



**HAL**  
open science

## Construction de masques faciaux pour améliorer la reconnaissance d'expressions

Taner Danisman, Ioan Marius Bilasco, Jean Martinet, Chaabane Djeraba

► **To cite this version:**

Taner Danisman, Ioan Marius Bilasco, Jean Martinet, Chaabane Djeraba. Construction de masques faciaux pour améliorer la reconnaissance d'expressions. *COMpression et REpresentation des Signaux Audiovisuels*, May 2012, Lille, France. pp.17. hal-00812319

**HAL Id: hal-00812319**

**<https://hal.science/hal-00812319>**

Submitted on 29 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Construction de masques faciaux pour améliorer la reconnaissance d'expressions

Taner Danisman, Ioan Marius Bilasco, Jean Martinet, Chaabane Djeraba

Université Lille 1 / Laboratoire d'Informatique Fondamentale de Lille  
{danisman.taner}@gmail.com, {marius.bilasco, jean.martinet, chabane.djeraba}@lifl.fr

## Résumé

*Ce travail propose une méthode pour détecter de manière automatique les régions qui contribuent le plus à une bonne classification des visages par rapport à des expressions prédéfinies : joie, surprise, etc. Notre méthode détermine les régions ayant le plus, (respectivement le moins) de pouvoir discriminant en utilisant un réseau de neurones de type MultiLayer Perceptron (MLP). A partir de régions de formes et de tailles quelconques, nous créons des masques à appliquer aux images avant de les classer. Ces masques éliminent les zones de visages non pertinentes pour le processus de classification, en augmentant ainsi la performance du système. Nous avons conduit des expériences sur les bases d'images FERET, GENKI et JAFFE. Les résultats montrent une augmentation du taux de classification en utilisant les masques désignant les pixels d'intérêt.*

## Mots clefs

Visage, Expressions, Masques d'expression, Classification, Réseaux de neurones, MLP

## 1 Introduction

La reconnaissance d'expressions faciales est un thème de recherche de prédilection dans différents domaines incluant la reconnaissance et l'animation des visages, l'interaction homme machine ou l'interaction sociale. Ces dernières années, la recherche autour de ce thème a impliqué de chercheurs issus de divers domaines qui ont concentré leurs efforts sur des méthodes précises et rapides pour la reconnaissance d'expressions faciales.

Les études récentes montrent d'importants progrès en ce qui concerne la classification d'expressions sur les bases d'images connues dans l'état de l'art. Cependant, des problèmes importants subsistent lorsqu'on s'intéresse à des images quelconques, prises dans des conditions naturelles ou extraites depuis des flux vidéos. Une réponse à ce problème consiste à réduire les zones de visages et les types d'information utilisés dans le processus de classification afin de réduire le bruit et augmenter ainsi le taux de reconnaissance. Dans ce sens, la majorité des précédentes études sur la reconnaissance d'expression considèrent le visage et les descripteurs associés comme

une combinaison de régions rectangulaires [1, 2]. Même si ces solutions offrent l'avantage d'une implémentation simplifiée, les régions incluent des informations inutiles et bruitées dans le processus d'apprentissage et de classification. Par conséquent, nous considérons utile d'identifier les pixels d'intérêts (POIs) du visage, les plus pertinents, en vue de reconnaître des expressions. Les regroupements de POIs constituent des masques génériques, non rectangulaires, qui peuvent être utilisés afin d'augmenter les taux de classification.

Dans ce travail, nous avons suivi une approche analytique qui réalise une sélection de pixels et de caractéristiques en recherchant de manière exhaustive les POIs sur tout le visage. Nous construisons un recouvrement complet du visage au moyen de fenêtres de recherche à position et à tailles variables. Pour chaque fenêtre nous évaluons le taux de reconnaissance. Nous étudions ainsi la qualité de chaque pixel du visage par rapport aux taux de classifications pondérés observés dans chacune des fenêtres contenant le pixel. Ainsi, les cartes que nous obtenons peuvent couvrir des régions éparées et pas forcément contiguës, telles que, les zones rectangulaires ou ellipsoïdales communément utilisées dans l'état de l'art. Pour chaque expression faciale, nous créons un masque défini par les POIs correspondants afin d'améliorer les performances de réseau de neurones de type Perceptron Multi-Couche (ang. MLP) [3] utilisé pour reconnaître une émotion. Les tests montrent des meilleurs taux de classification par rapport à des méthodes qui ne s'appuient que sur des masques rectangulaires.

L'article est organisé comme suit. La section 2 présente un aperçu des travaux connexes de l'état de l'art. Le processus de génération de masque, son application à la reconnaissance des expressions faciales, en se concentrant sur la joie, sont décrits dans la section 3. La présentation de résultats obtenus est réalisée dans la section 4, suivie par le bilan et les perspectives ouvertes par ce travail.

## 2 Etat de l'art

La reconnaissance d'expressions faciales, et plus généralement, la reconnaissance d'émotions est un domaine de recherche interdisciplinaire où les émotions peuvent être détectées à l'aide des capteurs physiologiques, des capteurs audio, des capteurs vidéos. Chaque type de capteur apporte ses propres défis relatifs

au traitement du bruit, à la qualité variable de descripteurs utilisés, à la quantité de données à analyser. D'autres facteurs impactent la reconnaissance au niveau de l'analyse des données enregistrées et à leur classification. Nous retenons ici : la taille des descripteurs, la qualité du prétraitement, la fusion a priori ou a posteriori des descripteurs, ainsi que les paramètres du processus de classification. A cette liste peuvent être ajoutés des facteurs spécifiques au domaine de la vidéo et de l'image : diverses conditions de luminosité, changements d'arrière plans, grandes variations dans les parties statiques et dynamiques du visage, les zones d'ombres. L'ensemble de ces éléments peut réduire de manière significative le taux de reconnaissance dans un contexte applicatif non contraint, si des traitements, limitant l'influence du bruit, ne sont pas intégrés dans le processus de reconnaissance. Parmi les éléments évoqués ci-dessous, nous nous intéressons uniquement à la réduction de la dimension des descripteurs, en sélectionnant les éléments apportant le plus d'information discriminante dans la classification.

Les approches analytiques sont largement utilisées dans l'analyse du visage. Elles s'appuient sur la détection des zones spécifiques telles que : les yeux, les sourcils, le nez, la bouche ainsi que des points caractéristiques : les coins des yeux, les coins de la bouches et leur disposition géométrique relative au sein du visage. Les systèmes analytiques sont construits à partir de sous-sytèmes spécifiques à des parties du visage. La classification est réalisée en intégrant les résultats des sous-systèmes [4].

Les méthodes holistiques traitent le visage dans son ensemble sans décomposition en zones. Les approches travaillent sur une représentation globale du visage définie dans un espace de très grande dimension. Dans la littérature, l'Analyse en Composantes Principales (ACP) [5], Analyse Discriminante Linéaire (ADL) [6] et Analyse en Composantes Indépendantes (ACI) [7] sont communément utilisés par les méthodes holistiques afin de réduire la dimensionnalité des descripteurs. Ces outils réduisent la dimensionnalité des données originales en agrégeant uniquement les éléments pertinents dans des nouvelles représentations. Toutefois, la capacité de réduire la dimension des données originales en les projetant dans un nouvel espace de dimensionnalité réduite n'est pas appropriée pour tout scénario holistique, car il peut induire des transformations dénaturant les données initiales.

Pour la reconnaissance d'expression faciale supervisée, nous devrions préserver en l'état les informations sur les zones et les points représentatifs du visage. Ainsi, nous nous intéressons à des solutions qui simplifient la représentation en conservant les caractéristiques initiales.

Une méthode qui s'appuie sur le feedback apporté par des caractéristiques est utilisée dans [8] pour la reconnaissance de visage. Dans cette méthode, les caractéristiques extraites sont analysées dans l'espace

d'origine et l'apport des caractéristiques dans le processus de classification est mesuré afin d'identifier des segmentations pertinentes des données d'origine.

Une méthode holistique pour l'analyse d'expression faciales se concentre sur certaines régions du visage (les yeux, la bouche, etc.) dont la géométrie est affectée le plus par les expressions faciales est présentée dans [9]. L'identification de ces zones a été faite en utilisant ACP.

Dans une étude récente [10], un masque elliptique a été construit afin d'éliminer les pixels de l'arrière-plan du visage. Ce masque est appliqué aux visages avant l'apprentissage et la classification. Ce type de masque élimine notamment les cheveux et la gorge de l'image retenue pour représenter le visage. Toutefois, parmi les pixels préservés par ce masque ils en restent qui n'apportent que peu d'information utile et qui de surcroît peuvent biaiser le processus de reconnaissance. Ainsi, des masques plus précis doivent être construit afin d'éliminer les pixels inadéquats. Dans la section suivante, nous présentons notre méthode de construction de masques génériques pour la reconnaissance d'expressions faciales.

### 3 Masques d'expression faciale

Nous construisons des masques non rectangulaires pour préserver uniquement les pixels apportant le plus d'information discriminante dans le processus de classification d'expressions faciales.

Notre méthode de construction de masques d'expressions faciales positionne de manière exhaustive, sur l'ensemble du visage, des fenêtres rectangulaires à la recherche des zones qui à elles seules apportent des bonnes performances de classification. En considérant l'ensemble de fenêtres rectangulaires, de différentes tailles, apportant, respectivement, les meilleurs et les pires taux de reconnaissance, nous construisons un masque à contour flou sélectionnant uniquement les pixels recouverts par un maximum de régions ayant de bonnes performances de classification et par un minimum de régions ayant des mauvaises performances. Nous supposons que les taux de reconnaissance élevés sont une bonne indication de la valeur intrinsèque des pixels délimités par les zones concernées. Alors que, les taux faibles sont des indications de l'absence d'intérêt des zones concernées.

Pour le processus de classification nous utilisons un réseau de neurones de type MLP (Multi-Layer Perceptron ou Perceptron Multi-Couche). Nous avons entraîné le réseau de neurones à distinguer la joie des autres expressions faciales regroupées toutes dans une deuxième catégorie (non-joie). Le pouvoir prédictif des masques construits est mesuré en termes de taux de mauvaise classification par rapport à l'ensemble d'apprentissage.

La Figure 1 illustre le fonctionnement général de notre méthode. Un premier module de détection permet

d'extraire les visages et de déterminer la position des yeux. Les informations sur la géométrie du visage et la position relative des yeux permettent de corriger l'orientation du visage (le roulis) de telle sorte à ce que le visage soit ramène à la verticale. Afin d'uniformiser l'ensemble de visages utilisés pour l'apprentissage et la classification, nous effectuons une mise à l'échelle (50px x 50 px) et un étalement de contraste.

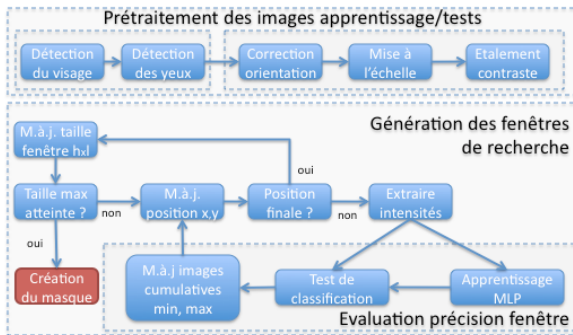


Figure 1 - Aperçu de l'approche

Par rapport à la taille de visage normalisé retenue, le module de génération de fenêtres de recherche explore toutes les possibilités de disposer des fenêtres de taille (l,h) à l'ensemble de positions (x,y) sur un visage normalisé. Chacune des fenêtres générées est ensuite appliquée à l'ensemble des visages utilisé pour l'apprentissage et le test. Nous récupérons les intensités des pixels situés à l'intérieur de la région déterminée par (x,y,x+l,y+h) pour servir à la fois dans le processus d'apprentissage et de test. Par taille de fenêtre, nous retenons les zones ayant les plus bas et plus haut taux de reconnaissance. Ces zones sont ensuite agrégées par le module d'agrégation de masques d'expression faciale. Nous détaillons dans la suite de cette section chacun des modules participant à la construction de masque.

### 3.1 Prétraitement des bases d'apprentissage

Afin de mieux illustrer notre approche, nous utilisons dans cette section des exemples de la base d'images GENKI [11]. La base d'images GENKI-4K de MPLab [11] contient des images organisées en deux classes "smiling" et "non-smiling" couvrant un large spectre de conditions d'illuminations, d'endroits géographiques, de personnes et d'ethnies. Elle contient quatre milles images annotées manuellement. Nous assimilons les images "smiling" à la classe "joie" et les autres à la classe "non-joie".

Comme précisé dans la Figure 1, la première étape consiste à construire l'ensemble d'apprentissage normalisé. Par rapport à la taille des images contenues dans la base GENKI (approximativement 180x192), avant de réaliser la détection du visage, nous augmentons artificiellement la taille des images en réalisant une interpolation linéaire pour arriver à des résolutions de l'ordre de 450x480 en préservant l'aspect de l'image. Les

étapes suivantes du prétraitement sont illustrées dans la Figure 2. Elles consistent dans la détection du visage (B), la rotation et au recadrage du visage (C) en considérant la distance inter-pupillaire (IPD), la mise à l'échelle de la taille (D) et l'étalement du contraste.

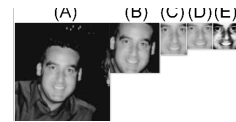


Figure 2 – Etapes du prétraitement pour les images.

La bibliothèque logicielle OpenCV est employée pour détecter et extraire le visage depuis les images initiales en utilisant un détecteur de visage en position frontale [2]. Le détecteur utilise les caractéristiques de type Haar qui considèrent les changements en termes de contraste observés entre groupes de pixels adjacents. Ensuite, la détection des pupilles est réalisée en utilisant un réseau de neurones [12] implémenté dans la bibliothèque logicielle STASM [13]. Par rapport, au visage tout entier, nous définissons des zones de recherche situées dans la partie haute du visage. La position des pupilles sert à réaliser une éventuelle rotation du visage en position verticale.

Nous estimons le roulis du visage en utilisant les positions relatives de pupilles. Nous calculons l'angle entre les centres des pupilles et nous appliquons une rotation en sens inverse. La rotation s'effectue par rapport au centre du visage. Nous corrigeons ainsi les orientations du visage jusqu'à  $\pm 25$  degrés par rapport à l'axe horizontal.

A partir du visage détecté initialement, nous découpons une zone dont la taille dépend de l'IPD afin d'élaguer les zones non-pertinentes telle que les cheveux, les oreilles, la gorge. Le découpage se fait selon une zone de largeur de 1.6 x IPD, de hauteur de 1.9 x IPD et dont l'origine se trouve au milieu du visage. Les valeurs scalaires (1.6 et 1.9) ont été obtenues suite à des nombreux tests expérimentaux. La taille du visage et ensuite ramenée à 50px x 50px en nuances de gris en étalant le contraste.

Les visages ainsi obtenus sont utilisés dans les nombreux réseaux de neurones qui seront construits pour les fenêtres de recherches générées à l'étape suivante.

### 3.2 Fenêtres de recherche

Pour chaque fenêtre de recherche nous construisons un réseau de neurones de type MLP comme illustré dans la Figure 3. Le nombre de neurones dans la couche d'entrée correspond au nombre de pixels de la fenêtre. La première couche intermédiaire contient vingt neurones. La deuxième couche contient quinze. Le nombre de neurones sur les couches intermédiaires a été fixé suite à des expériences que nous ne détaillons pas dans cet article. La couche de sortie contient deux neurones correspondant aux deux classes d'expressions considérés : joie / non-joie.

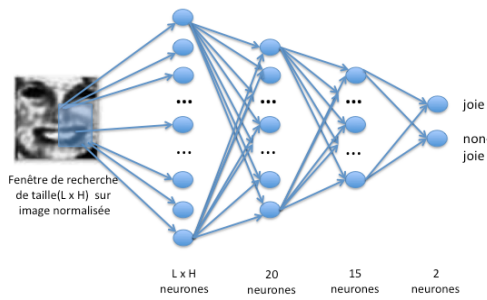


Figure 3 – Structure du perceptron multi-couche utilisé.

L'ensemble d'images normalisées obtenu à l'étape précédente est séparé en deux parties égales. Une moitié sert à entraîner en vue de classification le perceptron. L'autre moitié sera utilisée afin d'évaluer la performance de classification du perceptron ainsi entraîné. La fenêtre de recherche est utilisée comme un masque sur les images considérées. Les valeurs d'intensité des pixels se trouvant couverts par la fenêtre de recherche alimenteront seules, les L x H neurones de la couche d'entrée du perceptron.

Dû au très grand nombre des calculs pour l'ensemble de fenêtres de recherches nous avons essayé de réduire l'espace de recherche en sautant certaines zones de recherche. La fenêtre est avancée vers la droite à chaque fois d'un certain nombre de pixels. Une fois arrivé à l'extrémité droite de la fenêtre, nous appliquons le même décalage en nombre de pixels vers le bas. Suite aux expériences effectuées en modulant le décalage, sans perte de précision notable, nous sautons trois pixels à la fois.

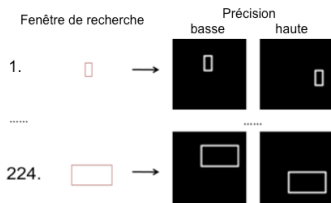


Figure 4 – Position des fenêtres pour les taux les plus haut et les plus bas de classification par taille de recherche

Ainsi pour une image de 50 × 50 pixels, nous avons considéré m=224 de type fenêtres de recherches de tailles différentes. Chaque type de fenêtre balaie ensuite l'ensemble de l'image ce qui produit un nombre total de 14,490 zones d'études distinctes sur les images. Chaque zone donne naissance à la construction d'un perceptron dont la précision de classification est évaluée. Pour chaque fenêtre nous retenons les zones d'études ayant eu le meilleur et le pire taux de classification. Dans la Figure 4, nous montrons les fenêtres de plus petite et de plus grande taille considérées, ainsi que les zones sur lesquelles les taux extrêmes de classification ont été observés.

Toujours dans un souci de réduire le temps passé à construire les masques, nous avons parallélisé le processus d'exploration et traitement des fenêtres. Nous avons implémenté une solution qui exploite au maximum les capacités de huit processeurs que nous avons eu à

disposition. Chacun des processeurs traite simultanément toutes les fenêtres ayant une même taille. Le processus d'exploration de toutes les zones de recherches sur les images de la base GENKI a demandé 23 heures.

### 3.3 Dédution du masque d'expression

Pour chaque taille de fenêtre de recherche, nous retenons les origines des zones ayant le plus haut et plus bas niveau de précision ainsi que les taux eux-mêmes. Une fois l'ensemble de l'espace de recherche totalement exploré, nous agrégeons les informations sur l'ensemble des pixels du visage. Soit  $A_{(Rk)m}$  et  $A_{(Rk)p}$  les moyennes obtenus sur les meilleures et pires zones respectivement suivant des distributions normales avec moyennes  $\mu_m$ ,  $\mu_p$  et variances  $V(\mu_m)$ ,  $V(\mu_p)$ . Les bons résultats (par rapport à  $\mu_p - V(\mu_p)$ ) obtenus en zones de basse précision sont écartés, au même titre que les moins bons résultats (par rapport à  $\mu_m + V(\mu_m)$ ) dans des fenêtres de haute précision. Le masque C est créé en suivant la formule ci-dessous :

$$C = (A_{(Rk)m} > \tilde{A}_{(Rk)m} + V(\mu_m)) \wedge (1 - (A_{(Rk)p} < \tilde{A}_{(Rk)p} - V(\mu_p)))$$

La Figure 5 présente le masque obtenu en appliquant l'expression ci-dessus sur la base d'images GENKI. La granularité observée est due aux zones de recherches sautées dans l'étape précédente (3 pixels à la fois). La partie (A) de l'image représente les images moyennes cumulant respectivement les pires et les meilleurs taux de précisions. Par rapport à l'expression ci-dessus, en pratique, nous réalisons une opération d'étalement du contraste (B) afin de rendre plus prononcées les parties intéressantes des images avant d'appliquer la formule (C).

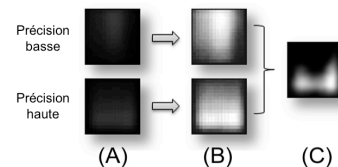


Figure 5 – Images moyennes représentant les taux (bas et haut) de classification et le masque d'expression.

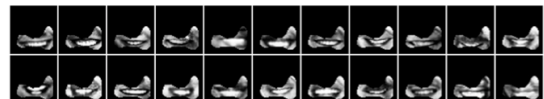


Figure 6 – Exemple d'application du masque d'expression

Le masque ainsi obtenu peut être utilisé pour extraire des images normalisées du visage un descripteur contenant uniquement les intensités des pixels sous l'emprise du masque (comme illustré dans la Figure 6).

La Figure 7 illustre différents masques générés utilisant d'autres méthodes de sélections de caractéristiques. De gauche à droite, nous présentons des masques obtenus sur la base d'images GENKI en utilisant : a) sélection des ensembles à base de corrélations [14]; b) algorithme de filtrage de type Las Vegas [15]; c) génération aléatoire d'ensembles (200 attributs); d) génération aléatoire

d'ensembles (100 attributs) ; e) sélection d'attributs avec le classifieur OneR [16] ; f) notre méthode.

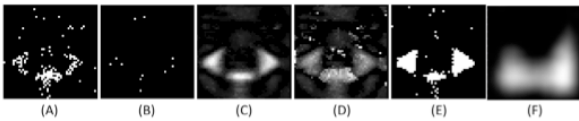


Figure 7 – Masques d'expressions créés sur GENKI

Dans la section suivante, nous présentons des expériences réalisées sur trois bases d'images différentes afin de mesurer le gain de précision avec notre masque.

## 4 Expérimentation

Nous commençons par discuter les résultats obtenus suite à la construction des 14,490 réseaux de neurones couvrant l'espace de recherche sur la base GENKI. Nous avons utilisé 50% des images GENKI pour l'apprentissage et la construction du masque. Ensuite, nous utilisons le masque construit suite à l'accumulation des zones de basse et haute précision sur les 50% de la base GENKI n'ayant pas participé à l'apprentissage ainsi que sur l'intégralité des bases JAFFE [17] et FERET [18].

Dans les premiers tests, la précision moyenne obtenue sur l'ensemble d'apprentissage de la base GENKI est de 93.6% avec une déviation standard de 14.1%. La précision moyenne pour l'ensemble de test est de 71.7% avec une déviation de 8.71%. Sur l'ensemble de réseaux 46.7% (6,773/14,490) génèrent des taux d'erreur inférieurs à 1%.

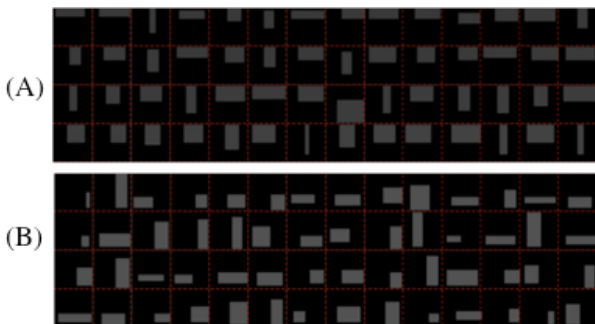


Figure 8 – Zone de recherches de précision basse (A) et hautes (B) sur les visages normalisés.

La Figure 8 (A) and (B) montrent différentes zones de recherches caractérisées, respectivement, par les meilleurs et les pires taux de classification pour les images normalisées à 50x50 pixels, par taille de fenêtre de recherche. Dans la Figure 8, les zones sont ordonnées selon les taux de précision obtenus. Les valeurs minimales sont généralement observées dans la partie haute du visage et du front (A). Alors que les valeurs maximales sont plutôt observées dans la moitié inférieure du visage autour de la bouche et dans la région médiane (B). Ces observations sont pertinentes car les expressions positives ont tendance à persister plus souvent dans la partie basse

du visage, alors que les expressions négatives sont plus souvent exprimées dans la partie haute de visage [10].

Une deuxième conclusion que nous tirons de l'analyse de l'ensemble de réseaux de neurones générés est que l'agrandissement de la fenêtre de recherche conduit généralement à des meilleurs taux de classification comme illustré dans la Figure 9. Dans cette figure, nous représentons les valeurs du coefficient de corrélation entre la taille des fenêtres et le meilleur taux de précision obtenu pour chacune. Approximativement 64% de la déviation en termes de précision est expliqué par la taille des fenêtres. Quelques unes des fenêtres de petite taille donnent une meilleure précision que des plus grandes. Ceci est dû également à la forme des fenêtres (plus larges ou plus hautes) qui permet de capturer certaines cooccurrences des pixels importants pour le processus. Mais, cela montre également qu'il est possible de réduire le nombre de caractéristiques en conservant des bons taux.

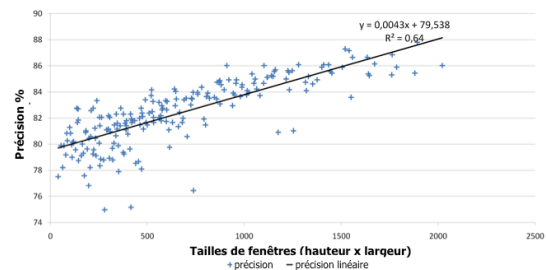


Figure 9 – Effet de la taille de fenêtres sur la précision pour les meilleurs taux obtenus sur la base GENKI.

Le masque obtenu en cumulant les zones de basse et de haute précision a été appliqué sur les trois bases d'images considérées (GENKI, JAFFE[17] et FERET[18]) afin d'évaluer son rôle dans l'amélioration des taux de classification tout en réduisant la dimensionnalité.

La base d'images JAFFE [17] (Japanese Female Facial Expression) contient 213 images couvrant sept expressions faciales : six expressions de base (joie, colère, peur, surprise, tristesse, dégoût) et une neutre. Dix femmes japonaises posent chacune trois à quatre fois pour chaque expression. La base d'images FERET [18] inclut des hommes et des femmes de différents âges et ethnies. Diverses expressions faciales sont répertoriées. La base de données contient 14,126 images et inclut 1,208 personnes. Chaque personne est photographiée sous différents angles et conditions d'illuminations.

Trois masques ont été considérés lors de cette expérience. Le premier masque couvre le visage tout entier. Le deuxième est constitué par la meilleure zone de recherche détectée par notre algorithme de fenêtres optimales. Le troisième correspond au masque construit. Nous avons mesuré les performances en terme de précisions (Figure 10) pour chaque masque et pour chaque base d'images.



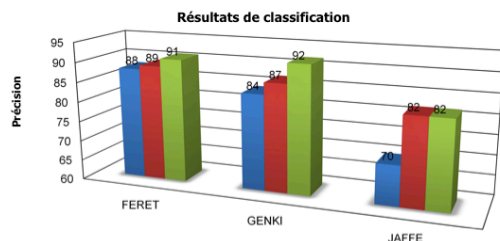


Figure 10 – Résultats expérimentaux sur différentes bases.

Les résultats montrent que les meilleurs résultats sont obtenus avec le masque construit sur la base GENKI. C'est uniquement sur la base JAFFE que la zone de recherche de précision maximale produit le même taux de classification que le masque. La plus grande amélioration induite par l'utilisation du masque est observée sur la base JAFFE où le masque construit permet de gagner environ 10 points par rapport au visage tout entier.

## 5 Conclusions

Dans le but de trouver les meilleures tailles et positions de fenêtres, nous avons mis en œuvre une recherche exhaustive sur le visage. Même si les plus grandes fenêtres donnent la plus grande précision, nos expérimentations ont montré que la taille et la position de la fenêtre ont un impact important dans le problème de classification d'expressions faciales. Dans certains cas, des petites fenêtres donnent des résultats plus précis que les grandes fenêtres. De plus, pour une taille de fenêtre donnée, il y a une grande différence de précision pour des zones très proches, ce qui met en évidence l'importance de notre étude. Dans ce travail, nous avons montré que la méthode proposée améliore les résultats de classification en sélectionnant les meilleurs POIs, sur trois bases de test.

L'approche que nous avons exposée ici est générique et peut s'adapter à la construction de masques pour d'autres types d'expressions. Pour construire des masques spécifiques à d'autres types d'expressions, il faut personnaliser l'ensemble d'images utilisé pour l'apprentissage, ainsi que la configuration du réseau de neurones pour distinguer l'expression cible.

## Références

[1] A. Hadid, M. Pietikinen, Combining appearance and motion for face and gender recognition from videos, *Pattern Recognition* 42 (2009) 2818 – 2827.

[2] P. Viola, M. J. Jones, Robust real-time face detection, *International Journal of Computer Vision* 57 (2004) 137–154.

[3] D. W. Ruck and S. K. Rogers, et al., "The multilayer perceptron as an approximation to a bayes optimal

discriminant function", *IEEE Trans. Neural Networks*, vol. 1 (1990), 296 - 298.

[4] K.-M. Lam, H. Yan, An analytic-to-holistic approach for face recognition based on a single frontal view, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20 (1998) 673 –686.

[5] M. Turk, A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* 3 (1991) 71–86.

[6] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, in: *Lecture Notes in Computer Science* 1206, 1997, 125–142.

[7] A. Hyvriinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Networks* 13 (2000) 411 – 430.

[8] G.-M. Jeong, H.-S. Ahn, S.-I. Choi, N. Kwak, C. Moon, Pattern recognition using feature feedback: Application to face recognition, *Intl. Journal of Control, Automation and Systems* 8 (2010), 141-148.

[9] V. Praseeda Lekshmi, S. Kumar, D. Vidyadharan, S. Naveen, Analysis of facial expressions using pca on half and full faces, in: *ICALIP 2008*, 1379 –1383.

[10] M.-Y. Chen, C.-C. Chen, The contribution of the upper and lower face in happy and sad facial expression classification, *Vision Research* 50 (2010) 1814 – 1823.

[11] <http://mplab.ucsd.edu>, The MPLab GENKI Database, GENKI-4K Subset, 2011.

[12] T. Danisman, I. M. Bilasco, N. Ihaddadene, C. Djeraba, Automatic facial feature detection for facial expression recognition, in: *VISAPP (2)*, pp. 407–412.

[13] H. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20 (1998) 23 –38.

[14] M. Hall, Feature subset selection: A correlation based filter approach, 1997.

[15] H. Liu, R. Setiono, A probabilistic approach to feature selection - a filter solution, *ICML*, 319-327.

[16] R. C. Holte, Very simple classification rules perform well on most commonly used datasets, *Machine Learning* 11 (1993), 63–90.

[17] S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model, in *Computer Vision ECCV 2008*, 504–513.

[18] P. Phillips, H. Moon, S. Rizvi, P. Rauss, The feret evaluation methodology for face-recognition algorithms, *Pattern Analysis and Machine Intelligence, IEEE Trans. on* 22 (2000), 1090 – 1104.

[19] M. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21 (1999) 1357 –1362.