GEOMETRIC MODEL FOR HUMAN BODY ORIENTATION CLASSIFICATION

Igi Ardiyanto
Department of Electrical Engineering and Information Technology,
Faculty of Engineering, Gadjah Mada University
Yogyakarta 55281, Indonesia
Email: igi@ugm.ac.id

Abstract—**This paper proposes an approach for calculating and estimating human body orientation using geometric model. A novel framework integrating gradient shape and texture model of the human body orientation is proposed. The gradient is a natural way for describing the human shapes, while the texture explains the body characteristic. The framework is then combined with the random forest classifier to obtain a robust class difference of the human body orientation. Experiments and comparison results are provided to show the advantages of our system over state-of-the-art. For both modeled and un-modeled gradient-texture features with random forest classifier, they achieve the highest accuracy on separating each human orientation class, respectively 56.9% and 67.3% for TUD-Stadtmitte dataset.**

Keywords: **Human Body Orientation; Histogram of Oriented Gradient; Local Binary Pattern; Geometric Model**

I. INTRODUCTION

Knowing human body orientation is useful for many applications. It may be used in a monitoring application and surveillance. It can also tell us about people interactions in the surveillance scenes. For example, we may predict that a group of persons facing each other for a long time are having conversation, or a group of persons facing to the road can be inferred as waiting for the bus. Yet, the human body orientation classification is a very difficult and challenging task.

There are several researches attempting to tackle the problem of estimating the human body orientation. Reference [1] utilized a HOG-based detector and SVM for solving the human body orientation problem. In a recent work, Ref. [2] employed HOG features and SVM Tree for solving the same problem. Yet, both researches above only consider shape features of the human body. Reference [3] made a notable proposal for classifying the human orientation, by using Pyramid-HOG features and sparse, combined with a soft-coupling technique between the whole body orientation and its velocity. Nevertheless, it is exclusively used for the surveillance system without any further information about how to use such method real-time and still focuses on the shape features.

Not only useful for the surveillance system, the human body orientation can also help the robot for obtaining a better prediction to avoid a moving person in a navigation task. It may assist the robot to build a social interaction with the human, such as approaching a person and asking the way in an outdoor setting. Such application surely needs the robot to have a good estimation of the person orientation for facing him/her.

We propose a system for detecting and classifying the human upper body orientation, as has been mentioned above. Here we exploit the upper body part of the human, in contrast to the whole body, for achieving better robustness under occlusion cases. The whole body detection is usually affected by the low and small things such as chair, table, bicycle, and so on.

Our main contribution resides in the use of the model-based gradient and texture features for estimating the human upper body orientation. A framework integrating both geometrical features above is also provided to improve the accuracy of predicting the human body orientation. Here we significantly simplify the methods described in our previous work [4].

The paper is organized as follows. First, the detection and estimation of human upper body orientation are described using models in the Method section. The comparison of several methods and the result of experiments are then provided in Result and Discussion section. Lastly, the work is concluded and some possible future works are discussed.

II. METHODS

Our proposed system is hierarchically built by first detecting and creating bounding boxes around the human upper body using a body detector. These detection results are then given to the orientation classifier

Received: February 24, 2015; received in revised form: March 30, 2015; accepted: April 13, 2015; available online: June 17, 2015.

part. Figure 1 explains the proposed framework for estimating the human body orientation.

A. Dataset

The body dataset (more specifically, human upper body) is created by cropping the INRIA [1] and Fudan-Penn [5] datasets into 48×64 pixels containing the upper-half body of persons. CALVIN upper body dataset [6] is also added into the dataset, so that 4250 positive samples of the human upper body are obtained. Subsequently, 3000 positive samples are used for training the upper body detector and the rest is for testing purpose. Two thousands and five hundred negative samples are then created from images which do not contain the human upper body, including the bottom part of the human body. For the orientation classification purpose, the training samples are separated into eight classes representing the eight orientation of the human body (see Fig. 2). The same treatment is also applied to the testing samples. Besides our testing set above, TUD Stadtmitte dataset was also used by Ref. [1] and it will be explained later in the experiment section.

B. Human Upper Body Detection

Histogram of Oriented Gradients (HOG) [7] is one of state-of-the-art descriptor for the person body detection. Since it is considerably slow for real-time applications, here the extended work of HOG by Ref. [8] is employed, which utilized Ada boost for selecting features and cascade rejection for speeding up the detection time. The bounding boxes of the human upper body (the detection results) are then fed as the input for the orientation estimation.

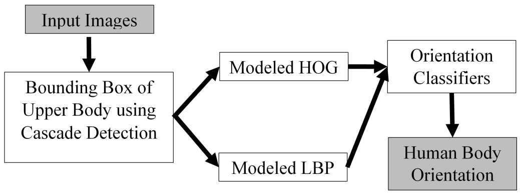

Fig. 1. The Diagram of the Human Body Orientation Classification System.

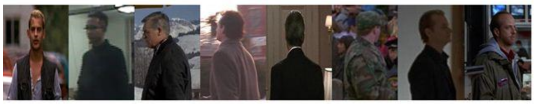

Fig. 2. The Eight Classes of the Human Upper Body Orientation Representing (from Left to Right) Front, Front-Left, Left, Back-Left, Back, Back-Right, Right, and Front-Right Directions.

C. Extracting Features of Human Body Orientation

The proper choice and model of the features will convincingly give a better impact to the orientation classification results, unlike the other works which are only based on the gradient features (e.g. as in Ref. [2] and [3]). Here, an integration of the gradient-based and texture-based features are proposed using geometric models which amplify the necessary cues of the human body orientation.

D. Shape Cue

For extracting the human upper body shape, HOG descriptor is used. Considering an image I, the HOG descriptor is obtained by computing the first derivative of the image with respect to x- and y-axis of the image. A convolution operation is then performed using 1-D mask [1 0 1], producing gradient magnitude of the image. The orientation of the gradient is then calculated by

$$\theta = \tan^{-1}\left(\frac{I_y}{I_x}\right) \tag{1}$$

where I_x and I_y respectively denote the gradient magnitude of each x- and y-axis.

Every sample is then divided into 6×8 blocks, and each block consists of four cells. The gradient orientation is subsequently quantized into nine bins, so now 1728 dimensional feature vectors of HOG descriptor are obtained.

E. Texture Cue

Local Binary Pattern (LBP) is used, adopting the work of Ref. [9], to build a texture descriptor. Image textures are calculated using $LBP_{8,1}$ operator for each pixel

$$I_{LBP_c} = \sum_{p=0}^{7} 2^p f(I_p - I_c), \tag{2}$$

where I_c is the center pixel from which the LBP value I_{LBP_c} and p is eight-surrounding pixels of I_c are calculated. The LBP image is then divided into 6×8 blocks, similar to the HOG features above. For each block, a histogram containing 59 labels based on uniform patterns is built. According to Ref. [9], the uniform patterns contain at most two bit transitions from 0 to 1 and vice versa. For an 8-bit data, there are 58 uniform patterns and the other patterns which have more than two bit transitions are grouped into one label, so the total is 59 labels. This procedure provides 2832 dimensional feature vectors after all histograms are concatenated.

F. Merging Features Geometrically using Models

The key of orientation estimator lies in the utilization of geometric models to illustrate the human upper body orientation for different features. For example, HOG features are very good for describing the global shape of the body, but they are poor for representing small details like the face. On the other hand, LBP is a good descriptor for catching the texture like the face outline, eyes, and nose. Another consideration is to suppress the background from the upper body shape which is caught by HOG descriptor and clothing textures captured by LBP descriptor which may vary from one person to others, to increase the classification results. For those purposes, a model for each HOG and LBP descriptor are made as follows:

1) For shape cue, the important features are edges separating the foreground (human body itself) and the background. The HOG features around those edges are then emphasized.
2) For texture cue, the face textures can be a basis for distinguishing the human orientation. One example is the fact that it is assumed one person is facing backward when there is no face textures in his/her head, or the person is facing left or right when only a half part of his/her face is examined, the LBP features around the head area is then emphasized.

For creating the models, 15 positive samples are chosen randomly from each orientation class. Each image is then divided into 6×8 blocks, the same with the features extraction above. Let $i = \{1, 2, \dots, M\}$ be the index of blocks in one image sample and M is the total number of blocks (48 blocks). Let $j = \{1, 2, \dots, N\}$ denotes the index of image samples with the total $N = 15 \times 8$ classes. We also define $b_{i,j}^{HOG}$ and $b_{i,j}^{LBP}$ as the block i at image j for HOG and LBP features respectively.

Each sample is then manually annotated for two models as follows

1) For HOG features, the block which contains the edge shapes of the human upper body is weighted as

$$b_{i,j}^{HOG} = \begin{cases} 1, & \text{if contains edges} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The yellow blocks in the middle images show the important cues for each feature. The right images show the final model of each feature (brighter means higher value).

2) For LBP features, the block which contains the head is weighted as

$$b_{i,j}^{LBP} = \begin{cases} 1 & \text{if contains the head} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

All samples are then averaged to get one model for each feature.

$$\bar{b}_i^{HOG} = \frac{\sum_{j=1}^{N} b_{i,j}^{HOG}}{N} \tag{5}$$

$$\bar{b}_i^{LBP} = \frac{\sum_{j=1}^{N} b_{i,j}^{LBP}}{N} \tag{6}$$

Figure 3 explains the procedure of creating models.

After the model is obtained, each feature is then weighted using its respective model. Let $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_M]^T$ denotes the HOG features, and $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_M]^T$ denotes the LBP features, with $\mathbf{F}_i$ and $\mathbf{G}_i$ are set of respective features at block i. The weighted features are then acquired by following equation:

$$\mathbf{F}'_i = \gamma^{\bar{b}_i^{HOG}} \mathbf{F}_i \tag{7}$$

$$\mathbf{G}'_i = \gamma^{\bar{b}_i^{LBP}} \mathbf{G}_i \tag{8}$$

where γ is a constant (currently we use $\gamma = 2$). Finally, the concatenated and weighted features $\mathbf{F}'$ and $\mathbf{G}'$ are fed to the classifier for training.

G. Random Forest Classifier

Estimation of the human upper body orientation is certainly a multi-class classification problem. One of the notable classifiers which works well on the

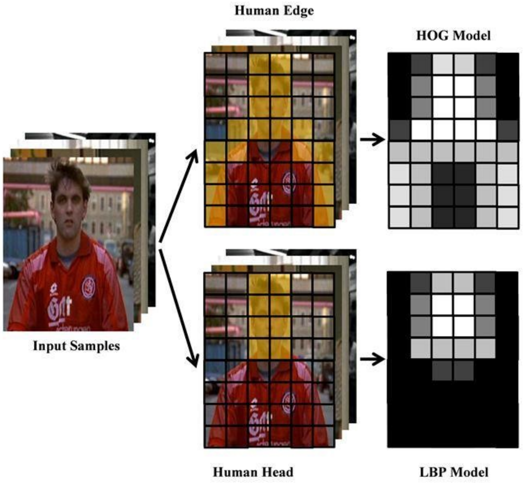

Fig. 3. The Procedure of Creating the Model for Gradient-based And Texture-based Features.

multi-class data is random forest, introduced by Breiman [10]. It is an ensemble learning method which combines the prediction of many decision trees using a majority vote mechanism. Random forest is devoted for its accuracy on the large dataset and multi-class learning. It becomes our reason to choose this algorithm for training our eight-orientation classification problem with a large set of features. The outline of random forest algorithm is as follows:

1) Let K denotes the number of trees to be generated.
2) For each tree k = 1 to K
    - a) Get a bootstrap sample φ_k from the training data
    - b) Grow an unpruned tree on the bootstrap φ_k
    - c) For i = 1 to number of nodes
        - i) Randomly sample m predictors.
        - ii) Choose the best split among those predictors
3) Output the prediction by taking the majority vote from all trees.

The readers are encouraged to refer to the original paper [8] for further explanations.

III. RESULTS AND DISCUSSION

Size of all images and camera sequences used in our experiments is 640 × 480, and all of implementations were done using C++ and a laptop PC (Core2Duo, 2.1 GHz, 2GB memory, Windows 7).

A. Evaluation of Human Orientation Estimation

First, the performance of our human upper body orientation estimator system is evaluated, and compared with the existing works [1–3]. Besides comparing with the existing methods, the comparison with several multi-class classifiers is also done, such as Decision Tree, SVM-Multiclass [11], and MultiBoost [12].

Figure 4 shows the results of our proposed human body orientation classification. From the figure, the human body orientation classification system runs well, even with various persons and poses.

Table I shows the evaluation results. For each method in the table, the training sets mentioned in subsection "Method: Dataset" are used to train the eight-class human upper body orientation. TUD-Stadtmitte dataset [1], which contains multiple persons crossing the street with a complex environment and many occlusions, and the testing sets (from subsection "Method: Dataset") are then used for the evaluation. The TUD-Stadmitte dataset is annotated and the bounding boxes containing the human upper body are given to the evaluation system. For HOG-LBP features in the table, the HOG and LBP features are directly concatenated without any weighting. As it is seen, the proposed models give advantages and make system outperform the other existing methods.

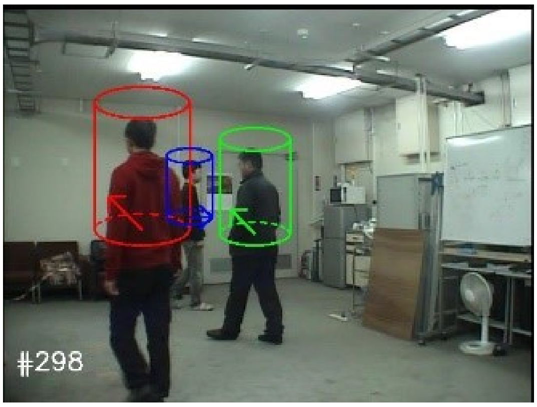

Fig. 4. The Examples of the Human Body Orientation Classification Results.

TABLE I
THE EVALUATION OF HUMAN UPPER BODY ORIENTATION ESTIMATION.

Method	Accuracy (%)	
	Our Set	TUD Set
HOG + SVM [1]	34.3	42.2
HOG + SVM-Tree [2]	42.1	47.1
PyrHOG + Spase-SVM [3]	50.6	55.0
HOG-LBP + Decision Trees	36.8	41.2
HOG-LBP + SVM-Multiclass [11]	43.3	48.0
HOG-LBP + MultiBoost [12]	45.5	52.6
HOG-LBP + Random Forest	52.1	56.9
Modeled HOG-LBP + Random Forest	60.2	67.3

IV. CONCLUSIONS

A framework of human body orientation estimation and classification has been described. Our detection and classification system of the human upper body orientation probably works better than any existing methods. The proposed method utilizes a model-based shape-texture features combined with the random forest. It also gives a possibility to be used in the real robot application such as the person tracking or surveillance system.

Possible future works for our system are to have integration with other sensors such as laser range finders and implementation on a real robot for specific purposes. A more robust system is expected to establish for person tracking and localization by applying multi-sensory fusion.

References

[1] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3d pose estimation and tracking by detection," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 623–630.

[2] C. Weinrich, C. Vollmer, and H.-M. Gross, "Estimation of human upper body orientation for mobile robotics using an svm decision tree on monocular images," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 2147–2152.

[3] C. Chen, A. Heili, and J.-M. Odobez, "Combined estimation of location and body pose in surveillance video," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*. IEEE, 2011, pp. 5–10.

[4] I. Ardiyanto and J. Miura, "Partial least squares-based human upper body orientation estimation with combined detection and tracking," *Image and Vision Computing*, vol. 32, no. 11, pp. 904–915, 2014.

[5] L. Wang, J. Shi, G. Song, and I.-F. Shen, "Object detection combining recognition and segmentation," in *Computer Vision–ACCV 2007*. Springer, 2007, pp. 189–199.

[6] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[8] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 1491–1498.

[9] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[10] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] K. Krammer and Y. Singer, "On the algorithmic implementation of multi-class svms," *Proc. of JMLR*, 2001.

[12] D. Benbouzid, R. Busa-Fekete, N. Casagrande, F.-D. Collin, and B. Kégl, "Multiboost: a multi-purpose boosting package," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 549–553, 2012.