Anthropometric Facial Emotion Recognition

Julia Jarkiewicz, Rafał Kocielnik, and Krzysztof Marasek

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa, Poland
{julia.jarkiewicz,rafal.kocielnik,kmarasek}@pjwstk.edu.pl

Abstract. The aim of this project is detection, analysis and recognition of facial features. The system operates on grayscale images. For the analysis Haar-like face detector was used along with anthropometric face model and a hybrid feature detection approach. The system localizes 17 characteristic points of analyzed face and, based on their displacements certain emotions can be automatically recognized. The system was tested on a publicly available database (Japanese Female Expression Database) JAFFE with ca. 77% accuracy for 7 basic emotions using various classifiers. Thanks to its open structure the system can cooperate well with any HCI system.

Keywords: emotion recognition, facial expression detection, affective computing.

1 Introduction

Effective computer image analysis was always a great challenge for many researchers. Tasks, usually quite simple for humans, such as object or emotion recognition proves to be very complicated in computer analysis.

Among the main problems are susceptibility to varying lightning conditions, color changes and differences in transformation. Effective detection of human faces is one of the greatest problems in image analysis. Therefore it is even more challenging to efficiently and effectively localize features of a face in analyzed image and to relate them to expression of emotions.

Hereby work is actually an attempt to create a computer system able to automatically detect, localize and recognize facial features. Sought features are, in this case, characteristic points placed in selected locations on human face model. The locations and distances between them change during facial expressions.

There are many, more or less effective solutions capable of detecting or recognizing faces, however only a few comprehensive and effective solutions exist connecting all those features together and, at the same time, able to cooperate with an emotion recognition system. The work is largely based on solutions presented in [1], [2], [3]. As our test-bed the JAFFE database [4] was used. It consists of 213 grayscale 256x256 pixels photos of 10 Japanese woman faces showing 6 basic emotions (fear, sad, angry, happy, surprised, disgust) each in 3 variants and one neutral face. Low resolution, grayscale only photos form a challenge even for advanced techniques.

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2009, LNCS 5611, pp. 188–197, 2009.

2 Face Detection

Before any attempt to detect facial features can be made, a precise position of a face must be determined. Number of approaches exists that address this task, see [3] for complete description. The one used for this work is presented below.

2.1 Cascade Haar-Like Feature Classifier

Solution uses Haar-like features based on a very simple Haar function:

$$H(t) = \begin{cases} 0 \; for \; t < 0 \\ 1 \; for \; 0 \leq t \leq 0{,}5 \\ -1 \; for \; 0.5 \leq t < \; 1 \\ 0 \; for \; t \geq \; 1 \end{cases} \tag{1}$$

The method works in different scales. For the sake of performance, the analysis begins at the highest scale and when the template is found, the region nearby is searched again in a more precise scale. This makes the detection a coarse to fine approach. The process is repeated until the most precise scale is reached. The method is one of the most popular thanks to its effectiveness and efficiency [2]. This is because of three main characteristics:

Cascade classification – The whole classification process is divided into stages that follow one another, greatly saving processing power and increasing efficiency.

Fig. 1. Haar-like image masks and their placement on analyzed image (after [7])

Image masks (weak Haar-classifiers) and boosting – each of the stages in cascade is a separate complex classifier based on boosting (Adaboost) [9] in the form of a decision tree with at least two leaves. The final determination is made based on the so called Haar-like features. The Haar-like features are in fact simple image masks that have been designed to find different features of the image like lines, corners, etc. The extended set of such masks is presented in Fig. 1.

Usage of integral image –for detection of objects in different locations a sliding widow technique is used. Finding objects in different sizes, was solved by usage of integral image which allows to resize classifier rather than the image itself [2].

The implementation of the algorithm used in this work is a part of the *OpenCV* [5] free image processing library. The results of the face localization are shown in Fig. 2.

Fig. 2. Examples of face localizations (*JAFFE database and Internet*)

2.2 Face Normalization

Currently two types of normalization are used by the system: illumination and geometrical normalizations.

Illumination normalization is in use at various stages of processing, beginning with normalization for the whole image and then followed by normalization for identified regions containing facial elements. Plain histogram equalization but also histogram matching and Retinex method [6] are used in different areas of this work.

Geometrical normalization is used in order to scale, crop or rotate a face. In this work only rotation and cropping of the face is performed since scaling could introduce some lose of the details that are important for precise feature detection.

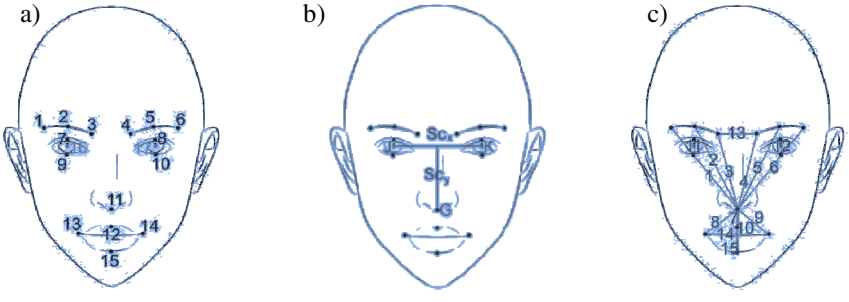

Fig. 3. Normalization of anthropometric distances: a) localization of points, b) angle normalization, c) distances computed

As can be observed in Fig. 2 faces are localized quite imprecisely and without information about rotation. For the next stages of the analysis it is a key feature to locate the face as precisely as possible, since this allows for correct placement of all face elements. In order to correct inaccuracies and acquire information about the rotation of the face a geometrical normalization based on eye center positions has been added. Although its length can vary for different faces, the perpendicular line placed in the geometric center between eyes almost always sets the face's symmetry axis (Fig. 3b). The very final information provided is the rotation of the face, which can be determined by measuring the angle between eye centers (Fig. 3b).

3 Localization of Characteristic Points in a Face

Anthropometry is a section of biology which deals with measuring human body and its parts. The authors of [1] have created the anthropometric model of human face measuring 300 facial en face images of more than 150 people originating from different parts of the world. The model can be used for localizing facial elements based on positions of eye centers. In Fig. 3a) localization of basic points is shown. They describe well face features which change during emotions' expression.

3.1 Localization of Eyes

For use with anthropometric model eye centers need to be localized very precisely. Unfortunately, most of the methods give location of the pupil rather than the eye center itself. The real difficulty starts when eye is half-shut or closed.

To meet all the requirements and deal with the aforementioned problems a Haar-like feature classifier has been used once again, trained this time specifically to detect left and right eye separately following [8].

Eye localization based on Haar-like features has numerous advantages (ability to operate on grayscale images, ability to detect eyes even if they are not fully open, resistance to changeable lightning conditions, effortless usage and efficiency). Initial attempts to use this method revealed some difficulties. In most cases detection process gives a number of possible candidate positions for one eye. Although several candidates are provided, the correct one is always among them. It has also been spotted that the correct localization of face itself is limited to the rotations of about $\pm 30^{o}$. Consequently, there is no need to look for eyes in the whole face area.

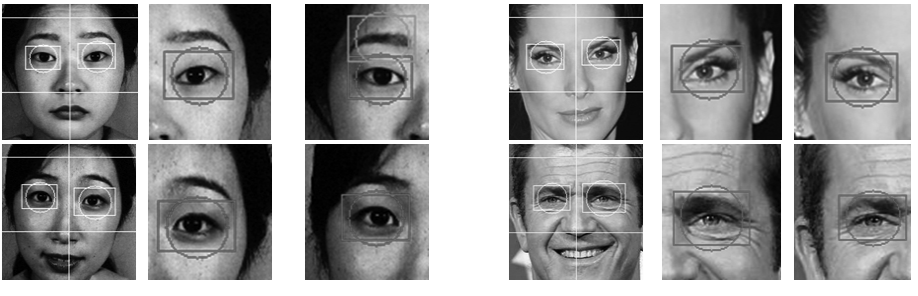

Fig. 4. Eye localization with usage of limited search space (*JAFFE database and Internet) in white* – final eye positions (white circle and rectangle), limiting regions (white lines) *in gray* – candidates for eyes inside limiting region

As shown in Fig. 4, in most cases, usage of limiting regions eliminates the problem of ambiguous localization. The limiting regions have been chosen to cover all the possible locations of eyes for every detected rotation of face. Further improvement can be achieved using a cost function which prefers best-matching candidates or/and those with smallest angle between left and right eye.

Since, at this stage, eye center positions are already known, face can be normalized geometrically. For this purpose the angle between eye centers must be determined and used for face rotations (Fig. 5).

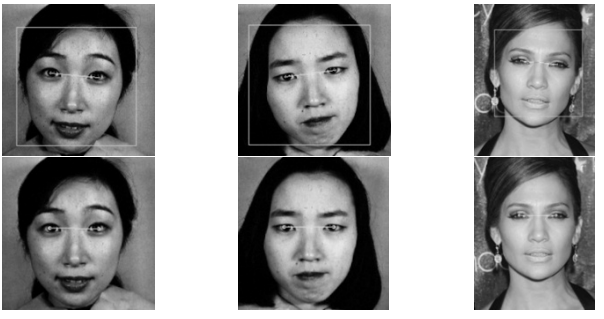

Fig. 5. Images before (*upper panel*) and after (*lower panel*) the corrective rotation based on the angle between eye centers (*JAFFE database and Internet*)

3.2 Localization of Characteristic Regions in a Face

The placements of face and eye centers are known at this stage. Furthermore, the face is already rotated to eliminate the initial rotation. The anthropometric face model is then used to localize remaining facial elements (mouth, eyebrows, nose).

According to the model [1], statistical proportions between certain distances on ordinary human face remain generally constant. Thus, areas of locations of points P2, P5, P11 and P12 (Fig. 3a) can be estimated based on the distance between centers of both eyes P16 and P17 as a shift from the geometrical center between two eyes SC_x (Fig. 3b). After the middles of facial elements have been localized it is necessary to define the size of the regions surrounding those elements. A distance between eye centers P16 and P17 has been used for this purpose. The sizes of regions were defined as to contain the element in every possible deformation. The rectangle surrounding mouth is, in this case, quite large and may, when the mouth is closed, contain large parts of face or even the surroundings of the face itself. The facial element rectangles for different facial expressions are presented in Fig.6.

Fig. 6. Localization of face elements based on anthropometric model (*JAFFE database and Internet*)

Because of the mentioned variations in facial expressions it has been decided that an additional stage of refinement for mouth and nose needs to be added. In those two cases the more precise localizations of elements are determined using specially trained Haar-classifiers as described in [8].

3.3 Localization of Characteristic Points in a Face

Having the isolated regions containing facial elements like eye, nose, mouth and eyebrows it is now possible to find 17 characteristic points selected specifically to represent facial mimic as fully and as economically as possible (Fig.3). For maximal efficiency a hybrid approach is used - every facial element is analyzed in different, individually designed way ensuring the best usage of its characteristics.

3.3.1 Analysis of Nose

Two points, placed in the centers of nostrils, need to be found. Nostrils are two very contrasting, dark circular or elliptical areas. The detection of nostril center points is done by filtration that is able to separate dark circular areas from the brighter background. Here a LoG (Laplacian of Gaussian) filter is used [3]. To make the process immune to different size of nostrils, the filter size is changed and the operation is repeated several times. Nostril central points are found by localizing local maxima and then performing verification based mainly on anthropometric conditions. To further improve the result horizontal Sobel filter is used to visualize nose ending line. The final decision is taken based on a specially formed goal function constructed from all the aforementioned information.

3.3.2 Analysis of Mouth

The mouth is probably the most changeable element of face. At first, the image contrast is adjusted in order to make the mouth darker and the surrounding skin brighter. After this the two points placed in the mouth corners are looked for. Two properties are used in this search. Firstly, these points are usually placed in strong image corners. Secondly, between mouth corners, irrespective of the state of the mouth, always exists a dark “valley”. This darker area can be strengthened using a horizontal Sobel filter. The candidates for corners are obtained as a result of search for strong image corners in the mouth image and then verified based on the properties mentioned above. Having the mouth corners localized it is possible to define the middle part of the mouth and look for outer points of the lips. Here own procedure is used. The image is inverted and binarized with a threshold calculated using an iterative procedure. The contour is further analyzed. Two points located on the mouth contour and having the x coordinates exactly in the middle between the corners are chosen. From those two, the one having the highest y coordinate value is set as the lower lip point, and the other one is chosen as the upper lip point.

3.3.3 Analysis of Eyes

As a result of eye analysis 4 points should be localized. Two in the eye corners, which are localized first, and another two in the middle of eyelids. In the first stage the intensity of eye image is adjusted to make the eyelids more apparent. The image is then converted to a binary version using iteratively calculated threshold [3]. The areas

obtained this way are surrounded by contours using 8-connected contour [11]. The largest contour is further analyzed to choose the outer and inner corner. To deal with misleading shadows in eye corners, the locations are checked against additional anthropometric conditions and if they are not met additional mechanism is used which uses localization of strong image corners. Having the locations of corner points already found, the search for middle eyelid points begins. The eye contour is divided to three parts. Points in the middle part with maximal vertical distance between them are chosen. Those are then verified against anthropometric conditions and further corrected if needed using the mechanism described in [12].

3.3.4 Analysis of Eyebrows

Three points are found for each eyebrow. In the first stage the image colors are inverted as to make eyebrows bright areas on dark background (in most cases). The facial hair that may be inside the region is eliminated by subtracting average background illumination image form the original one. In order to obtain this image morphological operation of opening is used with elliptical structuring element of as much as 10 pixels. Further, the image contrast is increased based on the pixels' cumulative distribution. Then using the Otsu's method [13], the binary image is obtained. The uniform areas in the image are approximated using the 8-connected contour algorithm [11]. The eyebrow contour, which is usually the largest one with its width greater than height, is further analyzed. The inner and outer eyebrow points are chosen first. Then the x coordinate of middle point is calculated as the average of the x coordinates of the aforementioned points. Y coordinate is retrieved as the top-most one. The identified points are checked against anthropometric conditions including the angles and distances between them and corrected if needed.

3.4 Test Results

On average for the JAFFE database the precision of automatic localization of all 17 points used in further emotion analysis was 2.63 pixels (computed as distance to reference hand labeling) or 95,58% (ratio of distance between automatic and hand-labeling divided by the distance between the eyes).

4 Data Normalization for Emotion Detection

In the recognition of emotions the distances between anthropometric points are used with assumption that they change during facial expression of emotions (Fig. 3c). Two types of data normalization are used to obtain person-independent and photo-independent classification of emotions.

Normalization of features [3] – is necessary because perception of emotions is influenced by differences in face proportions. To achieve this average neutral face is generated from all of the faces in the training set belonging to the class labeled as neutral. Feature vector of this face is then subtracted from every face in the data set.

Normalization of data space - it is similar to scaling and fitting the critical features of two photos (nose, eyes, mouth, eyebrows) in a way that would cause them to overlay with as much accuracy as possible. 100% accuracy is not possible for two

different photos, however, it is possible to manipulate any photo in such a way that the regions of eyes, brows and nose will overlay with any other photo. This type of normalization ensures that the photographs in the set do not have to be normalized with respect to position and size of the face.

Normalization of data space is done by averaging two distances between the faces that serve as horizontal and vertical scale factors for every set of face points. First of those distances is one between the two eye centers, while the other connects the averaged eye center to the central point between the nostrils.

5 Classification of Emotions

In experiments recognition of 7 basic facial expressions was tested (fear, sad, angry, happy, surprised, disgust, neutral). For classification of emotions three types of classifiers were used as well as two methods of averaging the results - voting and weighted average.

First of the used classifiers is a 15xQx7 RBF Neural Network [10], chosen for its ease of training as well as a good data fit. Q is a variable number, which depends on the exact data used to train the network. The hidden layer of the network is in fact composed of seven independent subnets, each of them responsible for one of the recognized emotions. Isolation of subnets helps with selection of weights and makes recognition of each emotion a separate task. Neurons' activation function is modeled as:

$$N = exp\left(-\sum_{i=1}^{M}\frac{(x_i - c_i)^2 w_i}{2r^2}\right) \tag{2}$$

where M is the number of features, c is a vector of neuron centers, w is a vector of neuron weights and r is radius of a particular neuron.

The network is trained by adding a new neuron for each averaged emotion of each face in the training set. For every incorrectly recognized sample a new neuron is added until all emotions are recognized correctly. Every neuron center is trained by averaging features of all samples that it should recognize, that is the samples of a particular emotion belonging to a particular face that the neuron was created for.

The radius for every neuron is trained by averaging the distances from all the other neurons in the subnet that the neuron belongs to. This spreads the influence of a particular subnet over the entire space in which the emotion recognized by that subnet could hypothetically appear, according to the training data set. Because RBF networks are linear in their weights it was possible to train it by using the SVD algorithm, without having to restore to time-consuming on-line training. In addition to the RBF network, two other classifiers were used: Naive Bayes Classifier and KNN. It was observed that various classifiers tended to classify emotions incorrectly for different samples. Combining them all together boosted the overall accuracy of emotion classification. This is especially evident for the weighted average method.

5.1 Results and Discussion

The system performs generally well even on automatically labeled data (Table 1 and 2). Please note that best results are obtained for weighted combination of KNN and RBF network. It confirms good generalization ability of RBF network needed for unknown face, while KNN is good for known faces, where exact matching is preferred.

Table 1. Data sets used in experiments

Name of data set	# photos (faces) in test data set	# photos (faces) in training data set
known face	70 (10)	143 (10)
unknown face	22 (1)	191 (9)

Table 2. Percentage of correct facial emotion recognition, [1]KNN size set to 1 for known faces, 11 for unknown; [2]KNN and RBF network only; [3]YM face from JAFFE used as unknown

name of data set	points extraction	Bayes	KNN[1]	RBF network	weighted average[2]	voting
known face	automatic	52.85%	71.42%	65.71%	**74.28%**	75.71%
	manual	64.28%	85.71%	75.71%	**90.00%**	87.14%
unknown face[3]	automatic	54.54%	63.63%	86.36%	**81.81%**	72.72%
	manual	63.63%	86.36%	77.27%	**86.36%**	81.81%

6 Conclusions and Future Plans

Needless to say, there is still room for improvement, especially for the part of the task that deals with classification. Better results can be expected if:

- rotation along Y axis is compensated for (the face is seen not perfectly en face)
- the number of neurons is reduced based on face similarity
- Mahalanobis distances are used instead of Euclidean
- the way in which expressions are classified is changed to allow for the use of different tools for different types of faces (known, unknown)

The facial emotion recognition system has been modified to allow for on-line usage. The feature detection part uses optical flow [14] to track locations and the classifier displays the scores of all 7 facial expression types. The implementation is able to work in real time for 320x240 15 frames/s videos on ordinary PC, without optimization. Further planned modification is to use color information for more precise location of the anthropometric points, computational load optimization and use of higher image resolutions. The system can be easily incorporated into any HCI framework.

Almost 200 years ago Charles Darwin pointed out the importance of emotional expression as part of human communication. Advances in computer vision and technology allows for emotional human-computer interaction. Our preliminary results show that use of anthropometric features eases the task of facial emotion recognition, the points can be precisely located and they contain enough information for recognition of seven basic facial emotional expressions.

References

[1] Sohail, A.S.M., Bhattacharya, P.: Detection of Facial Feature Points Using Anthropometric Face Model. In: Damiani, E., et al. (eds.) Signal Processing for Image Enhancement and Multimedia Processing. Springer, Heidelberg (2007)

[2] Viola, P., Jones, M.J.: Robust Real-Time Object Detection. In: Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing and Sampling, Vancouver, Canada (2001)

[3] Sohail, A.S.M., Bhattacharya, P.: Support Vector Machines Applied to Automated Categorization of Facial Expressions. In: Prasad, B. (ed.) Proceedings of the 3rd Indian International Conference on Artificial Intelligence IICAI 2007, Pune, India, December 17-19 (2007) ISBN 978-0-9727412-2-4

[4] The Japanese Female Facial Expression (JAFFE) Database, http://www.kasrl.org/jaffe.html

[5] OpenCV Wiki-pages, http://opencv.willowgarage.com/wiki/

[6] Jobson, D.J., Rahman, Z., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing 6(7), 965–976 (1997)

[7] Lienhart, R., Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection. In: Proceedings of the International Conference on Image Processing. IEEE, Los Alamitos (2002)

[8] Castrillón-Santana, M., Déniz-Suárez, O., Antón-Canalís, L., LorenzoNavarro, J.: Face And Facial Feature Detection Evaluation (2008), http://gias720.dis.ulpgc.es/Gias/Publications/visapp-2008-1.pdf

[9] Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1) (1997)

[10] Oukhellou, L., Aknin, P.: Optimization of Radial Basis Function Network for Classification Tasks. In: Neurap 1998 IV International Conference on Neural Networks and their Applications, Marseille (1998)

[11] Ritter, G.X., Wilson, J.N.: Handbook of Computer Vision Algorithms in Image Algebra. CRC Press, Boca Raton (1996)

[12] Kuo, P., Hannah, J.M.: An Improved Eye Feature Extraction Algorithm Based on Deformable Templates. In: IEEE International Conference on Image Processing, Genova, Italy, pp. 1206–1209 (2005)

[13] Otsu, N.: A Threshold Selection Method from Gray Level Histograms. IEEE Trans. Systems, Man, and Cybernetics 9(1), 62–66 (1979)

[14] Bouguet, J.-Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker. Intel Corporation, Microprocessor Research Labs (2002)