FACE ORIENTATION DETECTION IN VIDEO STREAM BASED ON HARR-LIKE FEATURE AND LQV CLASSIFIER FOR CIVIL VIDEO SURVEILLANCE

Zhiguo Yan[1, 2], Fang Yang[3], Jian Wang[2], Yongjie Shi[2], Chao Li[3], Mingxia Sun[2]*

[1]*Postdoctoral Research Station of Computer Science and Technology, Fudan University, 200433, Shanghai, P. R. China*

[2]*Research & Development Centre of Internet of Things, The Third Research Institute of Ministry of Public Security, 201204, Shanghai, P. R. China*

[3]*Department of information and electronic engineering, Shanghai Normal University, 200235, Shanghai, P. R. China*

yzgemail@gmail.com, [*]wjconan@yeah.net

Keywords: Face Orientation, Eye Detection, Haar, Lqv, Video Stream

Abstract

Most face recognition and tracking techniques employed in surveillance and human-computer interaction (HCI) systems rely on the assumption of a frontal view of the human face. In alternative approaches, knowledge of the orientation angle of the face in captured images can improve the performance of techniques based on non-frontal face views.

Face orientation detection plays important role in city surveillance video for the successive specific application, such as the face identification, face recognition and screening face snapshot image for saving the storage volume. In this paper，we propose a kind of method on face orientation by combining Haar-feature and LVQ technique. First, we execute the eye location based on the haar-like feature. Then, we divide the face image into several sub-images and statistical information of the binary sub-image at the eye location. After acquiring the statistical pixel distribution, we exploit the LVQ classifier to execute the classification on face orientation. According to the result, the algorithm we proposed can detect correctly up to 95%. By executing the face orientation classification, we can get the upright fontal face image with the best recognizable and distinctive quality for the further application.

1 Introduction

In the past two decades, automatic human face image analysis and recognition has become one of the most important research topics in computer vision and pattern recognition. Because of tremendous potential application, topics such as face detection, face identification and recognition and facial expression analysis have attracted more and more attention. Among these research topics, one fundamental but very important problem to be solved is face orientation detection. Face orientation detection is the premise of face recognition.

Face detection refers to determining the presence and location of faces in an image. Human face detection is very important in a face recognition system, quite useful in multimedia retrieval [1]. Numerous face detection methods have been proposed for frontal face detection such as region-based face detection [2], triangle-based approach [3], feature-based method [4] and template matching method [5]. The aforementioned methods limit themselves to dealing with human faces in frontal view. So, how to detect the face orientation and store the upright frontal face image for the further face recognition is of significance in civil video surveillance.

Orientation is one of the basic characteristics in image understanding and pattern analysis. Many approaches have been proposed to solve the above problem. Jie Zhou [6] proposed orientation histogram for orientation analysis. Chia-Feng Juang found that self-organizing fuzzy network with SVM [7] worked well in colour image detection. R. Brunelli [8] developed a good method to estimate the pose of a face, limited to in-depth rotations. By comparing matching in transparency displays, Paolo Martini [9] proposed an approach based on that for orientation tuning of human face.

However, these novel approaches limit in special situations. Rotated faces can only be detected within the image plane when using orientation histogram; Self-organizing fuzzy network with SVM method uses colour as an eigenvalue, so it's hard to work well in gray images; The algorithm in [8] needs appropriate template because the algorithm he presented requires the location of one of the eyes be approximately known, together with the direction of the interocular axis; The stimuli that used in Paolo's experiments is hard to obtain.

Face detection and recognition is based on the operation of eye localization. To some extent, the performance is determined by the correctness eye detection. Many researches have been conducted on human eye detection, and several algorithms have been proposed, such as region segmentation [10], Template Matching [11], AdaBoost algorithm [12] and so on. Region segmentation is a simple but rewarding algorithm, which attracts numerous researchers. The threshold is not easy to decide appropriately though it is the key to correct eye detection; Template matching method costs expensive computation when normalize the scale and orientation of face image [13]; Inho Choi used AdaBoost algorithm in eye detection and eye blink detection almost successfully, but the size of the sub-regions is hard to predict when dividing them to build image pyramid. As eye localization is concerned, making use of the distinct characteristic of the eyeball is the key step no matter what method is used. Document [14] devel-

oped a fast eye localization method based on a New Haar-like (ref) feature, which proved impressive in eye detection. However, it still needs further research to find a more adequate threshold segmentation method in order to make the algorithm more robust.
Most face recognition and tracking techniques employed in surveillance and human-computer interaction (HCI) sys-tems rely on the assumption of a frontal view of the human face. In alternative approaches, knowledge of the orientation angle of the face in captured images can improve the performance of techniques based on non-frontal face views.
Our approach is partly motivated by the work of[15, 16], in which they use the Haar-like feature and the boosting classification strategy to locate the eyes. On the basis of their achievement, we go a step further by using the Learning Vector Quantization (LVQ) classifier to detect the face orientation.
In this study, we present a facial orientation recognition method based on LVQ. In which, the features vector is comprised based on the features detection such as the edge of eyes, which was processed by edge detection method such as method based on Haar-like feature.
The organization of this paper is arranged as follows: Methodology will be introduced in Section 2. Section 3 elaborates the materials and the experiment setting. The experimental results and conclusions are given in Section 4 and 5, respectively.

2 Methodology

Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector.
Historically, working with only image intensities (i.e., the RGB pixel values at each and every pixel of image) made the task of feature calculation computationally expensive. A publication by Papageorgiou et al. [17] discussed working with an alternate feature set based on Haar wavelets instead of the usual image intensities.
Viola and Jones [18] adapted the idea of using Haar wavelets and developed the so-called Haar-like features. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. For example, let us say we have an image database with human faces. It is a common observation that among all faces the region of the eyes is darker than the region of the cheeks. Therefore a common Haar feature for face detection is a set of two adjacent rectangles that lie above the eye and the cheek region. The position of these rectangles is defined relative to a detection window that acts like a bounding box to the target object (the face in this case).
In the detection phase of the Viola–Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such a Haar-like feature is only a weak learner or classifier (its detection quality is slightly better than random guessing) a large number of Haar-like features are necessary to describe an object with sufficient accuracy. In the Viola–Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier.
The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time.
A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called 2-rectangle feature. Viola and Jones also defined 3-rectangle features and 4-rectangle features. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a 2-rectangle feature can indicate where the border lies between a dark region and a light region.
One of the contributions of Viola and Jones was to use summed area tables, which they called integral images. Integral images can be defined as two-dimensional lookup tables in the form of a matrix with the same size of the original image. Each element of the integral image contains the sum of all pixels located on the up-left region of the original image (in relation to the element's position). This allows to compute sum of rectangular areas in the image, at any position or scale, using only four lookups, see fig. 1:

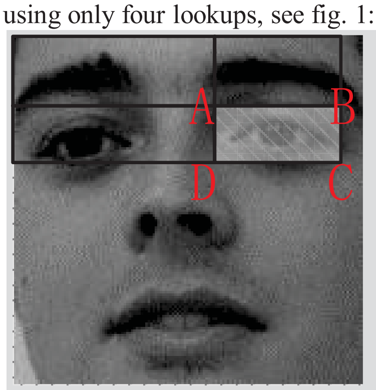

Figure 1. The sketch of computation on integral image

Equation 1 calculates the sum of the shaded rectangular area：

$$sum = I(C) + I(A) - I(B) - I(D) \qquad (1)$$

where points *A, B, C* and *D* belong to the integral image *I*, as shown in Fig. 1.
Each Haar-like feature may need more than four lookups, depending on how it was defined. Viola and Jones's 2-rectangle features need six lookups, 3-rectangle features need eight lookups, and 4-rectangle features need nine lookups.
Lienhart and Maydt [19] introduced the concept of a tilted (45°) Haar-like feature. This was used to increase the dimensionality of the set of features in an attempt to improve the detection of objects in images. This was successful, as some of these features are able to describe the object in a better way. For example, a 2-rectangle tilted Haar-like feature can indicate the existence of an edge at 45°.
Messom and Barczak [20] extended the idea to a generic rotated Haar-like feature. Although the idea sounds mathemati-

cally sound, practical problems prevented the use of Haar-like features at any angle. In order to be fast, detection algorithms use low resolution images, causing rounding errors. For this reason, rotated Haar-like features are not commonly used.

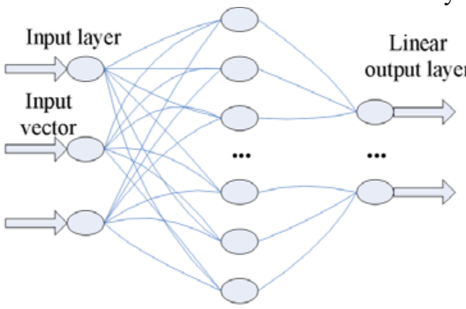

Figure 2. LVQ network model

LVQ network model is shown as Figure 2. A LVQ neuron network consists of three layers, i.e. input, competition and linear output. The network input layer is completely connected to the competition layer while the competition layer is partially connected to the linear output layer. Different connection exists between each output neuron group competition neuron group and the fixed value of them is 1.
The connection reference value of input and competitive neuron vector is established (a reference vector is appointed to each competitive neuron). In the training process, the network weights will be modified.
Both competition neurons and linear output neurons are of binary output value. When an input mode was sent to a network, when the reference vector is closest to input mode, the competitive neurons are started and win the competition.
Thus the generation of "1" is allowed, and other competitive neurons are forced to generate "0". Competition neurons, including the winning neuron group, connected to output neurons also generate "1" while other neurons output generates "0". The output that generated "1" issues the input mode to the output neurons, each output neuron produces different classes.
LVQ can be understood as a special case of an artificial neural network, more precisely, it applies a winner-take-all Hebbian learning-based approach. It is a precursor to Self-organizing maps (SOM). LVQ was invented by Teuvo Kohonen [21].
An LVQ system is represented by prototypes which are defined in the feature space of observed data. In winner-take-all training algorithms one determines, for each data point, the prototype which is closest to the input according to a given distance measure. The position of this so-called winner prototype is then adapted, i.e. the winner is moved closer if it correctly classifies the data point or moved away if it classifies the data point incorrectly.
An advantage of LVQ is that it creates prototypes that are easy to interpret for experts in the respective application domain. LVQ systems can be applied to multi-class classification problems in a natural way. It is used in a variety of practical applications [22].
A key issue in LVQ is the choice of an appropriate measure of distance or similarity for training and classification. Recently, techniques have been developed which adapt a parameterized distance measure in the course of training the system, see [23] and references therein.
There exist two classic approaches to train the LVQ, named LVQ1 and LVQ2 respectively[24].
LVQ can be a source of great help in various classification tasks.

3 Experiments

To verify the effective of the proposed method for face orientation detection, we utilized it on real-time video stream and face image database respectively. For the simplicity, we mount the USB camera on the laptop to test the effectiveness of the Haar-like feature on eye detection. Further study shows the proposed technique also has good performance on the IP camera and other types of surveillance cameras.

Figure 3. Eye location by using the Haar-like feature in real-time video stream

Fig. 3 shows the result of eye detection in a range of head rotation. We can infer coarsely that the Haar-like feature is advisable method to locate the eyes. Besides the above-mentioned photos acquired by the USB camera, photos of 10 people (5 male and 5 female) were obtained from PIE Face Database of CMU [25] from online website posted for scientific use. Each subject has five different head postures as “left”, “slight left”, “frontal”, “slight right” and “right”. One example of this part of data was presented as picture 1.

Figure 4. The illustration of 5 face orientations

As Fig. 4 shows explicitly, when people face different orientations, the eye location and the distance between two eyes on the images will vary dramatically following the head rotating. Considering this fact, we adopt the statistic information about eye location as the input to the LVQ classifier. We denote the 5 orientations as “1”、 “2”、“3”、“4” and “5” respectively as the output of the LVQ classifier.

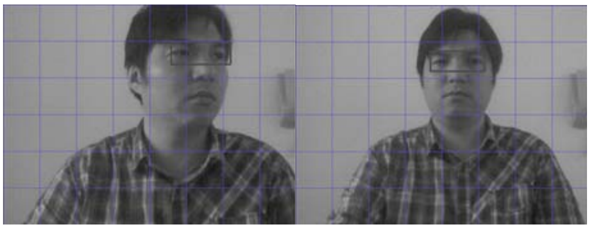

Figure 5. The illustration of eye location and the grid of sub-images

As Fig.5 shows, we divide the face image into 6*8 sub-images and execute the edge detection by Canny operator, then accord the eye horizontal location denoted by the rectangle we calculate the sum of “1” pixels in the corresponding horizontal 8 sub-images. The sum is adopted as the feature feed to the LVQ classifier.

As the above mentioned, the total face images we got is 50 (10 persons, each person with 5 images). We randomly select 30 images for training from the images. The left 20 images are used as the test samples to verify the performance of the proposed technique on face orientation detection.

4 Result

In the test phase, we respectively exploit the LVQ1 and LVQ2, the final recognition accuracy with LVQ1 is 90% (two test sample were classified wrongly among twenty test samples) and 95% with LVQ2. Because of the very limitation of experiment setting and sample number, we can’t affirm that LVQ2 is superior to LVQ2. But the experiment shows that the proposed technique on face orientation detection is a kind of promising technique. The further study will focus on the comparison between the proposed technique and other prevalent methods.

5 Conclusion

By utilizing the proposed face orientation detection technique, we can provide feedback to any user interface that could benefit from knowing whether a user is paying attention or is distracted to the camera or the video showed on the screen mounted with the camera. We can extent this technique to tutoring systems, computer games, even active safety protection systems that monitor driving fatigue. Besides these applications, this technique also can be used in evaluating the effect of commercial ads by detecting the upright face.

Acknowledgments

Our research was supported by the Innovation plan for practical application of the Ministry of Public Security (No. 2012YYCXGASS125), the plan of absorption and creation on the imported technology of Shanghai Municipal Commission Of Economy and Informatization (No. 12GA-19)，Major program of Science and Technology Commission of Shanghai Municipality (No.12510701900), National High-tech R&D Program of China “863 Program” (No. 2013AA010303) and Ministry of Public Security of China (No. 2012YYCXGASS125).

References

[1] Chellappa, R., Wilson, C. L., and Sirohey, S., "Human and machine recognition of faces: a survey" Proceedings of the IEEE, 1995. 83(5): p. 705-741.

[2] Ayinde, Olugbenga and Yang, Yee-Hong, "Re-gion-based face detection" Pattern Recognition, 2002. 35(10): p. 2095-2107.

[3] Lin, Chiunhsiun and Fan, Kuo-Chin, "Triangle-based approach to the detection of human face" Pattern Recognition, 2001. 34(6): p. 1271-1284.

[4] Yow, Kin Choong and Cipolla, Roberto, "Fea-ture-based human face detection" Image and Vision Computing, 1997. 15(9): p. 713-735.

[5] Jin, Zhong, Lou, Zhen, Yang, Jingyu, et al., "Face detection using template matching and skin-color information" Neurocomputing, 2007. 70(4–6): p. 794-800.

[6] Zhou, Jie, Lu, Xiao guang, Zhang, David, et al., "Orientation analysis for rotated human face detection" Image and Vision Computing, 2002. 20(4): p. 257-264.

[7] Juang, Chia-Feng and Shiu, Shen-Jie, "Using self-organizing fuzzy network with support vector learning for face detection in color images" Neurocomputing, 2008. 71(16–18): p. 3409-3420.

[8] Brunelli, R., "Estimation of pose and illuminant direction for face processing" Image and Vision Compu-ting, 1997. 15(10): p. 741-748.

[9] Martini, Paolo, McKone, Elinor, and Nakayama, Ken, "Orientation tuning of human face processing esti-mated by contrast matching in transparency displays" Vision Research, 2006. 46(13): p. 2102-2109.

[10] Shan, S.G, Gao, W., and Chen, X.L., "Facial Fea-ture Extraction Based on Facial Texture Distribution and Deformable" Journal of Software, 2001(04): p. 570-577.

[11] Shi, H.R, Liu, W.B, and Yu, S.L., "An Improved Template Matching Eye Location Method" Computer Engineering and Applications, 2004(33): p. 44-45+214.

[12] Inho, Choi, Seungchul, Han, and Daijin, Kim. Eye Detection and Eye Blink Detection Using AdaBoost Learning and Grouping. in Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on. 2011.

[13] Monzo, David, Albiol, Alberto, Sastre, Jorge, et al., "Precise eye localization using HOG descriptors" Machine Vision and Applications, 2011. 22(3): p. 471-480.

[14] [Lienhart, R. and Maydt, J., "An extended set of Haar-like features for rapid object detection" 2002 International Conference on Image Processing, 2002.: p. 900-903.

[15] Chen, Shuo and Liu, Chengjun, "Discriminant Analysis of Haar Features for Accurate Eye Detection" 15th International Conference on Image Processing, Computer Vision, and Pattern, 2011.

[16] Kasinski, Andrzej and Schmidt, Adam, "The Architecture of the Face and Eyes Detection System Based On Cascade Classifiers" Computer Recognition Systems 2: Advances in Soft Computing, 2007. 45: p. 124-131.

[17] Papageorgiou, Oren, and Poggio, "A general framework for object detection" International Conference on Computer Vision, 1998.

[18] Viola and Jones, "Rapid object detection using a boosted cascade of simple features" Computer Vision and Pattern Recognition, 2001.
[19] Lienhart, R. and Maydt, J, "An extended set of Haar-like features for rapid object detection" ICIP02, 2002: p. 900-903.
[20] Messom, C.H. and Barczak, A.L.C, "Fast and Efficient Rotated Haar-like Features Using Rotated Inte-gral Images" Australian Conference on Robotics and Automation 2006: p. 1-6.
[21] Kohonen, T., "Learning vector quantization" The Handbook of Brain Theory and Neural Networks, 1995: p. 537-540.
[22] Fahad and Sikander, "Classification of textual documents using learning vector quantization" Infor-mation Technology 2007. 6(1): p. 154-159.
[23] Schneider, P., Hammer, B., and Biehl, M., "Adaptive Relevance Matrices in Learning Vector Quanti-zation" Neural Computation, 2009. 21: p. 3532-3561.
[24] Somervuo, Panu and Kohonen, Teuvo, "Self-Organizing Maps and Learning Vector Quantization for Feature Sequences" Neural Processing Letters, 1999. 10: p. 151-159.
[25] "PIE face database"
[26] http://vasc.ri.cmu.edu/idb/html/face/.