International Journal of Social Welfare Promotion and Management
Vol. 2, No. 1 (2015), pp.1-6
http://dx.doi.org/10.21742/ijswpm.2015.2.1.02

Emotion Recognition on the Basis of Eye and Mouth

Arnima Chaurasia[1], ShailendraPratap Singh[2] and Divya Kumar[3]

[1]*Computer Science Department, Banasthali Vidyapith, Jaipur, India*
[2,3]*Computer Science and Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India*

Abstract

It is very interesting to identify human facial expression in the domain of image processing. Since last decade, a lot of research has been focused on this area. The role of this field in human-computer interaction is increased its importance multifold. In this paper, we have used manual database of eye and mouth which is further classified on the basis of emotion. The expressed emotions included anger, sadness, happiness, and neutral. This classified database is then used to retrieve corresponding facial emotion. We have used Haar-cascade technique for detecting the eyes and mouth and nearest-neighbor for classification of emotion using Euclidean distance. The accuracy of this technique is measured on 400+ sample images taken arbitrarily. From the result, we conclude that presented algorithm is able to precisely classify images on the basis of eye and mouth only.

Keywords- *Emotion, Facial feature, Haar-cascade, Nearest-neighbor, Image retrieval*

1. Introduction

Emotion is involved in our daily life. According to Cabanac [1], an emotion is a complicated psychological state that involves three different components a subjective experience, a physiological response, and a behavioral response. It is composed of two words, E-Motion that means Energy in Motion. It includes anger, joy, panic, grief and many more feelings. There is no universal accepted definition of emotion in literature. Emotions are considered as complex phenomena and have combination of subjective feeling, physiological response (body) and expressive behavior [4]. Subjective feeling can't be measured, however physiological response of emotion can be measured from pounding heart, facial expression and blood rushing to the face. Face, voice and physiological behaviour are modalities which are used to recognize the emotion. Each of which has its own weakness and strength. Facial expressions are important aspect in human- machine intersection. Human-computer interaction provides ability to the machines to recognize human emotion which helps in perception, decision making, learning and prediction which influence the rational thinking. Mehrabina (1968) [8] has shown that human convey messages only 5% times by language and 55% times by facial expression. Later, Ekman et. al [3] has reported that facial expression is universal tool to express their emotion. Kobayashi and Hara [5] studied a machine recognition method using neural network for static image facial expression. Kobayashi et al. [5] developed a dynamic recognition of human facial expression by taking images of 6 basic facial expressions and sequentially changing these images to detect facial expression. Metallinou et al. [9] have used facial as well as vocal modalities to achieve improved emotion recognition system. They have used Gaussian Matrix

Article history:
Received (July 02, 2015), Review Result (September 15, 2015), Accepted (October 11, 2015)

Print ISSN: 2205-8435, eISSN: 2207-5321 IJSWPM

Models (GMM) to model each modality and have used Bayesian classifier weight scheme and support vector machines to combine multiple modalities. Marc Lanze Ivan et al. [2] developed a multimodal emotion recognition system that was trained using a spontaneous Filipino emotion database. The system could extract voice and facial feature and then use support vector machine to classify correct emotion label.

In this paper, we have discussed a multiple attribute image retrieval system on emotion, in which we create a manual database of emotion and use this database to detect exact facial expression. We have described a general approach to retrieve images on the basis of facial expression which are detected on the basis of eye and mouth feature. To achieve this objective flow of the paper goes in this way: In Section-2 we have outlined basic functioning of haar-cascade for facial feature detection, in Section-3 we have described our proposed algorithm and result of experiments are discussed in Section-4.

2. Feature detection using Haar cascades

In this section we will discuss feature detection using Haar-cascade [12]. Face detection is first step in the face recognition system. Purpose of the face detection is to localize and extraction of the facial feature leaving out the background information. Paul Viola and Michael Jones proposed an effective object recognition method using Haar feature based cascade classifiers [6]. Haar-cascade is a feature detection method for visual object with higher detection rate. They have introduced a new image representation (known as Integral Image), which allows computation of feature very quickly. They have also proposed AdaBoost based learning algorithm [13] which usages a small number of essential visual feature from a large set and produces efficient classifier. Using Haar-cascade allow to detect faces up to 15 frames per second. It does not require any auxiliary information to be stored (information such as pixel color in image or image difference in video sequence). It works only with information present in gray scale image and does not work directly with image intensities [10]. Integral image is computed using basic operation on every pixel and allows fast feature evaluation on image. Paul Viola and Michael Jones [6] have also suggest a method to construct a classifier by using a small number of important features using AdaBoost. This technique reduces large extent of computation used in classification. To make classification efficient and fast, only a small number of features is selected in learning process. Cascade [7] function is trained using a large number of positive and negative images. Then this classifier is used to detect object in the sample image. For more details on Haar-feature, we redirect reader to [11]. Every feature is a single value which is calculated by subtracting sum of pixels under black area from sum of pixels under white area. There are large set of pixels in an image and for all possible size of rectangles, there are a greater number of feature then the number of pixels. To calculate this large number of features, integral image concept comes as a handy tool. To calculate sum of pixels using integral image requires just four pixels. Then we need to select fewer feature out of these large number of features using AdaBost algorithm. Initially we apply all the feature and find out best threshold to categorizes positive and negative faces. At starting stage, all images have identical weight but after every classification, miss-classified-image weight is increased. We repeat this process until we get required certainty with required number of features. Finally, classifier is a weighted sum of all these features. We don't apply all the feature on a window, instead we collect feature in different stages and apply next stage-feature only when it passes previous stage. The above-mentioned technique is called Cascading of Classifier.

Steps to create a Haar-like classifier

1. Set of positive and negative training images.
2. Mark confident images by using objectmarker.exe tools.
3. Create a .vec (vector) file which is based on positive marked images using createsamples.exe.
4. Guide the classifier using haartraining.exe.
5. Run the classifier using cvHaarDetectObjects().

3. Proposed work

This section details out the algorithm used to classify the frontal images on the basis of only eyes and mouth. Features are going to be extracted is one of the most significant steps to successfully analyze and recognize facial expressions automatically. The system can be broadly categorized in different stages: query processing, image classification and image retrieval stage. Haar-cascade and nearest-neighbor method is used for feature detection and image classification. Eyes and mouth are identified as the critical features and these features are used to classify the emotion. These features are detected with the help of Haar-cascade method. After feature detection is performed, a nearest-neighbor [14] approach is used to retrieve the emotions contained within the face. In this work, the system has efficiently recognized the four universal emotions from face images. First of all, we have created a database using algorithm outlined in algorithm 1. For this purpose, we have taken 91 images with happy faces, 81 images with sad faces, 77 images with angry faces and 78 images with neutral faces. Then we have used algorithm 2 to recognize emotion present in a given image. Calculated average Euclidean distance is depicted in the [Table 1] where edi represent calculated Euclidean distance for ith image and ED is minimum Euclidean distance for all value of i.

Algorithm 1 Database creation using Haar-cascade
Generate database *D* composed of *Di*
Where *i € {Happy, Sad, Neutral, Angry}* and *Di*, contains two set D_iE = *Set of eyes* and D_iM = *Set of mouths DiE = {DiE***1**, *DiE2,...,*DiE_n*}* and *DiM = {DiMi, DiM2,...,* DiM_n*}*
for each facial feature *i*
Generate Sub database D_iE and D_iM as Select Image with feature *i*
Apply Haar-cascade to detect facial feature (eye and mouth) Copy detected feature in D_iE and D_iM
end for

Algorithm 2 Multi-attribute Image Retrieval on Emotion
Step 1 for each experiment with facial feature *k*
Where *k € {Happy, Sad, Neutral, Angry}*
Take I_k as input image
Applying *haar — cascade* to I_k and generate template I_kE and I_kM**for** Vi € *{Happy, Sad, Neutral, Angry}*

$$ed_i = \text{^}2(edof I_kM \text{ with } D_iMj \; j=i \; +ed \text{ of } I_kE \text{ with } Di,Ej)$$

end for

Calculate Average value of *edi*
end for
Step 2 Find minimum $ED = MIN(edhappy, ed_{sad}, ...)$
Step 3 return label of I_k as *ED*

4. Experimental result

To demonstrate our algorithm, we have performed two experiments. In experiment 1, we have taken a single image of each category and then performed algorithm 1 to extract facial feature and then used algorithm 2 to calculate average Euclidean distance. It is clearly observed that for happy face image, average Euclidean distance is 2.1281 from happy templates, 2.5292 from sad templates, 2.7361 from angry templates and 2.6957 from neutral templates. Due to minimum distance form happy template, assigned label for this image is 'Happy'. Smiler calculation done for sad, angry and neutral faced images and results are depicted in Table 1. For the vague images, average calculated distance is much more then threshold value and hence we not assign any label to this type of image. Experiment 2 is performed by taking a set of 400 images of each category and we try to label all the image base on facial expression. When actual label is matched against predicted label, the accuracy for the happy faces image is 91.8%, for sad face image is 85.7%, for angry face image is 70.8% and for neutral face image is 70.9%. Result for experiment 2 is shown in figure 1 where H represent Happy, S represent Sad, A represent Angry and N represent Neutral. Accuracy of retrieved images are calculated as:

$$Accuracy \quad \frac{Total\ number\ of\ relevant\ images}{Total\ number\ of\ retrieved\ images} \tag{1}$$

Table 1. Calculated average Euclidean distance

Face to be labeled			Euclidean distance from templates			
			Happy	Sad	Angry	Neutral
		Eyes	1.0901	1.0170	1.4210	1.2734
	Happy	Mouth	1.0380	1.5122	1.3421	1.4223
		Total	2.1281	2.5292	2.7631	2.6957
		Eyes	0.2031	0.1612	1.8025	0.4010
	Sad	Mouth	1.3161	0.2232	0.8165	1.2160
		Total	1.5192	0.3844	2.6190	1.6170
Exp for		Eyes	1.0471	0.8042	0.3262	0.6240
	Angry	Mouth	1.4321	0.8012	0.2334	1.1440
		Total	2.4792	1.6054	0.5596	1.7680
		Eyes	0.5107	1.0610	1.0556	0.6132
	Neutral	Mouth	1.4001	1.0131	1.0542	0.5012
		Total	1.9108	2.0741	2.1098	1.1144
	Vague		47.2840	51.8419	38.0792	43.6952

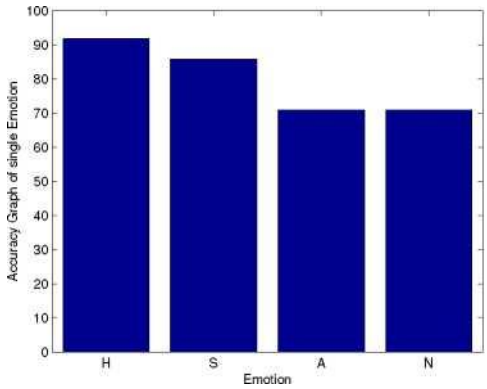

Figure 1. Accuracy graph of single emotion

5. Conclusion

Many facial features are being used to classify emotion in the image which are cheeks, eye, mouth, lips and nose. We have demonstrated that only eye and mouth can give accuracy up to 91%. We have used k-Nearest Neighbors and Haar Cascades to detect facial feature and classification. In this system, classification of images can be obtained on the basis of only two facial features which are eye and mouth. Result shows that the system was able to identify the expressions accurately from the images.

References

[1] M.Cabanac, "What is emotion?" Behavioural processes, vol.60, no.2, pp.69-83, **(2002)**

[2] M.L.I. C. Dy, I.V.L. Espinosa, P.P.V. Go, C.M.M. Mendez and J.W. Cu, "Multimodal emotion recognition using a spontaneous Filipino emotion database", In Human-Centric Computing (HumanCom), 2010 3rd International Conference on IEEE, pp 1-5, **(2010)**.

[3] P. Ekman and W.V. Friesen, "Unmasking the face: A guide to recognizing emotions from facial clues", Ishk, **(2003)**.

[4] P.R. Kleinginna Jr and A.M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition", Motivation and emotion, Vol. 5, No. 4, pp. 345-379, **(1981)**.

[5] H. Kobayashi, F. Hara, S. Ikeda and H. Yamada, "A basic study of dynamic recognition of human facial expressions", In Robot and Human Communication, 1993, Proceedings, 2nd IEEE International Workshop on, pp. 271-275, **(1993)**.

[6] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection", In Image Processing. 2002. Proceedings. 2002 International Conference on IEEE, Vol. 1, pp. I-900, **(2002)**.

[7] C. Lim, J.N. Bearden and J.C. Smith, "Sequential search with multiattribute options", Decision Analysis, Vol. 3, No. 1, pp. 3-15, **(2006)**.

[8] A. Mehrabian, "Nonverbal communication", Transaction Publishers, **(1972)**.

[9] A. Metallinou, S. Lee and S. Narayanan, "Audio-visual emotion recognition using gaussian mixture models for face and voice", In Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on IEEE, pp. 250-257, **(2008)**.

[10] C.P. Papageorgiou, M. Oren and T. Poggio, "A general framework for object detection", In Computer vision, 1998. Sixth International Conference on ,IEEE, pp. 555-562, **(1998)**.

[11] P. Viola and M. Jones,"Rapid object detection using a boosted cascade of simple features". In Computer

Vision and Pattern Recognition, CVPR 2001, Proceedings of the 2001 IEEE Computer Society Conference on IEEE, Vol. 1, pp. I-511, **(2001)**.

[12] P.I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers", Journal of Computing Sciences in Colleges, Vol. 21, No. 4, pp. 127-133, **(2006)**.

[13] Y.W. Wu and X.Y. Ai, "Face detection in color images using adaboost algorithm based on skin color information", In Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on IEEE, pp. 339-342, **(2008)**.

[14] H. Zhang, A.C. Berg, M. Maire and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition", In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on IEEE, Vol. 2, pp. 2126-2136, **(2006)**.