MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
3-21

Multi-view Human Head Detection in Static Images

Maolin Chen
CASIA-SAIT HCI Joint Lab.
Institute of Automation, Chinese Academy of Sciences
Beijing, P. R. China
Email: mlchen@hci.ia.ac.cn

Gengyu Ma
CASIA-SAIT HCI Joint Lab.
Institute of Automation, Chinese Academy of Sciences
Beijing, P. R. China
Email: gyma@hci.ia.ac.cn

Seokcheol Kee
Computing Lab.
Samsung Advanced Institute of Technology
Seoul, South Korea
Email: sckee@samsung.com

Abstract

Detecting humans in images is an important task in image or video processing field. This paper proposes an illumination and poses invariant method for human head detection. The principle stands on the evidence that human head has similar shape, which is composed of several edge segments at certain orientations on head contour. Firstly, head contour are labeled manually on the head sample images, then contour image is extracted and transformed into gradient space, finally gradient image is decomposed into several scalar magnitude images each with the same quantized phase. Standard boosting algorithm is utilized to search for weak classifiers. To our knowledge, we built the first multi-view head detector in the world with gradient information only. Experiment testifies good performance of proposed method.

1 Introduction

Human detection in static images or moving cameras is a difficult task. In static images, there is no motion information can be used for extracting human region. And face detection technology is limited by the frontal or profile face images. In moving cameras, the conventional background subtraction doesn't work anymore. Other features such as color and texture can help to detect the human, but it is also limited by the illumination or camera parameters configuration. Based on this, we proposed human detection solution in gradient space, which is less dependent on the human pose and illumination condition. We utilized the component based human detection, and introduced our work on head detection in this paper.

Many researchers concentrate on human detection and tracking. 3D human model is employed for segmentation of foreground regions into individual humans after the background subtraction [1]. Human can be simplified into composition of many line segments. For those line segments in image, which can be configured into human pattern, can be selected as human candidates [2]. Tomaso Poggio invented a method to search humans in image. Walvelet coefficients are extracted and SVM classifiers are trained to model the human intrinsic pattern [3]. Based on the face detection technology, Paul Viola extended his work into human detection in images or video [4]. He claimed the first detector integrating image intensity information and motion information for human detection. The method proposed in this paper also trains the human classifier with boosting algorithm, but in different feature space, which is less dependent on human appearance and illumination conditions.

2 Samples Preparation

As image samples can't be input directly for training process which requires preprocessing. In this section, samples preparation is introduced.

2.1 Variations generating

Generally, statistical learning method requires large number of training samples. In order to capture object patterns, we should label the truth data manually which is a very hard work. Although much effort is done, the samples are still limited comparing to the requirement. Here we try to model the movement human head to create more samples on the basis of labeled truth data.

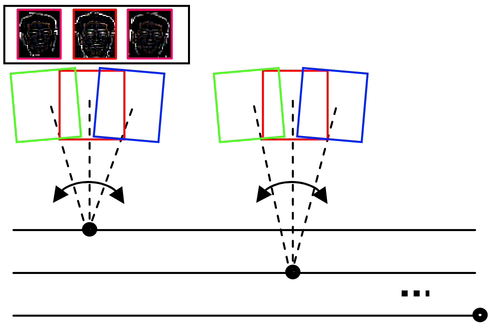


Fig.1 Variations generation of training samples

Head images are firstly transformed into gradient space. As shown in Fig.1, magnitude images are rotated at different radius to model the human head movement in real scene. For small radius, only head of human body rotates using human neck as the rotating pivot; for large radius, heads are rotated along with the rotation of human body. Based on this, head images are rotated at different radius to generate the sample variations.

2.2 Phase quantizing

After variations generating, more times head samples can be obtained as positive samples. Negative samples can also be collected from home albums, photos and internet images, and transformed into gradient space. All phase images of training samples are then quantized into predefined number of bins. In order to suppress the noise and reduce the noise lead by quantization, smoothing filter operation is applied to smooth the quantized orientation image [5]. So, there are two image matrixes representing the head contour magnitude and orientation respectively. For easy computation, the magnitude image is decomposed into multiple channels images each with the same orientation.

3 Model Training and Head detection

Similar to the methods in [4], this paper also uses the boosting method to train the human model, not in the image intensity space, but in the gradient space. To our knowledge, it is the first detector in the world using image gradient information to detect the human. Heads are modeled with profile view and non-profile views, which are represented by two different shape models. Fig.2 shows human head detection framework, including model training and head detection in images.

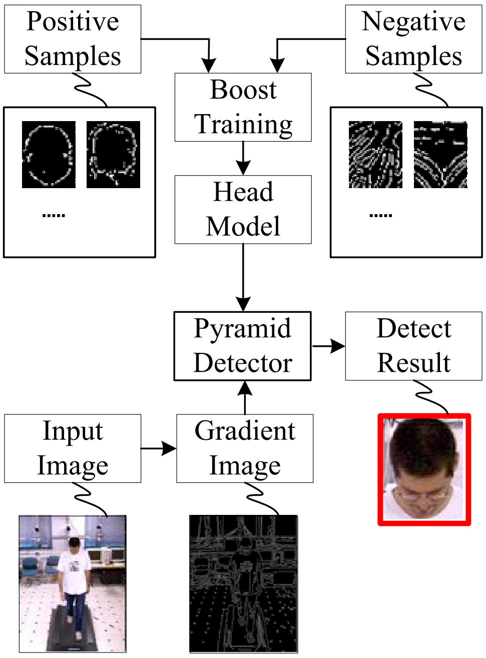


Fig.2 Framework for model training and detection

Firstly, positive and negative samples are prepared and input to the training module. In the training module, large number of head contour features is extracted to form a features pool. Then, standard boosting algorithm searches the pool for weak classifiers and boosts them into strong ones. On the detection, the input image is transformed into gradient space; pyramid detector will scan the image for head candidates.

3.1 Features definition

Feature window is composed of one or several sub-windows which are neighboring or symmetric. Detector will scale them along with sliding them in image space. As illustrated in Fig.3, four types of feature windows are generated, including single sub-window, double sub-windows with relative offset, three sub-windows with relative offset at horizontal and vertical directions.

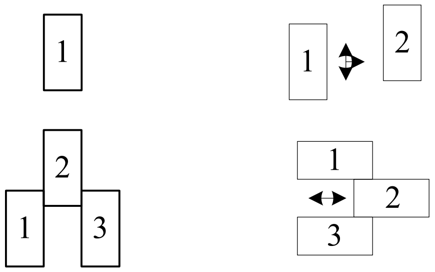


Fig.3 Generating feature windows

As the variability of local edges, local image gradient is quantized into N orientations, denoted as ‘Orient0’, ‘Orient1’, and ‘Orient2’… Inspired by the method in [5], assuming feature window is w, gradient feature in it can be defined as equation (1). In order to accommodate the diversity of each orientation image from all the samples, the operators defined in equation (2) are based on the local histogram.

$$f_i = \sum_{(x,y)\in W} G^i(x,y) \tag{1}$$

$$(OP)_{ij} = \{+, -, \times, \div, \ , \ \}(f_i, f_j) \tag{2}$$

The features with one, two, and three sub-windows are extracted respectively as equation (3), (4) and (5).

$$F_1 = \{(OP)(f_i, f_j), i \quad 1,2...N; j \quad 1,2,...N\} \tag{3}$$

$$F_2 = \sum_{i=1}^{N}\sum_{j\ 1}^{N} \frac{\left|(OP)(f_i, f_j)^{W_1} - (OP)(f_i, f_j)^{W_2}\right|}{w \times h} \tag{4}$$

$$F_3 = \sum_{i=1}^{N}\sum_{j\ 1}^{N} \frac{\left|(OP)(f_i, f_j)^{W_2} \times 2 - (OP)(f_i, f_j)^{W_1} \quad (OP)(f_i, f_j)^{W_3}\right|}{w \times h} \tag{5}$$

Equation (3) enumerates the relation matrix between different orientations in same feature sub-window. Equation (4) investigates the symmetric relation between two sub-windows; Equation (5) enumerates features among three sub-windows with operator defined in equation (1) and (2). These features extraction in equation (3) to (5) will create a feature pool with large number of features.

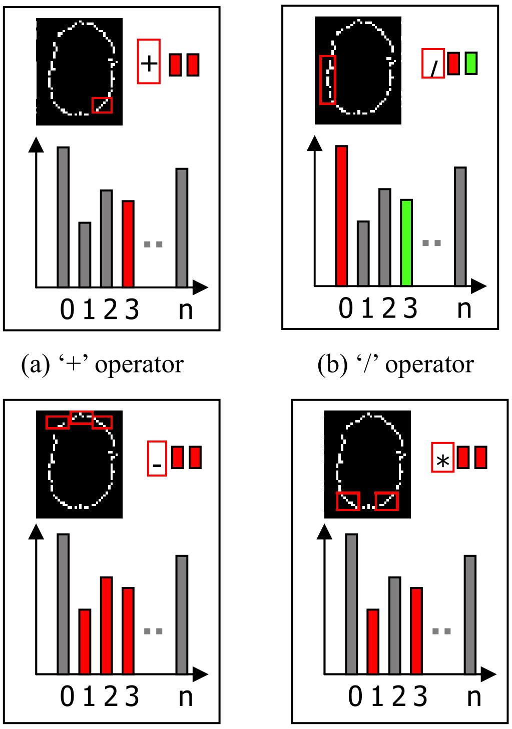
3.2 Features selection

In order to select features and create weak classifiers from the large features pool, real AdaBoost algorithm is employed [6]. The algorithm is briefly reviewed as following.

Four operator types selected in the features boosting are shown in Fig.4. Histogram indicates the statistics of edge magnitude in each orientation bin of feature sub-window. The feature windows are overlapped on a head sample contour image for easy understanding. Fig.4 (a) ~ (d) illustrate ‘+’ operator of orientation bin 3, ‘/’ operator of between orientation bin 0 and 3, ‘-’ operator among orientation bins 1, 2 and 3, '*' operator between orientation bin 1 and '3'.

3.3 Head detection

Referring to Fig.2, input image for head detector is firstly transformed into gradient space. Phase quantization in section 2.2 will work to calculate the orientation image. Head detector achieved in the section above will scan the image, evaluating each possible candidate with boosted strong classifiers. The postprocessor will merge and delete overlapped candidates in a simple way.



(a) '+' operator

(b) '/' operator

(c) '-' operator

(d) '*' operator

Fig.4 Four operator types selected

4 Experiment and Analysis

We use the CMU MOBO dataset as the training and testing samples [7]. CMU 3D room is configured to capture multi-view motion sequences of human body motion. The subjects walk on a treadmill positioned in the middle of the room. A total of 6 high quality (3CCD, progressive scan) synchronized cameras are used. The resulting color images have a resolution of 640x480. Each subject is recorded performing four different types of walking: slow walk, fast walk, incline walk and slow walk holding a ball. It composes of the images of 25 persons. There are more than 300 images captured under each pose of each person. In the experiment, more than 2000 multi-view human shapes are labeled manually as the training data. 3000 images sampled from the dataset act as the testing data. After training the head model, the detector scans each image in the pyramid way for the candidates. As shown in Fig.5 and Fig.6 are the samples of correct positive detection results and false positive detection results (false alarms). In our experiment, six stages of classifiers are used with 242 features in total. The numbers of weak classifiers in each stage are 17, 29, 30, 38, 58, and 70. Referring to the results of false alarms, current detector can't reject some of the clutter image patches, especially when they have similar edge contours to the head model. So, the following work is to refine the head model and utilize the information of other human components to improve the detector's discriminating ability. Fig.8 illustrates the performance of our head detector with ROC curve. Comparing to the best performance of face detection technology reported, our detector has 10 times false alarms. But our detector can work smoothly on multi-view head images without performance decreasing. Naturally, there are many objects with similar shape to the human head, but they maybe have no similar shape at its neighboring area to the shape of human body, such as torso.

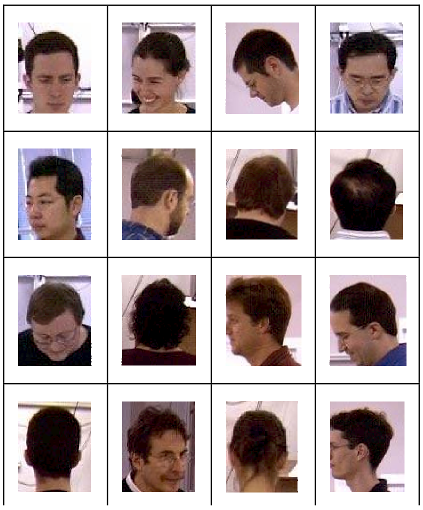


Fig.5 Correct positive results

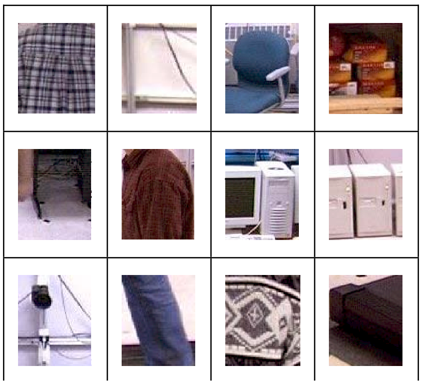
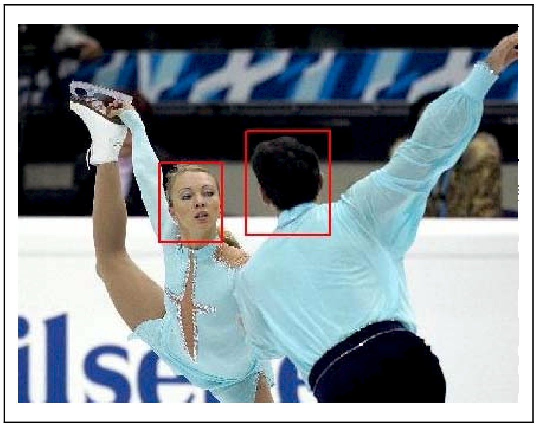


Fig.6 False positive results



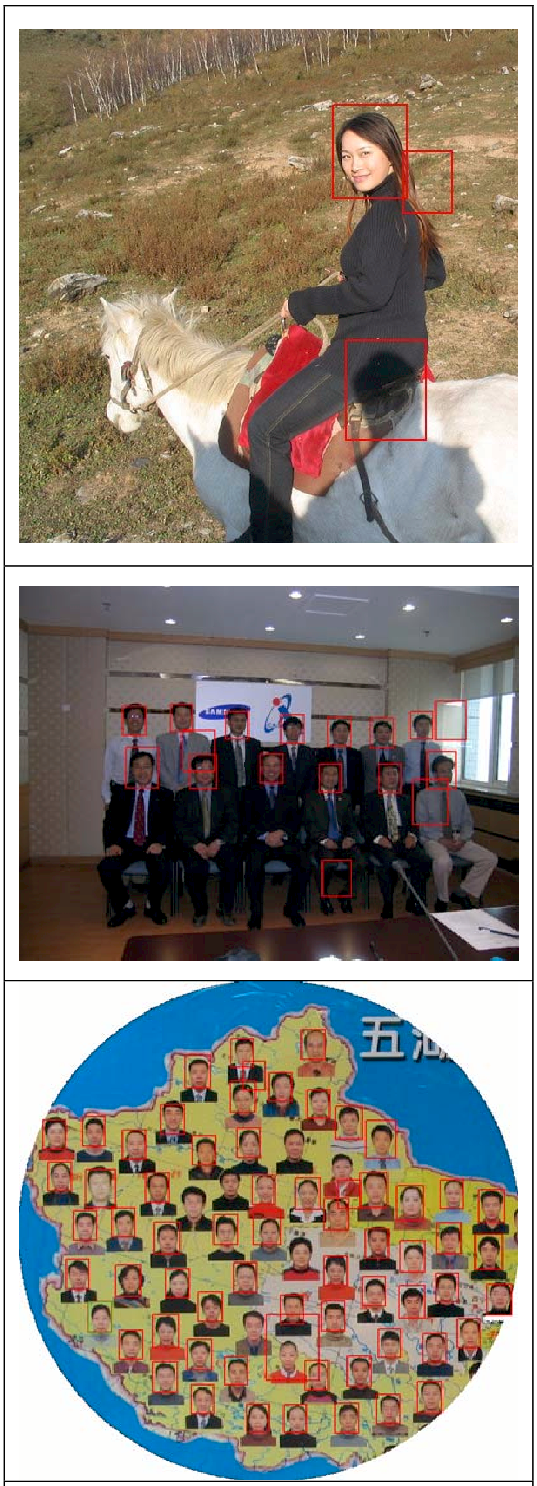


Fig. 7 Detection results on photos

5 Conclusion

In this paper, we present our method to detect multi-view human head in static images. Our method is not limited to work on static images. In crowed situations with fixed background, our detector can work to segment foreground regions provided by background subtraction and count the human number. Many researchers think that the edge information is not reliable and would rather use the color and motion information. But in our experiments, it DOES work well no matter on the detection accuracy and efficiency. If these extra appearance cues are employed in our experiments, we can further improve the detector's performance. Compare to existing head detection technology [8], our method is robust to illumination changes and pose changes out of plane without performance decreasing. In the video surveillance and media management applications, humans are not limited to expose their faces to the camera. Unlike the biometrics identification, user must cooperate with the machine. So, multi-view face detection has its limitation especially in these kinds of applications. Another important point, human head detection can help to segment the foreground region with group of people into individuals on the background subtracted image. To our knowledge, our head detector is the first multi-view human head detector in the world using only the gradient information.

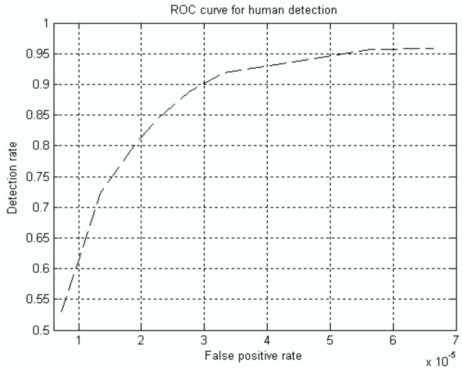


Fig. 8 Performance evaluation with ROC curve

References

[1] T. Zhao, Nevatia R. Tracking multiple humans in crowded environment. IEEE CVPR, pages 406-613, July, 2004

[2] S. Ioffe, D. A. Forsyth. Probabilistic methods for finding people. IJCV 43(1), 2001

[3] Tomaso Poggio, Oren Michael etc. Trainable system to search for objects in images. US Patent 6421463, July 16, 2002

[4] Paul Viola, Michael J. Jones, Daniel Snow. Detecting pedestrians using patterns of motion and appearances. ICCV, pages 734 – 741, Oct. 2003

[5] A. K. Jain, Aditya Vailaya. Shape-based retrieval: a case study with trademark image databases. Pattern Recognition, 31(9), pages 1369—1390, 1998

[6] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions", Machine Learning, 37, pages 297-336, 1999

[7] R. Gross, J. Shi. The CMU Motion of Body (MoBo) Database. Tech. report CMU-RI-TR-01-18 Robotics Institute, Carnegie Mellon University, June, 2001

[8] S. Li, Z. Zhang. Floatboost learning and statistical face detection. vol.26, no. 9, TPAMI 2004.