14-11

MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Estimation of Human Body Orientation using Histogram of Oriented Gradients

Kittipanya-ngam Panachit
Institute for Infocomm Research
1 Fusionpolis Way
Singapore, 138632
pngam@i2r.a-star.edu.sg

Ong Soh Guat
Institute for Infocomm Research
1 Fusionpolis Way
Singapore, 138632
sgong@i2r.a-star.edu.sg

Eng How-Lung
Institute for Infocomm Research
1 Fusionpolis Way
Singapore, 138632
hleng@i2r.a-star.edu.sg

Abstract

Estimation of human orientation is considered a key configuration to study and understand human behaviours. There have been many works focusing on this. However, most of previous works are sensor-based technologies but visual-based technologies. In this paper, we proposed a method to estimate the orientation of human in a given image patch. We have observed the performance of a feature called Histogram of Oriented Gradients (HOG) on human detection and found the sensitivity of this feature in in-plane rotation. It is so sensitive that it could be used in classifying images of rotated and non-rotated human body. Therefore we have drawn an idea to exploit a modified HOG descriptor with 'bin shifting' technique to estimate the orientation of human. The results show that the proposed method can effectively estimate the orientation of human pose inside an image with fast and simple additional operations required.

1 Introduction

Analysis of human behaviour from image sequences has attracted much attention for decades because it is very useful and helpful to have a system automatically observing human actions and behaviours via CCTV in many places. For example, detecting theft actions in shop store, alarms for intruders at home, detecting falls in hospital and alerting security guards for harmful actions in train station. Two main keys of this technology is to track configuration of human body along the image sequence and to analyse them through knowledge provided by experts for each applications. While the configurations of interest depends on applications concerned, position and orientation are basic configurations necessary for further analysis. Although there have been many proposed techniques on human detection and pose estimation to extract location and postures, it is noticed that there have been very few concerning orientation. Most of them assume 'upright' position of target and the variation of poses estimated is assumed perpendicular to the ground. However, in the real world there are many other postures than 'upright' based positions such as falling, bending, crawling and lying down. Then it is clearly that estimation of orientation is non trivial.

Estimation of human orientation is considered a key to study and understand human behaviours and there have been many works focusing on doing so. However, most of previous works are sensor-based technologies[1] but visual-based. Based on our knowledges, there are not many visual-based techniques for estimating orientation of human. Here we categorise techniques of orientation estimation in computer vision into two groups, feature-based and silhouette-based. While the feature-based techniques focus on estimating orientation degree from given image features, silhouette-based techniques estimate from foreground pixels obtained from background subtraction process[2, 3]. Iwasawa *et al.,* [2] track the change of the principal axis of the foreground area obtained from background subtraction in Thermal images. Lee *et al.,* [3] estimate ellipse parameters covering foreground image which contains orientation information. Bay *et al.,*[4] proposed to estimate the main orientation of the target from wavelet features. Chen *et al.,* [5] to estimate the orientation from Weber Local Descriptor.

Here we observed the performance of a feature called Histogram of Oriented Gradients (HOG) [6] on human detection and found the sensitivity of this feature in in-plane rotation. It is so sensitive that it could classify rotated and non-rotated human body. Therefore this paper suggests an idea to exploit a modified HOG descriptor to estimate the orientation of human in the given patch assumed there is a human inside. HOG is briefly described in section 2 and the proposed method is explained in section 3. The datasets used in this paper are introduced in section 4. Next, The study of Rotation Sensitivity and the evaluation of the proposed method are illustrated in section 5. Finally, the performance will be discussed and paper will be concluded in the section 6

2 Histogram of Oriented Gradients

In 2005 [6], Navneet Dalal and Bill Triggs proposed a descriptor representing local object appearance and shape in an image, called Histogram of Oriented Gradient (HOG). The HOG descriptor is described by the distribution of edge directions in the histogram bins. The common implementation begins by dividing the detection window into small a square pixels area, called cells, and for each cell estimating a histogram of gradient directions for those pixels within the cell. The final descriptor is the combination of all histograms in the detection window. In [6], Dalal and Triggs suggest to use 64x128 pixel detection window and 8x8 pixel cell. For detecting human in an image, Support Vector Machine (SVM) is introduced to handle the task of human/non-human classification in each detection window by training SVM with human and non-human images. In this article, HOG is slightly modified to deal with rotation. Mainly there are two major adjustments.

First, because our main objective is to estimate the orientation of the human body, human in the images have to appear in any orientations and poses. For this reason, all images used in this paper are resized to 128 by 128 pixels instead of 64 by 128 pixels suggested by Dalal and Triggs[6].

Second, our descriptor is not based on division of cells. The histogram of gradient directions is calculated from gradients of 128 x 128 pixels of the image patch. The number of bins used in this article is 18 bins. The bin size is 20°

3 Proposed Method on Orientation Estimation

3.1 Motivation

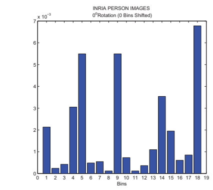
It is mentioned in [6] that HOG descriptor is limited to a certain range of geometrical variation not bigger than the bin size. HOG descriptor has been tested by detecting human on different orientations of human body to study the rotation invariance of HOG. We observed that HOG-based human detector can perform well if image is not rotated beyond the bin size of HOG, which is $\pm 20^\circ$ (See section 5.1). This means if the image patch with human inside is rotated beyond the bin size, it is likely to be classified a negative(non-human) image. This invariance characteristic inspired an idea to use it to estimate the orientation of human inside the image patch. For example, given an image patch with human inside with unknow orientation, we can estimate the orientation of human inside by rotating the patch around so that the detector returns positive (human detected). The positive result can be implied that the human has been rotated so that it is aligned with the main direction of trained data within the range of bin size away, $\pm 20^\circ$. However to do so with every image patch and every angle, it is too computational expensive and slow. Here we proposed a method called 'Bin Shifiting' to replace the rotating process.

3.2 Bin Shifitng

Knowing that Histogram of Oriented Gradients of Dalal and Triggs is sensitive to rotation, we exploit this property to estimate the angle of rotation of the human body. The fundamental thought is to use different angle of rotated test images and apply HOG-based human detector to obtain the confidence score of human detected. However in order to do this, we have to prepare images of different rotating angle of the given image and extract HOG of each image before passing to the classifier. The preparation of these training datasets is time consuming and tedious. Thus in order to overcome this problem, we approach the situation in another method called Bin Shifting to replace the rotation process.

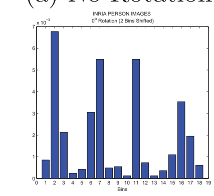
According to Dalal and Triggs, gradients are extracted and classified into different orientation bins. When images are rotated at a certain angle, we assume that all edges are shifted according to the angle of rotation. For instance, we have 18 orientation bins spaced evenly between -180° and 180°. An edge of 40° and 170° will be classified into the 12th and 18th bins respectively. So when this image is rotated by 40°, the mentioned edges will then be classified into the 14th and 2nd bins respectively. So when the image is rotated by 40°, it is equivalent to shift 2 bins of orientation bin. From Figure 1(b), we can clearly see that the bins are shifted 2 bins to the right as compared to Figure 1(a). Moreover, the distribution shape of Figure 1(b) is similar to Figure 1(c) which is the HOG decreiptor of actually-rotated image in Figure 1(a). Thus this shows that by shifting the bins of orientation is equivalent to rotating the image manually.

By using this idea, instead of rotating the target patch before extracting HOG, the bin shifting is applied just right after feature extraction and before classification.

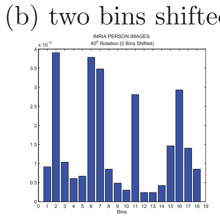


(a) No Rotation

(d) A sample image



(b) two bins shifted



(c)40° manually rotated

(e) Rotated image

Figure 1. Histograms of the oriented gradients of a sample

4 Datasets

There are two different sets of data used in this article: INRIA Person[6] and CVIU. INRIA Person database contains two groups of data, the Train Set and the Test Set. From the Train Set, we choose 664 images of positive training images; this number includes their left-right reflections. In addition, 700 images are used as negative training images. For the testing images of INRIA Person, we have prepared nine sets of images. Each set of images contain 208 positive images. However, these images are rotated according to a specific angle, which is every 40° from -160° to 160° for each test set correspondingly. Samples of rotated images are illustrated in figure 2

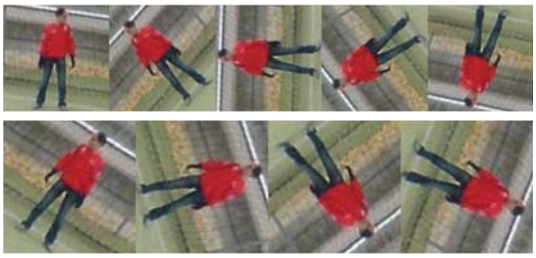


Figure 2. Sample images from INRIA Person Dataset with rotated images

CVIU is a dataset containing a video sequence of different orientations and poses, shown in figure 3.

Figure 3. Sample images from CVIU sequence

5 Experimental Results

5.1 Observation of Rotation Sensitivity

INRIA Person database is used in this section. Linear SVM Classifier was trained by modified HOG descriptors mentioned in section 2 of non-rotated positive images and non-rotated negative images from training set. Then it is used to classify the test data of positive images on each rotated angle, 0°, 20°, 40°, 60°, 80°, 100°, 120°, 140°, 160°, 180°, -20°, -40°, -60°, -80°, $-100\circ$, -120° and -160°). The number of images correctly classified on each rotated angle is then recorded and shown in table 1 and can be visualised on figure 4. The results show that within the orientation range of bin size, 20°, the performance of human detector is significantly better than others as mentioned above. It is also noticed that the performance of the angle in the range of bin size from opposite direction of the main orientation is surprisingly as good as those of the main direction. With this characteristic of HOG, we are proposing to estimate the alignment of human.

In figure 4 and table 1, it is noticed that the classification accuracies drop as the angle of rotation is further away from the 0° orientation. In other word, the HOG-based classifier will classify a given image patch positive only if the human in the target patch is well aligned with the main orientation of training images. Knowing this characteristic, we propose to apply HOG-based classifier to classify the rotated patch of target image, known surely there is a human inside. With this characteristic of rotation invariance, we expect that only the rotated patch of the right angle will be classified positive by the classifier and possess the highest score from SVM classifier. To do so, we here use a method so called *Bin Shifting* instead of actual rotating image before classifying.

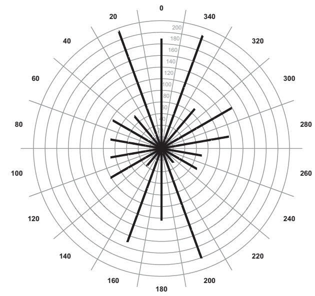


Figure 4. The number of images correctly classified over different orientation of human body

Table 1. The number of images correctly classified over different orientation of human body

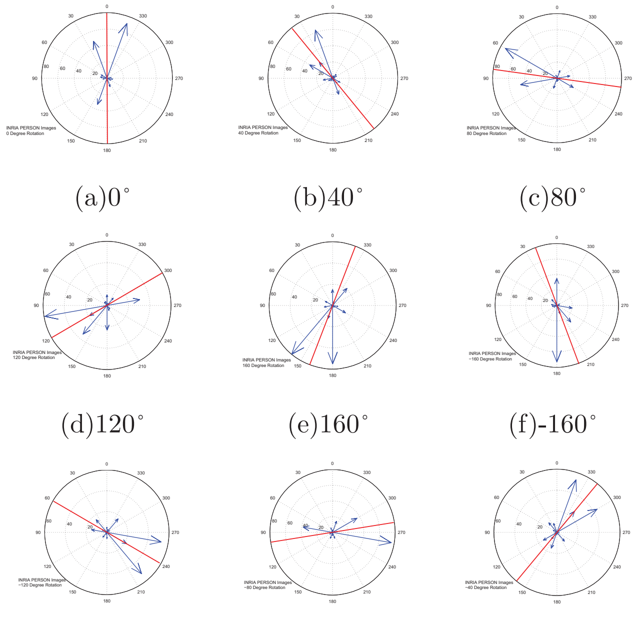
Rotation	Accuracy	Rotation	Accuracy
0°	**82.69%**	180°	54.33%
20°	**93.75%**	-20°	**90.38%**
40°	31.73%	-40°	38.94%
60°	42.31%	-60°	38.94%
80°	38.94%	-80°	51.44%
100°	38.94%	-100°	30.77%
120°	44.23%	-120°	30.77%%
140°	16.83%	-140°	13.94%
160°	**75%**	-160°	**87.98%**

5.2 Performance of Proposed Method

5.2.1 Performance on INRIA Person Database

Here we do the usual feature extraction by using the modified HOG suggested in section 2 on both the training set and testing data set of INRIA Person database mentioned in section 4. For our work, a total of 18 bins are used and they are spaced evenly over $-180\circ$ to $180\circ$. Hence we have 9 test sets of different orientations, each set contain the same images as others but rotated by different angle. Then we used linear SVM of LIBSVM presented by Chuh- Chung Chang and Chuh-Jen Lin[7] to learn the modified HOG descriptors of the training set and then to classify on test sets with 'bin shifting' method. Each test image will be bin-shifted through all 18 possible orientation. In this way, a test image will have 18 different classification scores given by SVM. We will then classify the test image into the angle of rotation base on the highest classification score. After estimation on a set of each orientation, the number of images assigned to each direction is recorded and displayed by the circular histogram. Hence, a set of each orientation contains 208 positive images.

Figure 5 shows the circular histograms of different rotation test images. The red line shows the actual orientation of the images. The blue line indicates the number of images classify under different angles of rotation. From the circular histograms, we see most of the images are classified into the correct angle of rotation with $\pm 20^\circ$ different from the actual angle of rotation, regardless of the direction of the human head. To illustrate the point, figure 5(a) shows the histogram of 0° rotation of the test images. Most of the images are classified as 20° rotated images. For figure 5(b) shows the histogram of 40° rotated test images. Most of the images are classified as either 20°, 40° or 60°.



(a)0° (b)40° (c)80°

(d)120° (e)160° (f)-160°

(g)-120° (h)-80° (i)-40°

Figure 5. the circular histograms of different rotation test images from INRIA Person database

5.2.2 Performance on CVIU Database

In this part, the same human detector with the same training set and 'bin-shifting' method as in section 5.2.1 is applied to estimate the orientation of human body in the CVIU video sequence. Each test image will be bin-shifted through all 18 possible orientation. In this way, each frame will have 18 different classification scores given by SVM. We will then classify that frame into the angle of rotation base on the highest classification score.

Figure 6. Correctly estimated CVIU test images.

Figure 7. Incorrectly estimated CVIU test images. CVIU test images.

Figure 6 and figure 7 show some of the CVIU images which are correctly and incorrectly categorized, respectively. The red lines represent the estimated angle of rotation. For CVIU test images, most of the upright images are categorized to the correct angle of rotation. However, there are quite a number of images that are misclassified. Most of the misclassified images are images with human in the upright with the hand outspread. The outspread hands could be one of the problems that cause the method to fail to estimate the orientation of the human body.

6 Discussions and conclusions

The performance of a feature called Histogram of Oriented Gradients (HOG) on human detection was studied and observed the sensitivity of this feature in in-plane rotation. With this feature, the score of confidence from linear SVM drops dramatically when the image is rotated away, beyond the bin size ±20○, from the direction of training data. This characteristic is used to classify rotated and non-rotated human body.

The proposed method is to rotate the target image with 'bin shifting' technique and then apply modified HOG-based human detector to calculate the score of confidence. Then the orientation of human is estimated based the highest score. This approach was tested on two datasets, INRIA Person Images and CVIU images.

'Bin shifting' technique is a simple and fast additional method but equivalent to actual rotating process.

The performance shows that this technique can effectively estimate the orientation of human body within the range ±20○, a bin size of HOG.

It is also noticed that such strong edges in other direction as those of hand outspread could cause wrong estimations. Probably, HOG-based human detector counts more on the local distribution of strong edges than on the global distribution of edges.

References

[1] Lee, J.K., Park, E.: Minimum-order kalman filter with vector selector for accurate estimation of human body orientation. Robotics, IEEE Transactions on **25** (2009) 1196 –1201

[2] Iwasawa, S., Ebihara, K., Ohya, J., Morishima, S.: Real-time human posture estimation using monocular thermal images. In: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. (1998) 492 –497

[3] Lee, M.W., Nevatia, R.: Body part detection for human pose estimation and tracking. In: Motion and Video Computing, 2007. WMVC '07. IEEE Workshop on. (2007) 23 –23

[4] Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: In ECCV. (2006) 404–417

[5] Chen, J., Shan, S., Zhao, G., Chen, X., Gao, W., Pietikainen, M.: A robust descriptor based on weber's law. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. (2008) 1 –7

[6] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1. (2005) 886 –893 vol. 1

[7] Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001)