International Journal of Electronics and Communication Engineering.
ISSN 0974-2166 Volume 6, Number 3 (2013), pp. 251-258

http://www.irphouse.com

An Efficient Method for Facial Expression Recognition : A SMOM ESTM Model : A Review

Miss. Shweta Wakchaure and Mr. Vilas Ubale

ME (Electronics), Amrutvahini College of engineering
E mail: shweta.wck@gmail.com, vilasubale1978@gmail.com

Abstract:

In this paper, an efficient method for human facial expression recognition is presented. first we have proposed a representation model for facial expressions, namely the spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of training facial images and has a powerful representation capability. Then the elastic shape–texture matching (ESTM) algorithm is used to measure the similarity between images based on the shape and texture information. By combining SMOM and ESTM, the algorithm, namely SMOM–ESTM, can achieve a higher recognition performance level. The recognition rates of the SMOM–ESTM algorithm based on the AR database and the Yale database are 94.5% and 94.7%, respectively.[1]

Introduction:

Humans interact with each other far more naturally than they do with machines. This is why face-to-face interaction cannot be still substituted by human-computer interaction in spite of the theoretical feasibility of such a substitution in numerous professional areas including education and certain medical branches. In fact, existing man-machine interfaces are perceived by a broad user audience as the bottleneck in the effective utilization of the available information flow. Hence, to improve man-machine interaction one should emulate the way in which humans communicate with each other.[4]

Human facial expression contains extremely abundant information of human's behavior and can further reflect human's corresponding mental state. [5]

As human face plays a crucial role in interpersonal communication, facial expression analysis is active in the fields of affective computing and intelligent interaction. Influenced by race, culture, personality etc, facial expression is extremely complex, much research on which has been limited to some prototype facial

expressions. However, we expect machine to recognize as many emotions as possible through facial expressions. To achieve the goal, abundant and effective data of facial expressions is necessary, and methodology of multiple facial expression recognition is to be studied.[5]

In this paper, a novel and accurate method is proposed for facial expression recognition. Our method includes two major techniques: the spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of the training set and can be used to describe the different facial expressions; and elastic shape–texture matching (ESTM), which is used to compute the similarity between two images. The combination of these two techniques, namely the SMOM–ESTM method, is used to classify the facial expressions. SMOM considers the spatial distribution of intensities in training images, and has a powerful representation capability to describe the expressions. However, SMOM does not concern the spatial correlation between neighboring pixels within an image. Therefore, ESTM is also adopted, which measures the similarity between images based on both the shape and the texture information. To measure the similarity, the positions of the two eyes and middle of the mouth are used for normalization and alignment. The shape and texture information about a face image are complementary to each other, and both are useful for expression recognition. The LEM, which mainly represents the shape information about a face, is used to describe an expression. Lyons et al. adopted the 2D Gabor wavelet to describe the texture, but the feature points, which represent the shape information, have to be detected manually.[1]

In our algorithm, ESTM is combined with SMOM for facial expression recognition. Compared with those methods based on the global features of a human face, e.g. PCA, our method considers the local information in an image, which can describe facial expressions more exactly. Furthermore, the proposed approach can be considered as a combination of template matching and geometrical feature matching, which not only possesses the advantages of feature-based approaches—such as low memory requirement—but also has the advantage of a high recognition performance in template matching.[1]

Recent Trends And Developments In The Field

Facial expressions are the means to convey emotions, feelings, warning signs of dangers, happiness, disappointments, confidence etc. of man. It is injected into the living things from the womb to tomb. Psychologists, Saints and Men of spirituality consider facial expressions as indications of hidden truth and exposition of sudden feelings, in the right way, at the right time without any reservations. In man, facial expressions were well studied, since 1971 by the pioneers Ekman and Friesen. Even in the theory of evolution of Darwin, there are reminiscence of the rule of automatic facial expression, to grab new shapes and intelligence in the transformation process of one animal into another. Ekman and Friesen are acclaimed of their contributions to the postulation of six primary emotions - happiness, sadness, fear, disgust, surprise and anger. These six distinctive facial expressions are unique in their feature. [7]

In the areas of research, lots of controversy still exist that Facial Expression Recognition is distinct to Human Emotion Recognition (B Fasel and J Luettin). However, it appears that facial expression is a mirror image that is being reflected on the face, which gives scope for facial expression recognition to expose the hidden truth and feelings of human mind. [7]

Origin and Scope Of Facial Expression Analysis

Emotions often come out as gestures, postures and even body languages in human beings. It may attain different forms with or without voice modulation to convey different needs, feelings, and anticipation. Initially, automatic facial expression was of great concern to psychologists but later it gained momentum due to its application for face detection, face tracking, face recognition, image understanding, facial nerve grading in medicine etc. Now, around the globe researches are being conducted on different areas like facial image compression, synthetic animation, video-indexing, robotics and virtual reality in addition to psychological studies. [7]

Various studies put forward many hypotheses, of which the most important one is that facial expression is a composite effect of mental state and physiological activities that attained exposition through verbal and non-verbal communications. Though mental state of the individual is of prime importance, it will be influenced by felt emotions, communication and cogitation. Similarly, physiological activities will be determined by manipulators, pain and tiredness. As a result of these composite influences and complexity, optimum accuracy still remains intricate. In fact, variety of facial expressions cannot be subjected to proper analysis and interpretation with the same type of facial expression measurements.[7]

Objectives:

Here, a novel and accurate method is proposed for facial expression recognition. Our method includes two major techniques: the spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of the training set and can be used to describe the different facial expressions; and elastic shapetexture matching (ESTM), which is used to compute the similarity between two images. The combination of these two techniques, namely the SMOMESTM method, is used to classify the facial expressions. SMOM considers the spatial distribution of intensities in training images, and has a powerful representation capability to describe the expressions. However, SMOM does not concern the spatial correlation between neighboring pixels within an image. Therefore, ESTM is also adopted, which measures the similarity between images based on both the shape and the texture information. To measure the similarity, the positions of the two eyes and middle of the mouth are used for normalization and alignment. The shape and texture information about a face image are complementary to each other, and both are useful for expression recognition. The LEM, which mainly represents the shape information about a face, is used to describe an expression. Lyons et al. adopted the 2D Gabor wavelet to describe the texture, but the feature points, which represent the shape

information, have to be detected manually.[1]

In our algorithm, ESTM is combined with SMOM for facial expression recognition. Compared with those methods based on the global features of a human face, e.g. PCA, our method considers the local information in an image, which can describe facial expressions more exactly. Furthermore, the proposed approach can be considered as a combination of template matching and geometrical feature matching, which not only possesses the advantages of feature-based approachessuch as low memory requirementbut also has the advantage of a high recognition performance in template matching.[1]

Mainly the design will be consisting of 2 parts.

1. Construction of SMOM
2. Construction of ESTM

Spatially Maximum Occurance Model

Human facial expression is a complex patternit relies on the emotion of the expressor and varies from person to person. On the one hand, the expression is determined by movements or changes in facial features, which means that

it is person-dependent and is affected by the characteristics of the expressor, such as the shapes or positions of the facial features, motion habits, and so on. On the other hand, for the same person, there are also variations

in the same expression due to different degrees of emotion. Therefore, the within-class variation of an expression is relatively large, and the betweenclass variation of different expressions is relatively small.[1]

In fact, even human beings sometimes cannot judge expressions correctly in a still image. In this case, knowing how to build proper expression models is very important. Using the mean image of a training set to represent a particular expression is simple, but most of the information is lost, and the within-class variations cannot be reflected. In this section, we will propose a new expression repre- sentation scheme, namely the spatially maximum occurrence model (SMOM), which is based on the statistical properties of the training set and contains most of the significant visual content.[1]

SMOM is constructed based on the probability of the occurrence of pixel values at each pixel position for all the training images, which is illustrated in Fig. Suppose that the number of training images is equal to N, and the size of an image is$M * H$. Therefore, there are N possible values at each pixel position (x, y). Ranking these N intensity values, we can obtain the histogram H(b) for the pixel position (x, y) as follows:

B is the number of bins in the histogram, and f(x, y) is the intensity value of the kth image at position (x, y). In general, B is equal to the number of intensity levels in the images. However, when the number of training images is small, the number of bins should be reduced and the histogram should be

$$H_{x,y}(b) = \sum_{k=1}^{N} \delta(f_k(x,y) - b),$$

where

$$\delta(m) = \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{if } m \neq 0, \end{cases} \quad \text{for } 0 \leqslant b < B. \qquad (1)$$

smoothed using a Gaussian filter as follows:

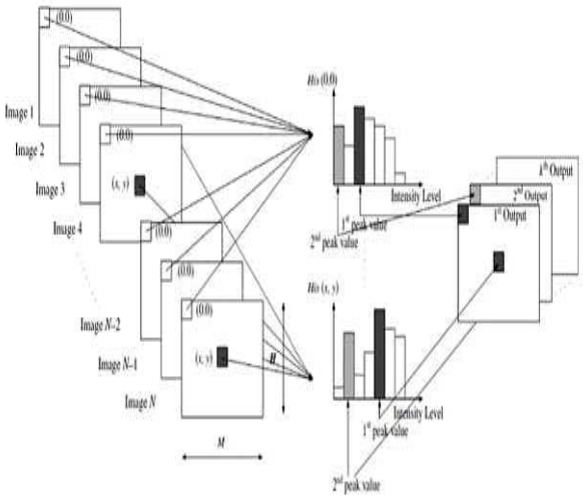
$$H'_{x,y}(b) = H_{x,y}(b) * G(\sigma, b), \qquad (2)$$

Where G(, b) is a Gaussian filter with variance , * is the convolution operator, and Hx, y(b) is the smoothed histogram of the pixel position (x, y). For each smoothed histogram, its peak values are identified and ranked in descending order. A peak occurs at a bin if its value is higher than its two adjacent bins. If a bin is the first (or the last) bin in a histogram, and its value is larger than the right (or the left) bin, we also consider it a peak. If m consecutive bins have the same value and this value is higher than the two adjacent bins of the consecutive bins, a peak also exists, and the bin value of the peak is set at the middle of the m consecutive bins. The gray levels corresponding to those bins that are the peaks of a histogram will be used in constructing SMOM. In other words, at each pixel position (x, y), the gray levels corresponding to the peaks are ranked according to their probabilities of occurrence. SMOM is therefore defined as follows:

$$SMOM(x,y,k) = \{b_1, b_2, \ldots, b_k\}, \quad \text{where } 0 \leqslant b_k < B, \quad \text{for } 0 \leqslant x < M \text{ and } 0 \leqslant y < H, \qquad (3)$$

where k is the number of peaks to be considered in the representation, b1, b2, ..., bk are the gray levels corresponding to the peaks of the histogram for pixel position (x, y), and the conditions Hx, y(b1)Hx, y(b2)Hx, y(bk)are satisfied. Usually, k is a small value. If the number of peaks p in a histogram is less than k, the remaining k − p values will correspond to those bins with the largest probabilities of occurrence.[1]

In our algorithm, the gray levels of those bins corresponding to the highest peaks, rather than the highest values, are used to represent the pixel intensities. As a histogram can be considered as a multi-cluster distribution and a peak is the representation of a bin cluster, so the peak values can provide useful statistical information at a pixel position, and are suitable for modeling complex patterns. An advantage of SMOM is its powerful representation capability. If each pixel position (x, y) in SMOM is represented by k values, the number of possible images that can be generated from a SMOM is kMH, where the image size is MH. Suppose that k = 2, M and H are both equal to 64, a SMOM can be used to represent 26464 different images. Furthermore, because the representation values are based on the statistical properties of the training images, most of the significant visual content of the training set is maintained in SMOM.[1]



SMOM considers intensity distribution at each pixel position for an expression class; however, the spatial correlations between neighboring pixels within an image are ignored, and they are important for the description and discrimination of the facial expressions. Therefore, ESTM is adopted complement SMOM.[1]

Elastic shapetexture matching ESTM is a method that measures the similarity between images based on their shape and texture information. The shape is represented by the edge map E(x, y), and the texture is characterized by the Gabor wavelets and the gradient direction of each pixel, which are described by the Gabor map G(x, y) and the angle map A(x, y), respectively.[1]

In this paper, the output of an image after edge detection is called an edge image, while after a thresholding procedure, the binary image produced is called an edge map of the image. The edge image is obtained by morphological operations ; and an adaptive thresholding scheme is adopted to produce the edge map E(x, y). The Gabor map of an image is obtained by concatenating the magnitudes of Gabor wavelet representations at different center frequencies and orientations. In our method, the center frequency is chosen to be_/2, and the orientation varies from 0 to 7_/8 in steps of _/8. Nastar et al. found that, when a facial expression varied, only the high-frequency spectrum was affectedthis is called a high-frequency phenomenon. This observation suggests that the high-frequency components are more discriminant for facial expressions.[1]

Therefore, in our method, we apply the Gabor wavelets on the edge images, instead of the original images, to obtain the corresponding texture information in the high-frequency spectrum. The angle map A(x, y) consists of the gradient direction of each edge point. For the edge map E(x, y), Gabor map G(x, y), and angle map A(x, y), our shapetexture Hausdorff distance is defined as follows:

Given two human face images A and B, then two finite point sets AP = a1, ..., aNA and BP = a1, ..., aNB can be obtained, where the elements in AP and B P correspond to the points in the edge maps EA and EB of the original images, and N A

and N Bare the corresponding numbers of points in sets A P and BP, respectively. Then, the shapetexture Hausdorff distance is

$$H(A,B) = \max(h_{St}(A,B), h_{St}(B,A)). \quad (4)$$

hst(A, B) is called the directed shapetexture Hausdorff distance, and is defined as follows

$$h_{St}(A,B) = \frac{1}{N_A} \sum_{a \in A_p} \max \left(I \min_{b \in N^a_{B_p}} d(a,b), (1-I)P(a) \right), \quad (5)$$

where NaBP is the neighborhood of the point a in the setBp, P(a) is an associated penalty function, and I is an indicator which is equal to 1 if there exists a point b ∈ NaBP , and which is equal to 0 otherwise. d(a, b) is a distance measure between the point pair (a, b), which consists of three different terms as follows

$$d(a,b) = \alpha \cdot d_e(a,b) + \beta \cdot d_g(a,b) + \gamma \cdot d_a(a,b), \quad (6)$$

where de(a, b),dg(a, b) and da(a, b) are the edge distance, Gabor distance and angle distance, respectively, for the pixel a ∈ A to a pixel b within the neighborhood of a in BP,and α,β and γ, are the coefficients used to adjust the weights of these three distance measures. All three measures are independent of each other and are defined as follows :

$$d_e(a,b) = \|a - b\|, \quad (7)$$

$$d_g(a,b) = \|\tilde{G}_A(a) - \tilde{G}_B(b)\|, \quad (8)$$

and

$$d_a(a,b) = \|A_A(a) - A_B(b)\|, \quad (9)$$

where k . k is an underlying norm, GA,GB, AA and AB are the Gabor maps and angle maps of the two images, respectively. Similarly, the penalty P(a) in Eq.(5) can also be considered as a combination of three parts, i.e

$$P(a) = \alpha \cdot P_e(a) + \beta \cdot P_g(a) + \gamma \cdot P_a(a), \quad (10)$$

where Pe, Pg,and Pa are the corresponding penalties for these three distance measures, and _,_ ,and have the same values as in Eq. (6). An advantage of using Eq. (10) is that it allows us to adopt different penalties for different distance measures. In our method, due to the act that the representation using Gabor wavelets magnitudes is less sensitive to lighting conditions, we define

$$P_g(a) = \|\tilde{G}_A(a) - \tilde{G}_B(a)\|. \quad (11)$$

The values of Pe(a) and Pa(a) are simply set as fixed values to compute the penalty P(a).[1]

Conclusion:

Thus we have seen the necessaty of facial expression recognition. There are various methods of recognizing facial expressions. Out of them we have seen here the method using shape and texture.

We have used here a combination of two models i.e. SMOM(Spatial maximum occurance model) and ESTM(Elastic shape and texture matching). The method gives an efficiency about 95 %.

References:

[1] Facial expression recognition based on shape and texture Presented by: Xudong Xie, Kin-Man Lam
[2] An automatic facial expression recognition approach based on confusion-crossed support vector machine tree Presented by-Qinzhen XuI, Pinzheng Zhang,Wenjiang Pei, Luxi Yang, Zhenya He
[3] Real time 2D + 3D facial action and expression recognition(2010) Presented by-Filareti Tsalakanidou SotirisMalassiotis
[4] Facial action unit recognition using temporal templates Presented by -michel valstar, ioannis patras and maja pantic
[5] Beihang university facial expression database and multiple facial expression recognition Presesented by-yu-li xue, xia mao, fan zhang
[6] Book: Digital image processing By- Gonzallez and Woods.
[7] Study of the Changing Trends in Facial Expression Recognition Presented by: Dr. S. Ravi Mahima S
[8] Elastic shape-texture matching for human face recognition Presented by: Xudong Xie Kin-Man Lam
[9] Automatic Target Recognition by Matching Oriented Edge Pixels Presented by: Clark F. Olson and Daniel P. Huttenlocher