STUDIA INFORMATICA 2012
Volume 33 Number 2B (106)

Ewa PIĄTKOWSKA, Jerzy MARTYNA
Uniwersytet Jagielloński, Instytut Informatyki

HUMAN EMOTION RECOGNITION BASED ON FACIAL TEXTURE INFORMATION

Summary. The paper presents the system for automatic emotion recognition. Firstly, face detection algorithm [5] is performed on input image to create face representation. Then, face texture is encoded with Local Binary Patterns [11] and used as a feature set in emotion recognition. The Support Vector Machine [15] is used as a classifier. The proposed system was tested with spontaneous emotions.

Keywords: spontaneous emotions recognition, facial expressions, local binary patterns, Haar-like features, support vector machine

ROZPOZNAWANIE LUDZKICH EMOCJI NA PODSTAWIE INFORMACJI ZAKODOWANEJ W TEKSTSTURZE TWARZY

Streszczenie. W niniejszym artykule opisany został system do automatycznego rozpoznawania emocji. Pierwszym etapem systemu są detekcja i lokalizacja twarzy [5] na obrazie wejściowym. Następnie tekstura twarzy kodowana jest przy użyciu Local Binary Patterns [11] i zastosowana jako zbiór cech opisujących emocję. Maszyna wektorów wpierających [15] pełni rolę klasyfikatora w rozpoznawaniu emocji. Skuteczność przedstawionego systemu została zbadana dla zadania rozpoznawania emocji spontanicznych.

Słowa kluczowe: rozpoznawanie emocji spontanicznych, ekspresje mimiczne, local binary patterns, cechy Haar-podobne, maszyna wektorów wspierających

1. Introduction

The field of automatic emotion recognition is widely explored in the area of computer vision. The main subject of study is human face which plays crucial role in social communica-

tion. According to psychological research [10] almost 55% of information is carried by facial expressions.

What is more, the idea of automatic emotion recognition systems became eligible since Ekman et al. [3] introduced the theory, that there are six basic emotions universal for all people, despite of culture or nation. Those emotions are: joy, sadness, anger, fear, disgust and surprise.

Facial expressions recognition systems (FERS) are used in many fields starting with human computer interaction applications such as user mood driven software, through the medical support in pain detection or newborn children monitoring, ending with surveillance software, like drivers fatigue detection systems. Although emotion recognition is highly applicable, it is still very challenging problem in computer vision and much effort is put to improve the performance of FERS.

1.1. Related works

During the last two decades facial expression recognition was an active research topic in the area of computer science. The task of facial expression analysis involves three main phases: face detection, feature extraction and expression recognition. Many different approaches were proposed for each phase however in this paper we will only mention few of them to outline the basic idea of emotion recognition problem. More detailed description of work that was done can be found in [12, 18].

Face detection is the first stage of facial expression recognition system in which face is to be localized in the input image. Face can be perceived as an integral object – holistic approach or as a set of facial landmarks (eyes, mouth, nose) – analytic approach. An example of holistic face representation can be found in work by Huang [6] who introduced Point Distribution Model (PDM) based on mean geometry of human face. Moreover, Pantic and Rothkrantz [13] detected face by analysis of vertical and horizontal projections as well as skin color segmentation. Analytic face detection was used in work by Kobayashi and Hara [7] where face region was determined by iris location in monochrome images, and in Kimura and Yachida [8] who searched for eyes and mouth corners. Face could be also represented by a set of features for instance Haar-like features in Viola and Jones [16] algorithm or eigenfaces in Essa and Pentland [4] approach.

Depending on face representation different feature extraction techniques are used to describe facial expression. In Pantic and Rothkrantz [13] face is regarded as a set of points and expression is measured by displacement of those points in the initial and peak image. Littlewort et al. [9] used Gabor wavelets to encode facial texture to describe emotion by appearance fea-

tures. The Active Appearance Model, introduced by Edwards et al. [2], combines shape and texture information.

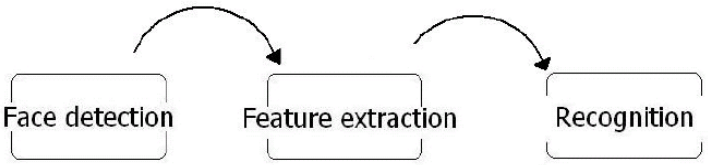

Fig. 1. System steps diagram
Rys. 1. Diagram elementów system

The final stage of the system is the classification task where expression described by a set of features is assigned to one of several classes. Emotions are usually categorized in terms of six basic emotions: anger, sadness, joy, surprise, disgust and fear, however sometimes there is seventh class included for neutral expression. Different machine learning methods can be used at this stage for instance: *k*-nearest neighbors [6], neural networks [7], expert systems [13], Support Vector Machines or boosting algorithms [9].

1.2. Outline of the proposed system

In this paper we present fully automatic system for emotion recognition. System's performance is tested in terms of spontaneous emotion recognition which is usually hard problem because spontaneous expressions are subtle and person specific. In our approach, the information about expression is retrieved from facial appearance. To encode texture we use Local Binary Patterns [11] which are efficient as well as less memory and time consuming than commonly used Gabor wavelets. The goal of this work is to learn how significant is the information carried by facial texture in task of spontaneous emotions recognition.

This paper is organized as follows: firstly we provide some theoretical background about the algorithms used in the proposed system. In third section, the flow of the system is presented with the results of our tests. Finally, we present conclusions and indicate possible improvements which can be addressed in the future work.

2. Emotion recognition problem

The proposed system for emotion recognition consists of three steps, namely: face detection, feature extraction and emotion recognition (Fig. 1). In the first step, the input image is processed in order to detect the occurrence of face and to create face model. Next step includes emotion representation with a set of well-suited features.

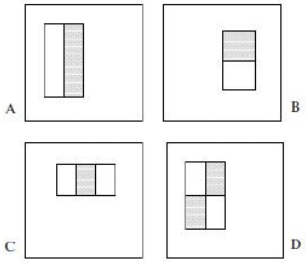

Fig. 2. Examples of Haar-like features
Rys. 2. Przykłady cech Haar-podobnych

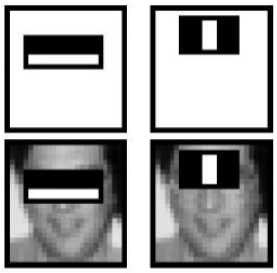

Fig. 3. Features detected on face region
Rys. 3. Cechy wykryte na twarzy

In our system, emotion is described by texture of face region. The last step of a system is concerned with classification task where detected emotion is assigned to one of the seven classes (neutral + six basic emotions). The output of the system is properly labeled image.

2.1. Face detection

Face detection algorithm used in first stage of the system is based on work by Viola and Jones [16] who proposed a method for rapid object detection. In this approach image is represented by a set of rectangular Haar-like features (see Fig. 2), which are calculated by subtracting a sum of pixels covered by white rectangle from a sum of pixels covered by gray rectangle. Depending on the type of feature we can detect different elements in the image.

Two-rectangular features detect contrast between two vertically or horizontally adjacent regions. Three-rectangular features detect contrasted region placed between two similar regions and four-rectangular features detect similar regions placed diagonally (Fig. 3).

To reduce computational effort put into feature calculation, input image is transformed into integral image in which each pixel is a sum of pixels above and to the left. Pixels in integral image are calculated by the formula:

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y'), \tag{1}$$

where $ii(x,y)$ is integral image and $i(x,y)$ is input image.

Using integral image improves the efficiency of the algorithm because the value of each rectangle (in terms of Haar-like features) requires up to four pixel references. Considering image representation, the number of features is much higher than number of pixels in the original image. However it was proven, that even small set of well-chosen features can build a strong classifier for object detection. Viola and Jones [16] used the Adaboost algorithm which iteratively selects the most discriminative feature to separate positive and negative examples in training set. The Viola and Jones [16] method is widely used in the area of face detection because of its efficiency and robustness.

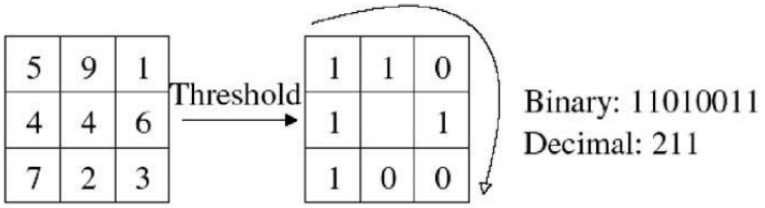

Fig. 4. LBP encoding scheme
Rys. 4. Kodowanie LBP

2.2. Facial expression representation

Second stage of the proposed system is concerned with the task of emotion description by the set of well-chosen features. In this paper, we examine the significance of texture information in emotion recognition. To encode texture we use the Local Binary Patterns (LBP) method, proposed originally by Ojala et al. [11] and later extended by Ahonen [1] and Hadid [5]. The LBP method allows for transforming the image into a representation, thanks to the application of the special operator which assigns the value on the basis of a value of the particular pixel *P*. The LBP operator is given by:

$$\forall_{n\in N} t(n) = \begin{cases} 1, & n \le P \\ 0, & n > P \end{cases} \tag{2}$$

where N is the number of pixels in the neighbourhood. Values of pixels from the neighbourhood, making a binary sequence, constitute a code which, when transformed into the decimal system, is assigned to a pixel.

The classical LBP operator has analyzed the neighbourhood with dimensions 3×3, yet the relatively small size of the operator was its basic limitation. For the purposes of the features extraction, a classical LBP operator was used here. The dimensions of the neighbourhood for the operator, e.g., a circular neighbourhood with radius *R* and any number of pixels *P* are determined. As a result of LBP transformation carried out in this way, binary standards, coded local primitives of texture, the micro-patterns or textons also called. Examples of such micro-patterns are as follows: spot, spot/flat, line end, edge, corner.

Sliding window is applied to the face region detected in the previous stage to transform it into LBP representation (see Fig. 4). Value of each pixel in the neighbourhood is thresholded with the value of the central pixel in the sliding window.

Neighborhood pixels after thresholding are formed into a binary code and the decimal value of this code is assigned to the central pixel in the corresponding LBP image. Binary codes are called 'micro-textons' because they represent texture primitives such as curved edges, flat or convex areas.

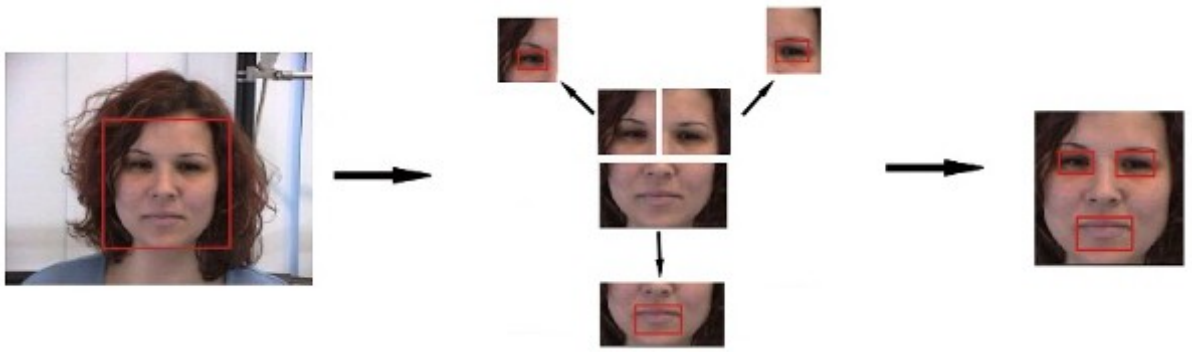

Fig. 5. Face detection and representation
Rys. 5. Lokalizacja twarzy i utworzenie jej reprezentacji

2.3. Emotion recognition

The last stage of the proposed system is emotion recognition in which the Support Vector Machine [15] with Radial Basis Function (RBF) kernel is used as a classifier. The Support Vector Machine is a machine learning system which receives labeled training data and transforms it into higher dimensional feature space. Then separating hyperplane with respect to margin maximization is computed to determine the best separation between classes. The greatest advantage of SVM is that even with small set of training data it has good performance in generalization.

3. Experimental Results

In the proposed system, the algorithm was used for face, eyes and mouth detection. After the acquisition, input image is being searched in order to detect face. In case of finding more than one face in the image, the biggest one is chosen for emotion analysis. Inversely, when there is no face in the image, further processing is omitted.

If the face is detected in the image (see Fig. 5), the classifiers for landmark detections are applied to find mouth, left and right eye. The Viola-Jones framework, used in the first stage

of our system, supports the trained classifiers in the XML files of OpenCV which can be download as part of the OpenCV software accessible on [http://opencv.willowgarage.com]. To improve algorithm efficiency, each landmark is searched in narrowed region, for instance, the left eye is searched only in upper left region of the detected face. Having locations of the face and its landmarks, we can form the face representation which is used in the next stage of the proposed system.

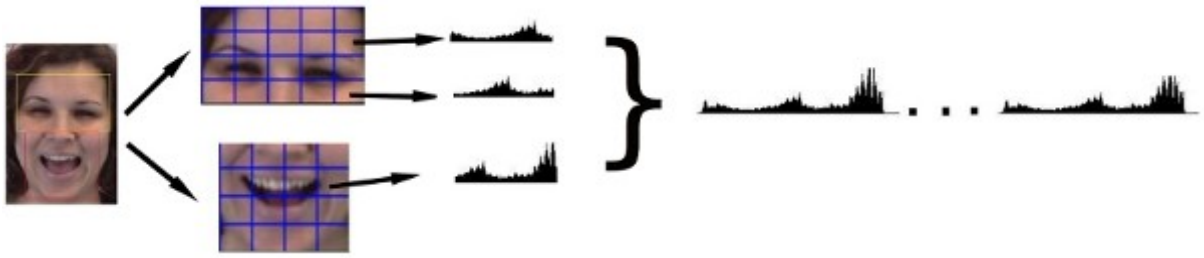

Fig. 6. Feature extraction procedure
Rys. 6. Procedura ekstrakcji cech

In the second stage of our system, we encode face texture in an analytic way to use also the spatial information about texture. In order to reduce the size of feature set, only the parts of the face which are highly involved in emotion expressions are encoded. Those parts are chin – mouth – cheeks and forehead – eye regions (see Fig. 6). Particular face regions are normalized to the same size, namely: 90x48 for upper region and 72×48 for lower face region. Next, regions are divided into grids of sizes: 4x4 in the lower part, 5×4 in the upper part. Each patch in the grid is then encoded with LBP operator. We apply the basic version of LBP which uses 3x3 sliding window thus the range of possible codes are from 0 to 255. The original image texture is described by 256-bin histogram of the corresponding LBP image. Therefore, the feature set in our system consists of 36 histograms and particular emotion is described by 9216 features.

For the purpose of training we used the FG-NET Facial Expression and Emotion Database [17] which consists of MPEG video files with spontaneous emotions recorded. Database contains examples gathered from 18 subjects (9 female and 9 male). What is more, each subject shows particular emotion three times, thus first two samples are used for training while third is used for testing.

The proposed system was trained with captured video frames in which the displayed emotion is very representative. Training set contains 675 images and testing set – 330. Both training and testing sets consist of images with seven states: neutral, surprise, fear, disgust, sadness, happiness and anger.

To make dimensionality reduction, a Principal Component Analysis (PCA) [14] was used to find the vectors that best account for the distribution of face images within the entire image

space. We recall that these vectors define the subspace of face images, and the subspace is called face space. All faces in the training set are projected onto face space to find a set of weights that describes the contribution of each vector in the face space.

The key procedure in PCA is based on the Karhumen-Loeve transformation. If the image elements are considered to be random variables, the image may be seen as a sample of a stochastic process. The PCA basis vectors are defined as the eigenvectors of the scatter matrix S_T, namely:

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T, \tag{3}$$

where μ represents the mean vector of the training images, x_i is a face vector of dimension n and N is the number of different face images in the training set.

Since the proposed method in the third stage of our system is based on the SVM classifier, the necessary SVMs are trained with the preprocessed data to obtain the Support Vectors associated to each class. It is well-known that SVM generalization performance (estimation accuracy) depends on a good setting of meta-parameters C and ε. Parameter C determines the trade off between the model complexity and the degree to which deviations larger than ε are tolerated in optimization. Parameter ε controls the width of insensitive zone, used to fit the training data [15]. We assumed RBF kernel function $K(x_i, x) = \exp(-\| x - x_i \|^2 / (2\rho^2))$, so that $K(x_i, x) \le 1$. Hence we obtain the following upper bound on SVM function:

$$| f(x) | \le C \cdot n_{SV}, \tag{4}$$

where n_{SV} is the number of Support Vectors (SVs). In order to estimate the value of C independently of the number of n_{SV}, we assumed that $C \ge | f(x) |$ for all training samples.

Based on well-known results, the value of ε should be proportional to the input noise level $\varepsilon \propto \sigma$ [15]. We assumed that the standard deviation of noise σ can be estimated from data. The choice of ε should also depend on the number of training samples. This suggests the prescription for choosing $\varepsilon \propto \sigma / \sqrt{n}$. According to these expressions, we used following values: C is equal to 15, ε is equal 0.15, σ is equal to 3.9.

System's performance was measured with accuracy rate that is the proportion of properly classified images to all images in the test set. Proposed system recognizes emotions with accuracy rate of 71%. Additionally, the recognition results were presented by confusion matrix (see Table 1), which not only shows recognition accuracy of each emotion but also indicates the emotions which are commonly confused. The best recognition rate was obtained for sadness (91%), the worst one for anger (51%) that was usually confused with sadness.

Table 1
Confusion matrix for emotion classification

%	neutral	joy	sadness	surprise	anger	Fear	disgust
neutral	61	0	33	0	0	6	0
joy	3	78	3	10	0	3	3
sadness	3	1	91	0	4	1	0
surprise	3	3	3	76	0	12	3
anger	0	2	36	0	51	4	7
fear	2	2	28	4	0	64	0
disgust	0	0	20	0	7	7	64

4. Conclusion

In this paper, we proposed fully automatic system for spontaneous emotion recognition. The system consists of 3 stages: face detection, feature extraction and classification. Firstly, acquired image is processed in order to detect occurrence of face and to create its representation with use of Viola and Jones [16] algorithm. Then, face region is encoded by Local Binary Patterns [11] method and passed to the SVM [15] classifier for emotion recognition. Our main goal was to investigate the performance of Local Binary Patterns as a texture descriptor. The system can recognize emotions with accuracy rate of 71% therefore information encoded in facial texture is significant.

In the future, we intend to improve the performance of classifier by reducing the feature set to contain the most discriminative features for particular emotion description. Some other classification methods could be considered as well.

BIBLIOGRAPHY

1. Ahonen T., Hadid A., Pietikainen M.: Face Recognition with Local Binary Patterns. Computer Vision – ECCV, LNCS, Vol. 3021, 2004, p. 469÷481.
2. Edwards G. J., Cootes T. F., Taylor C. J.: Face Recognition Using Active Appearance Models. Proc. European Conf. Computer Vision, Vol. 2, 1998, p. 581÷695.
3. Ekman P., Huang T., Sejnowski T., Hager J.: Final Report To NSF of the Planning Workshop on Facial Expression Understanding, http://face-and-emotion.com/dataface-/nsfrept/ nsf_contents.html. 1992.

4. Essa I., Pentland A.: Coding, Analysis Interpretation, Recognition of Facial Expressions. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.19, No. 7, 1997, p. 757÷763.
5. Hadid A., Pietikainen M., Ahonen T.: A Discriminative Feature Space for Detecting and Recognizing Faces. Proc. Computer Vision and Pattern Recognition, 2004, p. 797÷804.
6. Huang C. L., Huang Y. M.: Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification. J. Visual Comm. and Image Representation, Vol. 8, No. 3, 1997, p. 278÷290.
7. Kobayashi H., Hara F.: Facial Interaction between Animated 3D Face Robot and Human Beings. Proc. Int'l Conf. Systems, Man, Cybernetics, Vol. 3, 1997, p. 732÷737.
8. Kimura S., Yachida M.: Facial Expression Recognition and Its Degree Estimation. Proc. Computer Vision and Pattern Recognition, 1997, p. 295÷300.
9. Littlewort G. C., Bartlett M. S., Chenu J., Fasel I., Kanda T., Ishiguro H., Movellan J. R.: Towards Social Robots: Automatic Evaluation of Human-robot Interaction by Face Detection and Expression Classification. Advances in Neural Information Processing Systems, Vol. 16, 2004, p. 1563÷1570.
10. Mehrabian A.: Communication without Words. Psychology Today, Vol. 2, No. 4, 1968, p. 53÷56.
11. Ojala T., Pietikainen M., Maenpaa T.: Multiresolution gray-scale and rotation invariant texture with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 7, No. 7, 2002, p. 971÷987.
12. Pantic M., Rothkrantz L.: Automatic Analysis of Facial Expressions: The State of the Art. IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, 2000.
13. Pantic M., Rothkrantz L.: Expert System for Automatic Analysis of Facial Expression. Image and Vision Computing Journal, Vol. 18, No. 11, 2000, p. 881÷905.
14. Solomon Ch., Breckon T.: Fundamentals of Digital Image Processing. A Practical Approach with Examples in Matlab, Hoboken, John Wiley and Sons, 2011.
15. Vapnik V.N.: Statistical Learning Theory. John Wiley & Sons, 1998.
16. Viola P., Jones M. J.: Robust Real-time Object Detection. International Journal of Computer Vision, Vol. 57, No. 2, 2004, p. 137÷154.
17. Wallhoff F.: Facial Expressions and Emotion Database. Technische Universität München, http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html, 2006.

18. Zeng Z., Pantic M., Roisman G.I., Huang T. S.: A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 1, 2009, p. 39÷58.

Wpłynęło do Redakcji 31 stycznia 2012 r.

Omówienie

W niniejszym artykule opisany został system do automatycznego rozpoznawania emocji. Rozpoznawanie emocji odbywa się w trzech krokach: lokalizacja twarzy na obrazie, ekstrakcja cech opisujących widoczną ekspresję i klasyfikacja, czyli rozpoznanie emocji. Obraz po wczytaniu do systemu jest przeszukiwany w celu wykrycia i lokalizacji regionu twarzy. Do tego celu użyty został dobrze znany algorytm [16], wykorzystywany do szybkiej detekcji obiektów na obrazie. Następnie system analizuje region twarzy pod względem tekstury, aby za pomocą kodów LBP [11], stworzyć reprezentację emocji w postaci wektora cech. W ostatnim kroku dla zadania klasyfikacji użyty został mechanizm Maszyny Wektorów Wspierających [15], który, mając dany wektor cech, przypisuje emocji jedną z 7 etykiet (twarz neutralna lub emocje: radość, smutek, strach, zaskoczenie, wstręt oraz złość).

Skuteczność systemu została zbadana dla zadania rozpoznawania emocji spontanicznych, które są dość trudne do klasyfikacji ze względu na ich subtelność i różnorodność. Badania wykazały, że system potrafi rozpoznawać spontaniczne emocje z dokładnością 71%. Dodatkowo wykazane zostało, że tekstura pełni bardzo ważną rolę w analizie ekspresji i, w połączeniu z innym metodami ekstrakcji cech, może istotnie poprawić dokładność klasyfikacji.

Addresses

Ewa PIĄTKOWSKA: Uniwersytet Jagielloński, Instytut Informatyki Stosowanej, ul. Reymonta 4, 30-059 Kraków, Polska, ewa.piatkowska@uj.edu.pl.
Jerzy MARTYNA: Uniwersytet Jagielloński, Instytut Informatyki, ul. Prof. S. Łojasiewicza 4, 30-348 Kraków, Polska, martyna@softlab.ii.uj.edu.pl.