Feature-Based Facial Expression Recognition: Experiments With a Multi-Layer Perceptron

Zhengyou Zhang

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Feature-Based Facial Expression Recognition: Experiments With a Multi-Layer Perceptron

Zhengyou Zhang

N° 3354

February 1998

THÈME 3

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Feature-Based Facial Expression Recognition: Experiments With a Multi-Layer Perceptron

Zhengyou Zhang

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Robotvis

Rapport de recherche n° 3354 — February 1998 — 22 pages

Abstract: In this paper, we report our experiments on feature-based facial expression recognition within an architecture based on a two-layer perceptron. We investigate the use of two types of features extracted from face images: the geometric positions of a set of fiducial points on a face, and a set of multi-scale and multi-orientation Gabor wavelet coefficients at these points. They can be used either independently or jointly. The recognition performance with different types of features has been compared, which shows that Gabor wavelet coefficients are much more powerful than geometric positions. Furthermore, since the first layer of the perceptron actually performs a nonlinear reduction of the dimensionality of the feature space, we have also studied the desired number of hidden units, i.e., the appropriate dimension to represent a facial expression in order to achieve a good recognition rate. It turns out that five to seven hidden units are probably enough to represent the space of feature expressions. Then, we have investigated the importance of each individual fiducial point to facial expression recognition. Sensitivity analysis reveals that points on cheeks and on forehead carry little useful information. After discarding them, not only the computational efficiency increases, but also the generalization performance slightly improves. Finally, we have studied the significance of image scales. Experiments show that facial expression recognition is mainly a low frequency process, and a spatial resolution of 64 pixels $\times$ 64 pixels is probably enough.

Key-words: Facial expression recognition, learning, Gabor wavelets, multilayer perceptron, sensitivity analysis, image scale

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Reconnaissance d'expressions faciales à base d'attributs: Expériences avec un perceptron à multicouches

Résumé : Dans ce papier, nous présentons des expériences menées sur la reconnaissance d'expressions faciales en utilisant un perceptron à deux couches. Nous examinons l'utilisation de deux types d'attributs extraits d'images faciales : les positions géométriques d'un ensemble de points fiduciels sur un visage, et un ensemble des coefficients des ondelettes de Gabor sur ces points. Ils peuvent être utilisés indépendamment ou conjointement. L'expérience comparative en performance de la reconnaissance montre que les coefficients des ondelettes de Gabor sont beaucoup plus puissants que les positions géométriques. Par ailleurs, comme la première couche du perceptron effectue en effet une réduction non linéaire de la dimension de l'espace d'attributs, nous avons aussi étudié le nombre optimal des noeuds cachés, c'est-à-dire, la dimension appropriée pour représenter une émotion faciale afin d'obtenir un bon taux de reconnaissance. Il se trouve que cinq à sept noeuds cachés sont probablement suffisants pour représenter l'espace d'expressions faciales. En suite, nous avons mené des études sur l'importance de chaque point fiduciel. L'analyse de la sensitivité montre que les points sur le front et sur la joue n'apportent pas d'information importante. Après l'élimination de ces points, non seulement le calcul est plus rapide, mais aussi la performance en généralisation s'améliore. Enfin, nous avons étudié l'impact de l'échelle image. Nos expériences montrent que la reconnaissance d'expressions est plutôt un processus de basse fréquence, et une résolution spatiale de 64 pixels $\times$ 64 pixels est probablement suffisante.

Mots-clés : Reconnaissance d'expressions faciales, apprentissage, ondelettes de Gabor, perceptron à multicouches, analyse de sensitivité, échelle image

Contents

1 Introduction

There are a number of difficulties in facial expression recognition (FER) due to the variation of facial expression across the human population and to the context-dependent variation even for the same individual. Even we human beings may make mistakes [7]. On the other hand, FER by computer is very useful in many applications such as human behavior interpretation and human-computer interface.

An automatic FER system needs to solve the following problems: detection and location of faces in a cluttered scene, facial feature extraction, and facial expression classification.

Face detection has been studied by many researchers, and it seems that most successful systems are based on neural networks [22, 20]. Once a face is detected in the image, the corresponding region is extracted, and is usually normalized to have the same size (for example, the same distance between two eyes) and the same gray level. In this paper, we do not address the face detection problem.

Facial feature extraction attempts to find the most appropriate representation of the face images for recognition. There are mainly two approaches: holistic template-matching systems and geometric feature-based systems [4]. In holistic systems, a template can be a pixel image or a feature vector obtained after processing the face image as a whole. In the latter, principal component analysis and multilayer neural networks are extensively used to obtain a low-dimensional representation. In geometric feature-based systems, major face components and/or feature points are detected in the images. The distances between feature points and the relative sizes of the major face components are computed to form a feature vector. The feature points can also form a geometric graph representation of the faces. Feature-based techniques are usually computationally more expensive than template-based techniques, but are more robust to variation in scale, size, head orientation, and location of the face in an image. The work to be described in this paper is, to some extent, an hybrid approach. We first locate a set of feature points, and then extract a set of Gabor wavelet coefficients at each point through image convolution.

Compared with face recognition, there is relatively a small amount of work on facial expression recognition. The first category of previous work uses image sequences. Suwa et al. [21] did a preliminary analysis of facial expressions by tracking the motion of twenty identified spots. Mase [15] uses the means and variances of optical flow data at evenly divided small blocks. Yacoob and Davis [25] use the inter-frame motion of edges extracted in the area of the mouth, nose, eyes, and eyebrows. Bartlett et al. [2] use the combination of optical flow and principal components obtained from image differences. Essa and Pentland [8] builds a dynamic parametric model by tracking facial motion over time, which can then be used for analyzing facial expressions. The second category of previous work tries to classify facial expressions from static images. Turk and Pentland [23] represent face images by eigenfaces through linear principal component analysis. Padgett and Cottrell [17] use an approach similar to eigenfaces but with seven pixel blocks from feature regions (both eyes and mouth). Cottrell and Metcalfe [5] use holistic representations based on principal components, extracted by feed forward networks. Rahardja et al. [18] also use holistic representations with neural networks, but the images are represented in a pyramid structure. Lanitis et al. [12] use parameterized deformable templates (flexible models) which take into account both variations in shape and grey-level appearance.

In this paper, we extract two types of features from face images in order to recognize facial expressions (Sect. 2). The first type is the geometric positions of a set of fiducial points on a face. The second type is a set of multi-scale and multi-orientation Gabor wavelet coefficients extracted from the face image at the fiducial points. They can be used either independently or jointly. The architecture we developed is based on a two-layer perceptron (Sect. 3). The recognition performance with different types of features will be compared in Sect. 4. Since the first layer of the perceptron actually performs a nonlinear reduction of the dimensionality of the feature space, we will also study the desired number of hidden units, i.e., the appropriate dimension to represent a facial expression in order to achieve a good recognition rate. The importance of each individual fiducial point to facial expression recognition is studied in Sect. 5 through sensitivity analysis. Finally, we investigate the significance of image scales for facial expression recognition in Sect. 6.

We note that a similar representation of faces has been developed in Wiskott et al. [24] for face recognition, where they use a labeled graphs, based on a Gabor wavelet transform, to represent faces, and face recognition is done through elastic graph matching.

2 Data Set and Representation

The database we use in our experiments contains 213 images of female facial expressions. The head is almost in frontal pose. Original images have been rescaled and cropped such that the eyes are roughly at the same position with a distance of 60 pixels in the final images (resolution: 256 pixels $\times$ 256 pixels). The number of images corresponding to each of the 7 categories of expression (neutral, happiness, sadness, surprise, anger, disgust and fear) is roughly the same. A few of them are shown in Fig. 1. For details on the collection of these images, the reader is referred to [14].

Each image is represented in two ways. The first uses 34 fiducial points as shown in Fig. 2. They have been selected manually. Development of a technique for automatically extracting these points is under way. (An automatic technique for building a similar representation has already been reported in the literature [11, 24].) The image coordinates of these points (geometric positions) will be used as features in our study. Therefore, each image is represented by a vector of 68 elements.

The second way is to use features extracted with 2-D Gabor transforms [6, 13]. A 2-D Gabor function is a plane wave with wavevector $\mathbf{k}$, restricted by a Gaussian envelope function with relative width σ:

$$\Psi(\mathbf{k},\mathbf{x}) = \frac{\mathbf{k}^2}{\sigma^2}\exp\left(-\frac{\mathbf{k}^2\mathbf{x}^2}{2\sigma^2}\right)\left[\exp(i\mathbf{k}\cdot\mathbf{x}) - \exp(-\frac{\sigma^2}{2})\right]$$

We set $\sigma = \pi$ for our 256×256 images. We use a discrete set of Gabor kernels which comprise 3 spatial frequencies, i.e., scales, (with wavenumber $k = \|\mathbf{k}\| = (\pi/4, \pi/8, \pi/16)$ in inverse pixels) and 6 distinct orientations from 0° to 180°, differing in 30° steps. Two examples with three of the total 18 even Gabor kernels are shown in Fig. 3. Each image is convolved with both even and odd Gabor kernels at the location of the fiducial points as shown in Fig. 2. We have therefore 18 complex Gabor wavelet coefficients at each fiducial point. In our study, only the magnitudes are used, because they vary slowly with the position while the phases are very sensitive. In summary, with Gabor wavelet coefficients, each image is represented by a vector of 612 (18×34) elements.

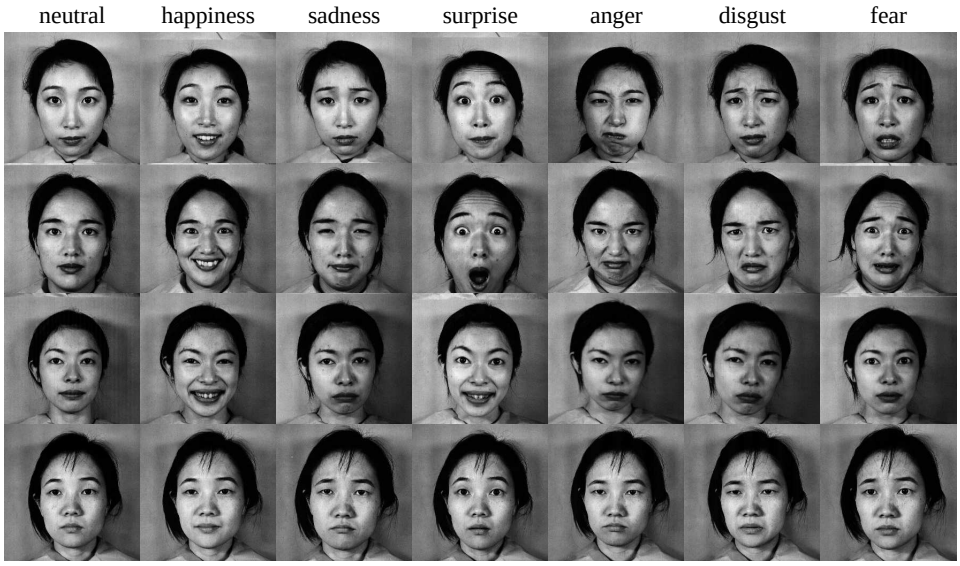

Figure 1: Facial expression database: Examples

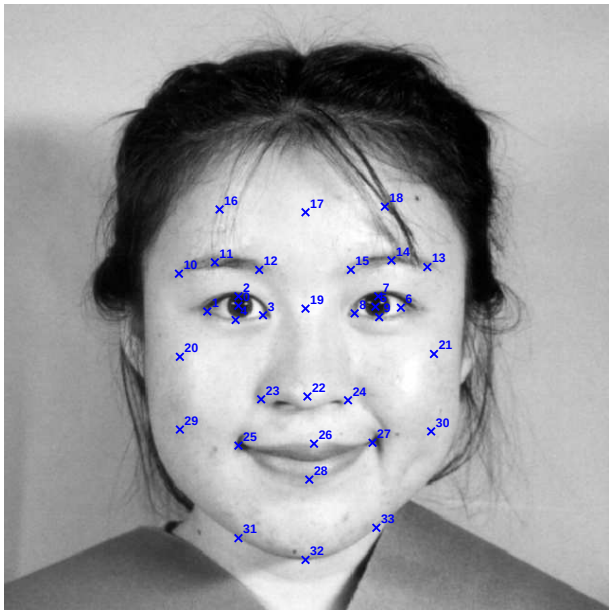

Figure 2: Geometric representation: 34 fiducial points to represent the facial geometry

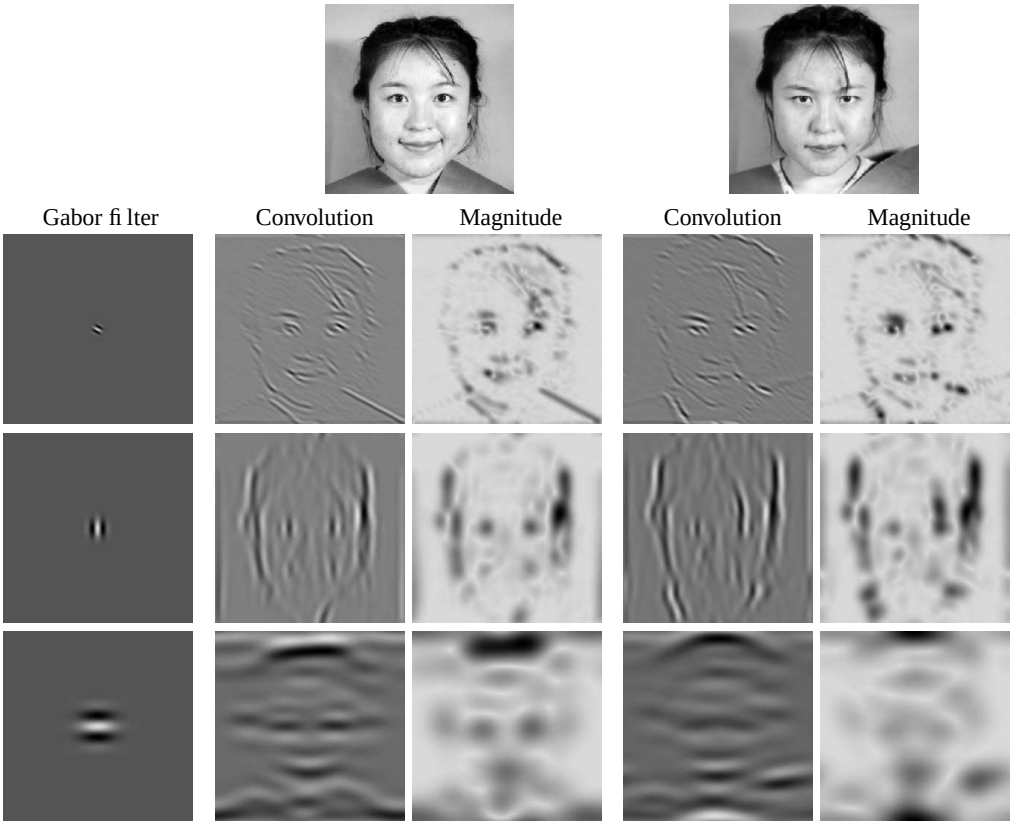

Figure 3: Gabor wavelet representation: Examples of three kernels