Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 1104–1111
Pattern Recognition Letters
www.elsevier.com/locate/patrec

A similarity-based neural network for facial expression analysis

Kenji Suzuki [a,*], Hiroshi Yamada [b], Shuji Hashimoto [c]

[a] *Department of Intelligent Interaction Technologies, University of Tsukuba, Japan*
[b] *Department of Psychology, Nihon University, Japan*
[c] *Department of Applied Physics, Waseda University, Japan*

Received 2 July 2004; received in revised form 24 July 2006
Available online 27 January 2007

Communicated by R. Manmatha

Abstract

In this paper, we introduce a novel model for the measuring of human subjective evaluation by using *Relevance Learning* based on a similarity-based multilayer perceptron. This work aims to achieve a multidimensional perceptual scaling that associates the physical features of a face with its semantic vector in a low-dimensional space. Unlike the conventional multilayer perceptron that learns from a set of an input feature vector and the desired output, the proposed network can obtain a nonlinear mapping between the input feature vectors and the outputs from a pair of objects and their desired *relevance* (distance). We conducted a facial expression analysis with both a psychological model of line-drawing image of facial expression and a real image set. Regarding the construction of semantic space, the proposed approach not only shows a good performance as compared with the conventional statistical method but is also able to project new data that are not used during the training phase. We will show some experimental results and discuss the obtained mapping function.

Keywords: Similarity-based neural network; Multidimensional perceptual scaling; Facial expressions; Stable dynamic parameter adaptation; Nonlinear mapping

1. Introduction

Judging and categorizing emotional facial expressions have long been subjects of research, mainly in experimental psychology. In classical works of psychology, Schlosberg proposed both two-dimensional (Schlosberg, 1952) and three-dimensional models (Schlosberg, 1954) of facial expression analysis; this analysis states that facial expression is located along three scales: pleasant–unpleasant, attention–rejection, and sleep–tension. In addition, some other psychological works on two-dimensional models for facial expression have been reported in recent years, for example (Russell and Fernández-Dols, 1997). In these models, the facial expressions are arranged on a low-dimensional semantic space by using statistical methods.

These models, however, do not include the relationship between the physical features and the semantic parameters although later models do include semantic scaling. We consider the projection of high-dimensional physical features perceived by humans onto another low-dimensional perceptual space in a nonlinear fashion as the more significant perspective of multidimensional perceptual scaling.

Facial expression has several physical features such as the shape and location of each facial part, the skin color, and the wrinkles. Many research studies concerning the relationship between a facial expression and its impression, particularly in the field of psychology, cognitive science, and engineering have been reported thus far (Russell and Bullock, 1986; Ekman, 1992). In these cases, a neural network is appropriate for the acquisition of nonlinear mapping between qualitatively different data. The nonlinearity in the perceptional processing of humans can be effectively

* Corresponding author. Tel.: +81 29 853 5761; fax: +81 29 853 6471.
E-mail address: kenji@ieee.org (K. Suzuki).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2007.01.005

performed using a multilayer perceptron classifier (Zhang et al., 1998; Padgett and Cottrell, 1997), as compared with the conventional statistical approaches which are based on linear functions (Katsikitis, 1997). More recently, some neural network-based approaches to facial analysis have achieved positive results (Du and Lin, 2003; Hu et al., 2002).

In addition to metric variables such as measurable physical features of objects, a measure of the similarity/dissimilarity of objects often plays an important role in pattern recognition. There are some methods that accept a measure of the similarity/dissimilarity of objects. For instance, multidimensional scaling (MDS) (Kruskal and Wish, 1978) is a typical statistical method that uses a matrix containing the similarities between objects. MDS aims to provide an arrangement of a set of objects in a geometric configuration from the pattern of proximities (i.e., similarities or distances) between every two objects. MDS is also used for facial expression analysis (Russell and Bullock, 1985).

We have already proposed a method for analyzing the human perception by using a unique neural network to determine the nonlinear mapping between the physical feature and its impression with regard to the timbre of musical instruments (Suzuki and Hashimoto, 1997) and facial expressions (Suzuki et al., 2001). In addition to the learning from proximities such as MDS, the proposed network enables the construction of a nonlinear mapping that includes the relationship between input and output patterns. By obtaining a measure of dissimilarity between objects, the system produces the outputs in a low-dimensional space in order to satisfy the given dissimilarity measure.

In order to achieve MDS by a nonlinear mapping, an algorithm for data structure analysis has been proposed by Sammon (1969). A similar work, known as Hayashi's quantization method, was reported by Hayashi (1952) In addition, certain learning methods utilizing similarity-based distance have been reported, for instance, for image database organization (Squire, 1998), the classification method (Duch, 1997).

The proposed method differs from the other existing approaches with respect to the training of the network. In particular, the network is trained by a pair of inputs and mutual distance between them is regarded as the relationship between two given patterns. The human impressions of faces can thus be visualized by the physical facial parameters that are projected on the constructed semantic space of arbitrary dimensions. Moreover, after training the network, the new data, which are not used in the training of the network, can also be evaluated by the generalization ability of the network. In this framework, the proposed network is basically similar to Sammon's nonlinear mapping and Hayashi's quantization method mentioned above. In particular, the proposed method realizes the MDS algorithm by a modified multilayer perceptron. In the visualization of the inherent data structure, it must be important to obtain the data arrangement in possible approximately low dimensions by using the relationship between objects, even by taking less notice of a strictly mathematical definition.

In this paper, we first describe our approach to analyze the facial expressions of emotion, including the description of facial expressions and the representation of human impressions. Next, the learning algorithm of the proposed similarity-based learning network is introduced along with some experimental results. The discussions and conclusions are then finally given.

2. Approach to the facial expression analysis

The following are the characteristics of the proposed multidimensional perceptual scaling of facial expressions:

(1) A nonlinear mapping represents the relationship between the physical features of a face and its impression.
(2) The mapping rule is determined by the similarity evaluation that is regarded as the desired distance between any two given face images.

It is appropriate to say that the human impression of a facial expression is due to the perceived physical feature of the face. There are many methods to arrange the given facial images by the dimension reduction technique of particular statistical methods such as factor analysis, PCA, and MDS. Generally, the subjects face difficulty while arranging them in the low-dimensional space even if the meanings of scale are given in advance. We therefore focus on the Scheffe's paired comparison method in order to reduce users' questionnaire task. Our approach aims to obtain an arrangement by a nonlinear mapping from the facial physical feature space to a low-dimensional semantic space utilizing a similarity-based neural network.

With regard to the description of facial expression, FACS (Facial Action Coding Unit) (Ekman and Friesen, 1978) is widely used by many researchers. We, however, adopted a psychological model of line-drawing facial image (Yamada, 1993a) in the present work. This is an abstract facial model that comprises nine physical facial parameters, which correspond to facial movements such as raise/lower inner/outer eyebrows and eyes and upper/lower lips as illustrated in Fig. 1. Each face has line symmetry with respect to the vertical central line on which the nose is located and fixed. Each feature point is consistently connected to the others, introducing the spline interpolation. The line-drawing model focuses not on the multiple shapes of the face but on the essential expressions. The advantage is the ease of quantitative operation. Unlike FACS that contains redundancy due to the facial muscle system, this model can directly operate with the geometrical features of facial expression.

Using this line-drawing model, six faces of typical emotional categories such as *happiness*, *anger*, *disgust*, *fear*, *surprise*, and *sadness* are acquired on the average of images drawn by 36 subjects in the previous work (Yamada,