_j = \lambda_j \psi_j .$$

Soulignons que de nombreux travaux ont montré ou constaté que les propriétés structurelles du graphe sont capturées principalement par les vecteurs propres de  $P$  liés aux plus grandes valeurs propres [SAR96,SHI97] et leur visualisation fournissent une représentation des données dans un espace où il est bien séparé et peut facilement être groupé. Chose qu'on espère de la voir dans notre expérimentation.



**Figure 3.1: La segmentation d'image par le clustering spectral.**

Par conséquent, on résume dans le schéma suivant les principale démarche de notre système de catégorisation:

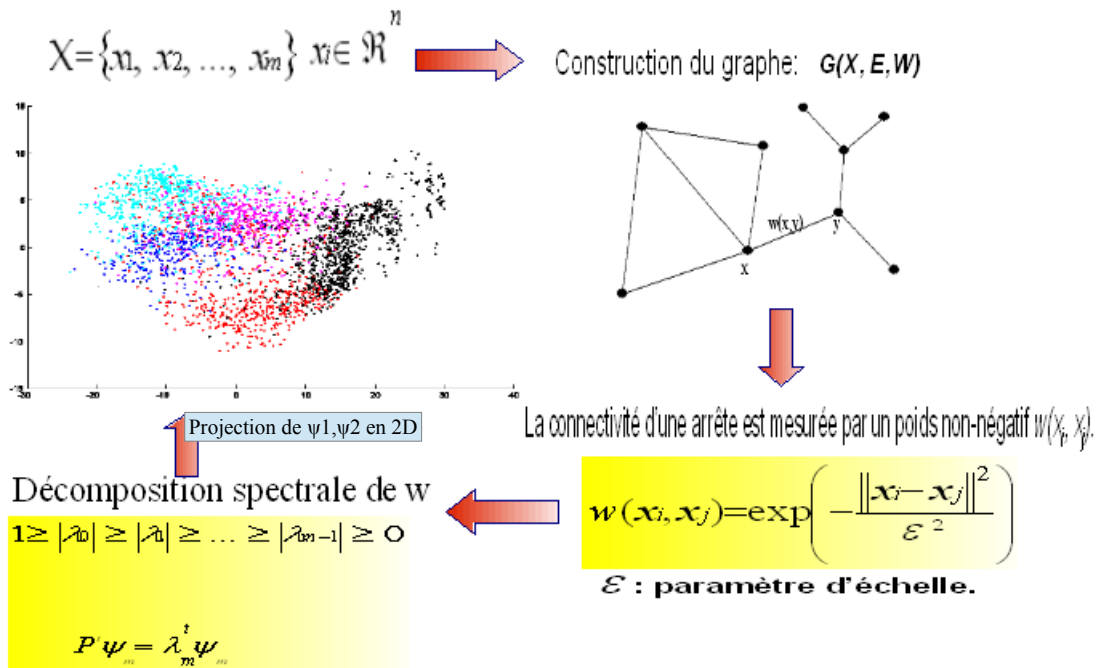


Figure3.2. Les étapes de l'Algorithme de clustering spectral

### 3.2 Tests et résultats

Après avoir présenté, théoriquement, les propriétés spectrales des marches aléatoires sur graphe, nous souhaitons tester leurs adéquations pour la révélation de l'information géométrique liées à l'ensemble de vecteurs obtenues dans le chapitre précédent et qui sont de nature non linéaire .

Pour cela, nous présentons dans cette section deux applications démonstratives de catégorisation basées sur les *vecteurs d'intensités* et les *vecteurs de distances* ont été mise en œuvre.

#### 3.2.1 Catégorisation avec les vecteurs d'intensités:

Motivé par l'interprétation du paragraphe (1.2..1) (indépendance des muscles) on a découpé l'analyse en deux zones, la zone des composantes faciales yeux et la zone de composante faciale bouche.

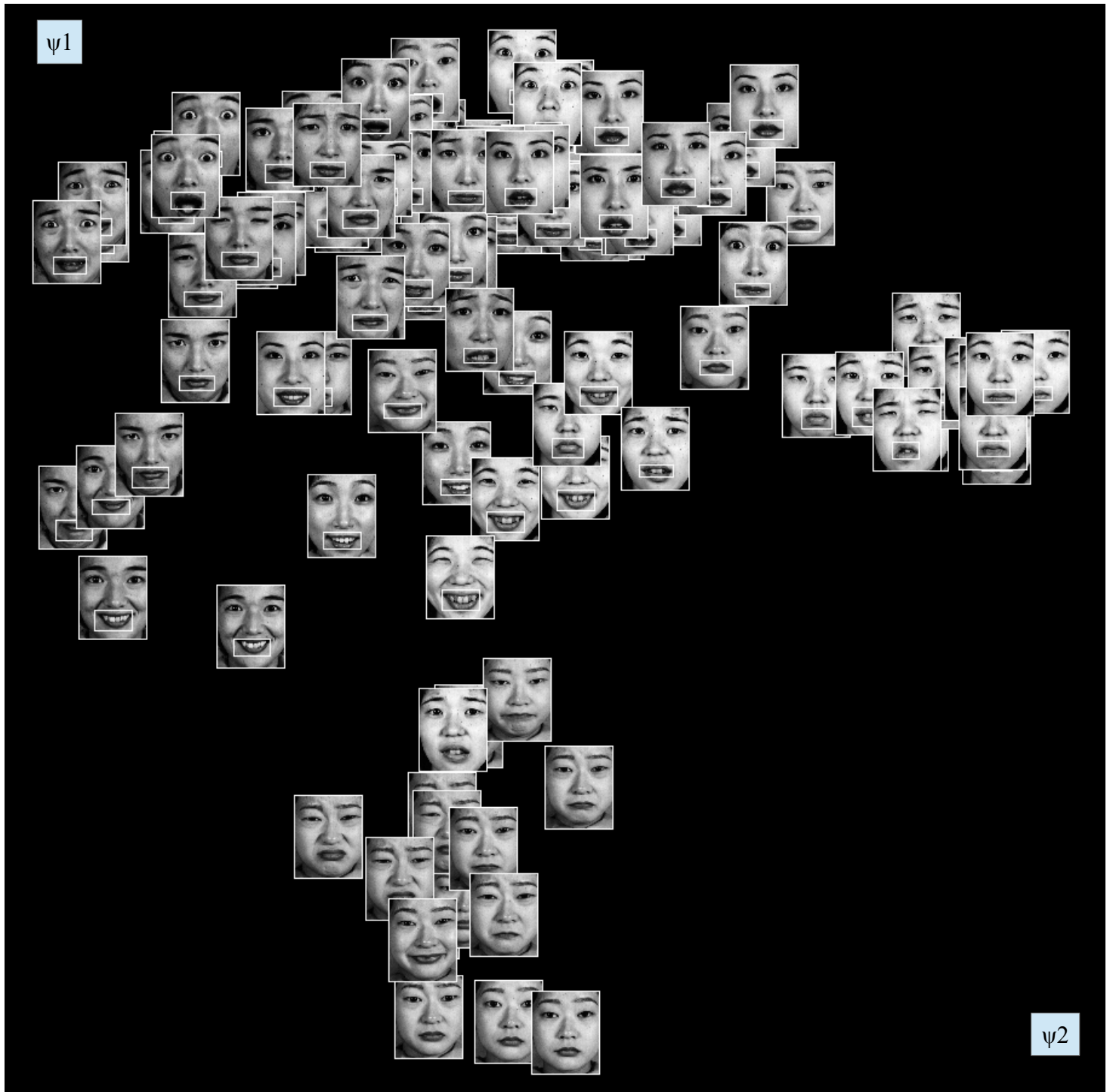


Figure 3.3: La projection des deux premiers vecteurs propres dans l'espace 2D selon la forme de la bouche.



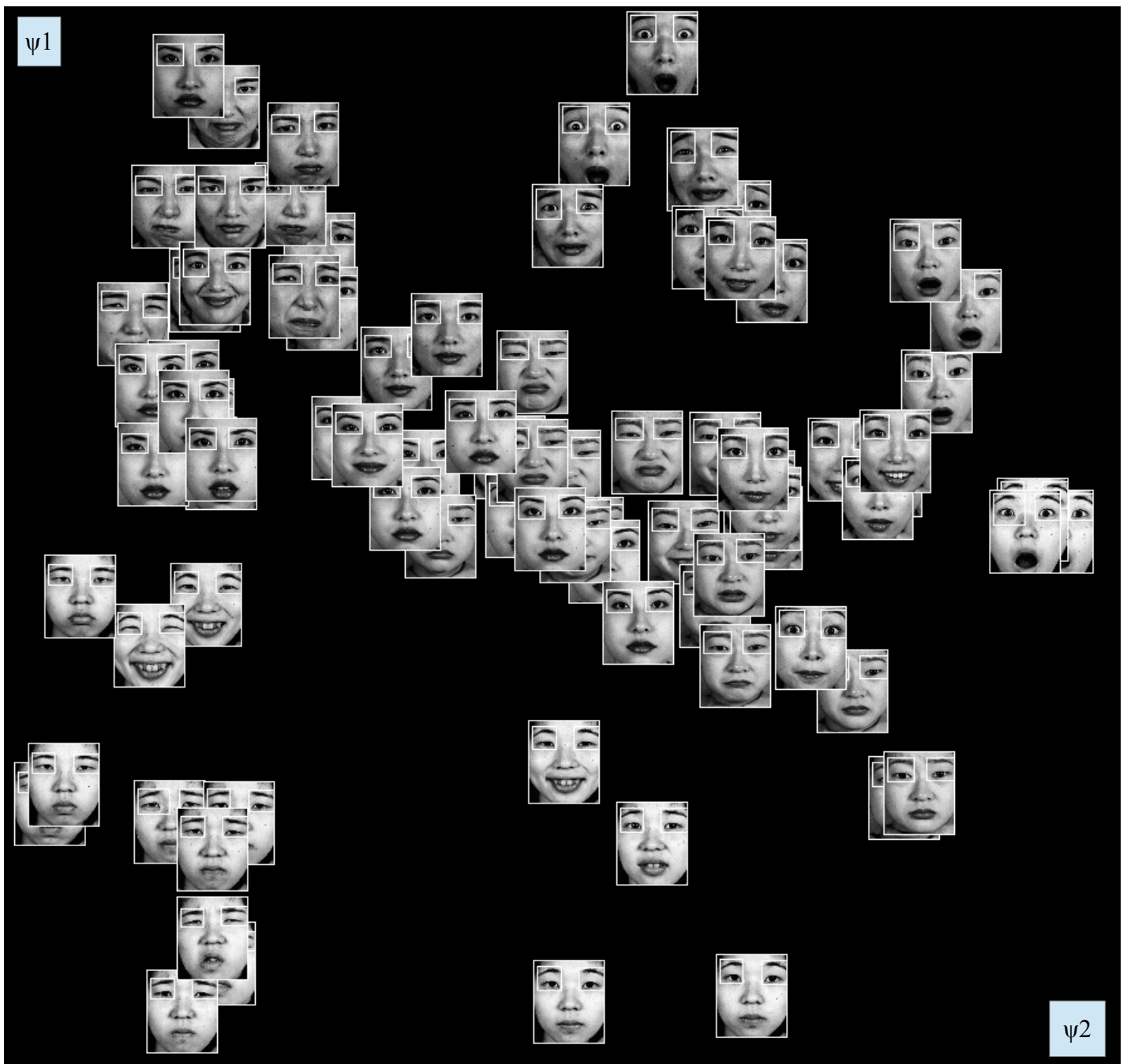


Figure 3.4: La projection des deux premiers vecteurs propres dans l'espace 2D selon la forme des yeux.

Basant sur le vecteurs d'intensités de bouches, on aperçoit un espèce de groupe lié à l'expression faciale «Sourire» et à l'expression faciale de «Surprise»(Figure 1). Cette situation est du à la différence d'intensité existe entre les deux formes correspondant à la composante bouche ouvert essentiellement les parties milieu de bouches. En revanche, le reste de composantes correspondant aux bouches fermées se focalise en groupe selon la forme.

De même, la projection de la composante faciale de yeux, montre une organisation selon la forme, et ça est du de fait que la déformation entre l'oeil et celui du sourcil sont

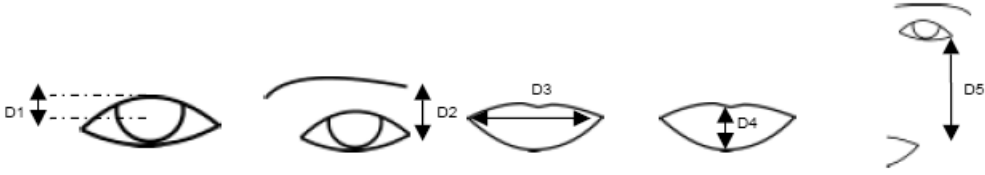
estimés par une simple différence d'intensité entre chaque deux composantes faciales. Cependant, il s'avère important de prendre en compte la variation de distance entre les deux mimiques faciales sourcil et yeux, et faire lié cette distance à la déformation de la bouche. D'où la nécessité des distances issues de la l'information des points caractéristiques lié à chaque composante faciale(1.2.2).








### 3.2.1 Catégorisation avec les vecteurs de distances:

À partir de ces distances caractéristiques, on va ajouter des états symboliques où on associe l'un des trois états symboliques suivants à chacune des valeurs de distances .

- 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 du 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.

Après cela, chaque expression faciale est caractérisée par une combinaison d'états symboliques. Nous avons défini une combinaison d'états pour chaque expression.



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 3.1:Etats symbolique associés à chaque expression**

Fondé sur ces états symbolique, la construction de la matrice de similarité prendre un autre sens où les lignes va se présenter les expressions neutre de chaque visage et les colonnes l'ensembles de visages qui vont virifier les états décrits dans le tableau en dessus. En conséquence, le nouveau calcul de la matrice de similarité se ferai de la manière suivante:

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

$\Omega_1$  : ensemble des composantes faciales correspondantes aux visages d'expression neutre.

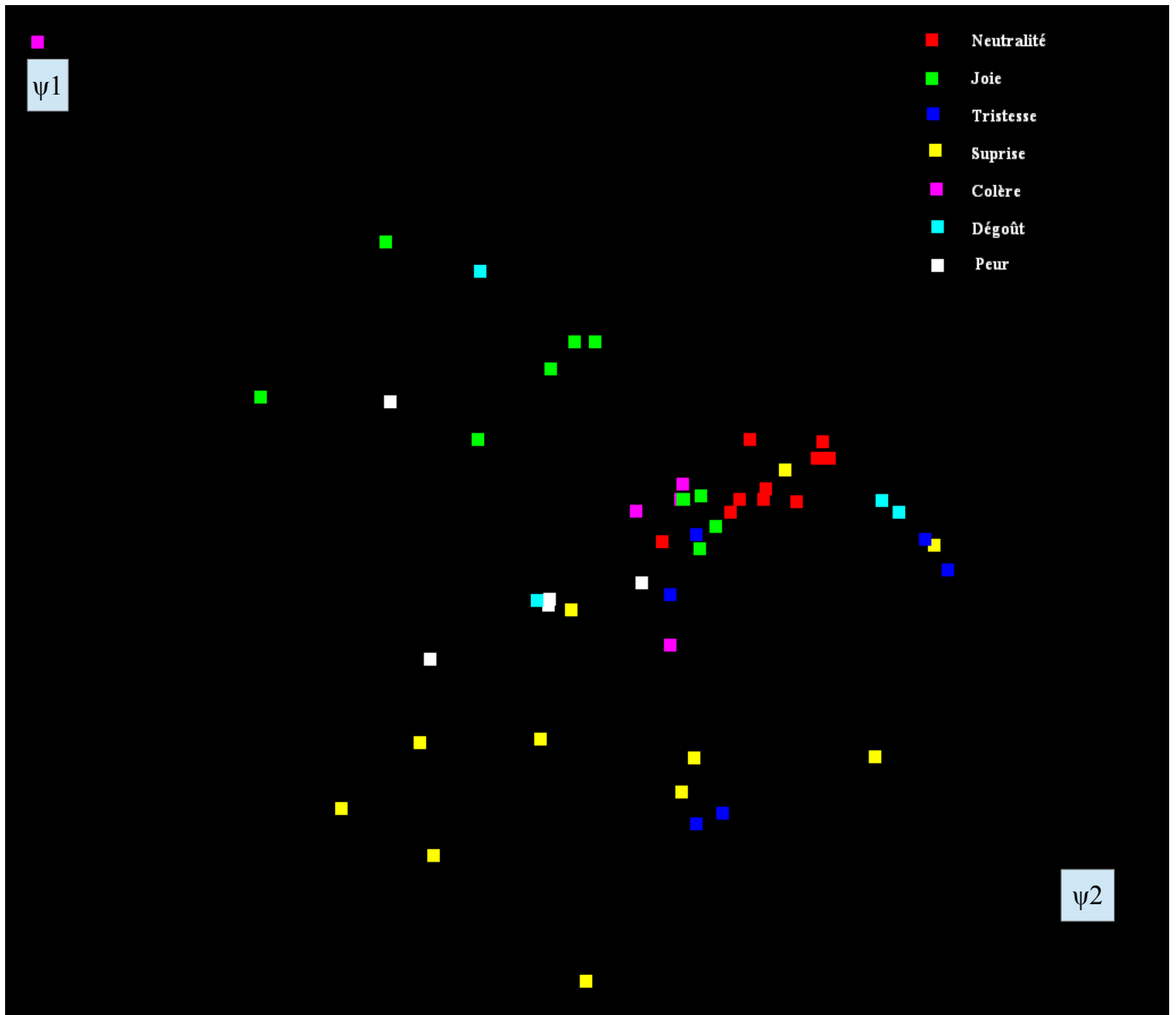
$\Omega_2$  : ensemble des composantes faciales correspondantes aux visages de différentes expressions.

$$W' = W.W^t$$

Figure 3.5: Le nouveau calcul de la matrice de similarité.



Figure 3.6: La projection des deux premiers vecteurs propres dans l'espace 2D selon les distances issues de la l'information des points caractéristiques.



**Figure 3.7:La répartition des composantes faciales dans l'espace 2D.**

Visuellement, la répartition des visages donne un résultat plus au moins acceptable dans le cas d'expression neutre, joie et surprise. En revanche, dans certains cas, des erreurs de confusion sont apparues (voir figure ci-dessus) entre les différents types d'expression. Ces confusions sont dûes principalement à l'imprécision de la détection automatique de points caractéristiques de chaque composante faciales. Les fausses détections de ces points sont suscités par des erreurs de segmentation des composantes faciales, notamment avec les composantes faciales des bouches ouvertes et les sourcils.

Un grand avantage, c'est que l'expression neutre se trouve sous forme d'un groupe au milieu du nuage de points, ce qui confirme la validité de la proposition concernant le nouveau calcul de la matrice de similarité décrite précédemment.

## Conclusion

Dans ce travail, nous avons présenté notre contribution en répondant aux deux questions qui se sont levées pour la catégorisation des expressions faciales. La première partie de ce travail, présente deux algorithmes de la détection et la segmentation des composantes faciales des yeux et de la bouche. Les deux algorithmes ont montré une meilleure efficacité pour la détection que pour la segmentation.

Le premier modèle qui est basé sur la partition locale de graphe de grande dimension a montré son efficacité du fait que ce modèle caractérise non seulement les composantes faciales mais aussi les déformations musculaires autour de ces composantes. Cette 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 outre, le deuxième modèle qui est basé sur la variation de luminance de pixel, permet aussi de focaliser les composantes faciales. En revanche cette méthode montre des limites dans les zones qui montrent une variation d'intensité sombre (exemple de bouches ouvertes). Donc, il est nécessaire de s'orienter vers les approches de contours actifs....

En se basant sur les composantes faciales issues de la phase de la segmentation, un algorithme de clustering spectral (Graphe Laplacien) a été choisi afin de visualiser la répartition de ces composantes qui sont de nature dimensionnelle. 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 extrêmes correspondant aux composantes faciales.

La projection des deux vecteurs propres, calculés par la décomposition matricielle de la matrice de similarité, calculée d'une manière différente du mode classique, construite à partir des vecteurs d'information des composantes faciales, a montré que le groupement des expressions faciales «Sourire», « Surprise» et «Neutre» est plus au moins réalisé malgré des erreurs de confusion dû principalement aux erreurs de la segmentation.

L'amélioration de la segmentation de visage est possible avec les modèles basés sur les courbes notamment la nouvelle approche inventée par Zakia Hammal [HAM06] et qui vise à extraire les contours caractéristiques du visage qui sont suffisants pour la catégorisation des expressions faciales.

## Bibliographie

- [ABD02] Mahamed Abdel-Mottalebz, Rein-Lien Hsuy, Anil K. Jainy. Face Detection in Color Images. 2002
- [ABO06] Bouchra Abboud, Franck Davoine, Mô Dang. Reconnaissance et synthèse d'expressions faciales par modèle d'apparence Facial expression recognition and synthesis based on an appearance model. 2006.
- [AND01] Andrew Y. Ng, Micheal I. Jordan, Yair Weiss. On Spectral Clustering Analysis and an algorithm. NIPS 2001: 849-856
- [BEN03] Y. Bengio, P. Vincent, J.-F. Paiement, O. Delalleau, M. Ouimet, and N. Le Roux. Spectral Clustering and Kernel PCA are Learning Eigenfunctions. Technical Report 1239, Département d'informatique et recherche opérationnelle, Université de Montréal, 2003.
- [CHE03] Ya-Chun Cheng, Shu-Yuan Chen. Image classification using color, texture and regions. *Image and Vision Computing* 21 (2003) 759–776
- [CHU97] F. R. K. Chung, *Spectral graph theory*, CBMS regional conference series in mathematics, AMS 1997
- [HAM06] Zakia Hammal, Facial Features Segmentation, Analysis and Recognition of Facial Expressions using the Transferable Belief Model, thèse de Zakia Hammal, Juin 2006.
- [HAM06] Zakia Hammal, Alice Caplier. Facial features boundaries segmentation and facial expressions analysis. 2006.
- [HAM04] Z. HAMMAL, N. EVENO, A. CAPLIER, PY. COULON. Parametric models for facial features segmentation. 2006.
- [HAR01] David Harel, Yehuda Koren. Clustering Spatial Data Using Random Walks, 2001.
- [GEN05] Xin Geng, De-Chuan Zhan, and Zhi-Hua Zhou, *Member, IEEE*. Supervised Nonlinear Dimensionality Reduction for Visualization and Classification. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*.
- [MEI01] Marina Meilä et Jianbo Shi. Learning segmentation by Random Walk. *Neural Information Processing Systems, NIPS*, 2001
- [MER06] Hugo MERCIER. Analyse automatique des expressions du visage. Application à la Langue des Signes. Travail de DEA 2006, TOULOUSE. FRANCE.
- [SAR96] S. Sarkar and K.L. Boyer. quantitative measures of change based on feature organization: eigenvalues and eigenvectors. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1996.
- [SHI97] J. Shi and J. Malik. Normalized cuts and image segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 731-737, 1997.
- [TEN00] A global geometric framework for nonlinear dimensionality reduction. J. B. Tenenbaum, V. De Silva and J. C. Langford (2000). *Science*, 290 (5500), 2319-2323.
- [WEI99] Yair Weiss. Segmentation using eigenvectors: a unifying view. *Proceedings IEEE International Conference on Computer Vision* p. 975-982, 1999.