Human Detection Method Combining HOG and Cumulative Sum based Binary Pattern

Jong Gook Ko', Jin Woo Choi', So Hee Park', Jang Hee You',

' Electronics and Telecommunications Research Institute
218 Gajeongno, Yuseong-gu, Daejeon, 305-700, Korea
{jgko@etrise.lcr}

Abstract. In this paper, we introduce a novel method to detect human in image/video. For human detection, we use Histograms of Oriented Gradients (HOG) and Cumulative Sum based Binary Pattern (CSBP) as the feature sets. And Support Vector Machine (SVM) is employed for classification of training data. HOG and CSBP features are combined for better performance of human detection. In this paper, human detection system based on HOG and CSBP with SVM is compared with the current state-of-the-art human detection system: SVM using histograms of oriented gradients (HOG). The results show that HOG+CSBP classifier achieves better result than HOG.

Keywords: Human detection, Histograms of Oriented Gradients, Support Vector Machine.

1 Introduction

Human detection has very important role in video surveillance, automotive safety system, and content-based image/video retrieval and so on. However, human detection in image and videos is a challenging task because of their non-rigid appearance and the wide range of poses.

The complexity of the human class is handled via strong machine learning methods by computing discriminative features inside an image region. The first approach that showed excellent results was the work of Dalal [1] who used local histograms of oriented gradients (HOG) as features and a linear support vector machine classifier. It is notable because it was the first paper to report impressive results on human detection. Their works use HOG as low-level features, which were shown to outperform features such as wavelet[3], PAC-SIFT[4], and shape contexts[5]. To improve detection speed, Zhang et al. [2] propose a rejection cascade using HOG features. Li et al [7] applied the histogram of Oriented Gradients feature to detect the head-shoulder. Wang et al [8] combined the HOG feature with the Local Binary Pattern(LBP) feature to detect the holistic body of human. They reported that the

HOG-LBP feature gains more that 20% improvement for upper body detection compared to the HOG feature.
In this paper, we propose a novel method to detect human in image using HOG and Cumulative Sum based Binary Pattern (CSBP) features. Our experiments show that the HOG-CSBP descriptors are more discriminative than HOG only. This paper is organized as follows. Section 2 describes Cumulative Sum based Binary Pattern features extraction method. In section 3, experimental results are presented, and conclusions are given in section 4.

2 Human Detection using HOG and CSBP feature Descriptors

In this paper, HOG and cumulative sum based binary pattern is used to extract features form image for human detection.

Histograms of Oriented Gradients (HOG).
HOG features have been successfully used in pedestrian classification by Dalal/Triggs[1]. It seems to be one of the best features for capturing edge and shape information, while being sensitive to noisy background edges and clutter. In the original approach, pixel gradients are extracted from a spatial grid of 7 x 15 = 105 overlapping blocks. Each block consists of four 8x8 pixel cells, for which orientation histograms with 9bins are computed. Four cell vectors are concatenated to one block vector and normalized at feature level, All block vectors together form the final feature vector for one instance. We make use of integral images in our approach similar to the one proposed in[8]. An important parameter in the original approach is the spatial pixel weighting, which is difficult in the integral image approach since spatial information on pixel levels is usually lost. In this paper, one block consists of 16 x 16 pixels and the blocks make one histogram. So, 945 dimensions (7 x 15 x 9) of HOG features for a 128x64 window are generated.

Cumulative Sum based Binary Pattern (CSBP)
We propose a cumulative sum based binary pattern features for human detection descriptors. Similar to Local Binary Pattern (LBP), CSBP uses 8 pixels surrounding center pixel. Cumulative sums are calculated as follows. Suppose that X1, X2, ... X5 are five representative values of each around pixels of center pixel.

- Calculate the average AVG(X) = (X1 + X2 + + X5) / 5
- Calculate cumulative sum from 0: $S_o = 0$.
- Calculate the other cumulative sums by adding the difference between the current value and the average to the previous sum:
  S, = S + (X, - AVG(X)) for i = 1,2, ... 5.

After calculation, local pattern are generated for each pixel using the following algorithm. 8 neighbor pixel of center pixel are used for cumulative sum based binary pattern.

```
Local Pattern Generation
  MAX = max(S1, S2, ...     MIN = min((S1, S2, ... S5);
  S5) if (upward slope)
       If( neighbor pixel between MAX and MIN
          set neighbor     pixel between MAX and MIN to 1
          else
              set neighbor pixel between MAX and MIN to 0
  if (downward slope)
        If( neighbor pixel between MAX and MIN
          set neighbor     pixel between MAX and MIN to 0
          else
              set neighbor pixel between MAX and MIN to 1
end.
```

The algorithm generates cumulative based local patterns that present texture. And we only use not 58 uniform binary patterns but 9 uniform binary patterns for CSBP local pattern to reduce vector dimension. 9 binary pattern consist of the patterns which defined by the number of '1'.

3 Experiments

We perform all of the experiments on the INRIA human database[], which is one of the most widely used database for human detection in still images, consisting of thousands of cropped human images in urban scenes. This database contains 2416 human annotations and 1218 non-human images for the training stage, and similar number of samples for testing. Moreover, there are a variety of variations in human pose, clothing, lighting, clutters and occlusions, thus challenging and suitable as a benchmark for comparison between different algorithms and features.

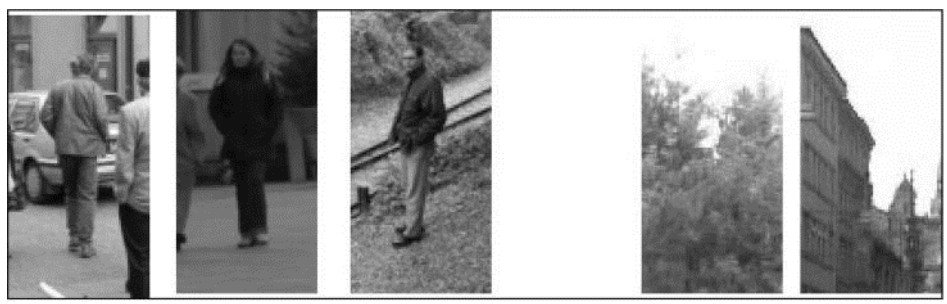

Fig. 1. Some of the positive and negative samples from INRIA DB

We use libSVM to train and classify on the INRIA dataset. And we evaluate the performance of HOG-CSBP human detection. For reducing feature vector dimensions, we make cell by 16 x 16 pixels. From the results shown in Table 1, we conclude that proposed HOG + CSBP method better performance than only HOG method.

Table 1. Performance Test Result

DB	HOG	HOG + CSBP
INRIA test set images (person:1126, nonperson: 1194)	87%	92%

4 Conclusion

In this paper, HOG + CSBP feature extraction method for human detection was proposed. CSBP employs feature extraction using a cumulative sum based change analysis. HOG and CSBP features are combined for better performance of human detection. Experimental results show that the proposed approach has good performance for human detection. Later, we need to test more by changing the cell size.

Acknowledgments. This work was supported by the Industrial Strategic Technology Development Program, 10039149, Development of Basic Technology of Human Identification and Retrieval at a Distance for Active Video Surveillance Service with Real-time Awareness of Safety Threats funded by the Ministry of Knowledge Economy (MKE, Korea)

References

1. N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. CVPR, pp. 886--893. (2005)
2. Zhu, Q., Avidan, S. Fast human detection using a cascade of histograms of oriented gradients. In: CVPR, pp. 1491-1498 (2006)
3. A. Mohan, C. Papageorgiou, T. Poggio, Example-based object detection in image by components, IEEE conference on PAMI, pp. 349-361 (2001)
4. Y. ke, R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors, IEEE conference PAMI, pp. 506-513, (2004)
5. A. Mykhaylo, R. Stefan roth, Pictoral structures revisited: People detection and articulated pose estimation, IEEE Conference on CVPR, (2009)
6. Q. Zhu, M.C, S. Avidan, Fast human detection using a cascade of histograms of orientated gradients, IEEE Conference on CVPR, pp. 1491-1498, (2006)
7. M. Li, Z. Zhang, T.N. Tan, Estimating the number of people in crowed scenes by MID based foreground segmentation and head-shoulder detection, ICPR, (2006)
8. X.Y. Wang, X. Han and S.C. Yan. An HOG-LBP human detector with partial occlusion handling . ICCV, (2009)