Human Detection Algorithm Based on Discriminative Local Feature

申請者
Jiu XU

情報生産システム工学専攻
マルチメディアシステム研究

2014年 6月
Human detection is one of the most active research areas in computer vision and is quite useful for many applications, such as smart vehicles and video surveillance systems. It is also an important preliminary step for content analysis, such as for behavior recognition and human tracking. Due to the variations in human body poses and clothing together with as well as those in cluttered backgrounds and environmental conditions, human detection remains challenging work. In particular, because of its huge computational workload, real-time human detection now seems the most critical step in intelligent video surveillance systems. To overcome these problems, a robust method with a real-time technique is required.

A number of feature extraction algorithms have been proposed so far. For example, several gradient-based methods including the histogram of orientated gradient (HOG) (N. Dalal, CVPR 2005) and the covariance matrix (COV) (T. Oncel, PAMI 2008), have achieved excellent performance using gradient information. Certain texture-based methods have shown promise as well. Semantic local binary patterns (SLBP) (Y. Mu, CVPR 2008) and Center-symmetric local binary/trinary patterns (CS-LBP/LTP) (Y. Zheng, ACCV 2010) were proposed as effective applications of conventional Local Binary Patterns (LBP) (T. Ojala, Pattern Recognition 1996) to the problem of human detection. Some other combinational methods such as histogram of template (HOT) (S. Tang, ICASSP 2010) and gradient local binary patterns (GLBP) (N. Jiang, ISCAS 2013) have also contributed great results to the human detection. However, all these existing methods have some limitations such as not considering the gradient and texture information entirely or being sensitive to the relative changes between the background and foreground.

In order to solve these problems, in this dissertation, several proposals of more powerful and efficient method are given. Moreover, in order to make our proposals suitable for real-time applications, acceleration approaches are also proposed to reduce the running complexity.

The rest of this dissertation is organized as follows:

In chapter 1 [Introduction], the background knowledge of object detection and human detection is introduced. An overview of this dissertation is also presented in this chapter.

In chapter 2 [Bidirectional Local Template Patterns Method], bidirectional local template patterns (B-LTP) method is proposed as a combination and modification of HOT method and CS-LBP method using both texture and gradient information. For each pixel, B-LTP method defines four templates, each of which contains the pixel itself and two neighboring center-symmetric pixels. For each template, it then calculates texture and gradient information from the relationships among these three pixels and from the two directional transitions across these pixels.
based on three novel formulas. These calculations take the gradient direction and the intensity variance information into account, make it possible to improve detection accuracy with limited pixel numbers.

In contrast with other related combinational methods, B-LTP method makes full use of the most important four templates rather than eight templates in HOT method. Furthermore, B-LTP method not only the calculates intensity and gradient information relationships between the three pixels as is also considered in HOT method, but also the information of two directions across these pixels are considered. This enables B-LTP method exceed HOT method in detection performance. On the other hand, B-LTP method calculates the intensity and gradient variations along two directions based on four predefined templates, while in GLBP method, only the information of single direction is applied without templates. Thus, B-LTP method helps reflect local properties of different types of human bodies without becoming over sensitive to environmental intensity variations.

Besides, the feature dimensions of B-LTP are considerably smaller than that of other methods. For instance, the length of HOG, HOT and GLBP methods for a 64 × 128 image is 3780, 3360, and 5880, respectively, while our proposed B-LTP method comprises only 1008 dimensions, thereby increasing the computation speed.

Experiments are evaluated on INRIA dataset. We measured the overall detection rates of our method and compared the results with the detection rates of other related methods, for the same conditions and dataset. The results show that B-LTP method achieves the highest detection rate than other referred method, with a 95.3% hit rate at 10⁻⁴ false positive per window (FPPW) and a 98.8% hit rate at 10⁻³ FPPW under radial basis function (RBF) kernel based detector using support vector machine (SVM).

In chapter 3 [Non-Redundant Gradient Semantic Local Binary Patterns Method], another method is proposed named Non-Redundant Gradient Semantic Local Binary Patterns (NRGSLBP) for human detection as a modified version of conventional SLBP. Feature extractions of this method are carried out for both intensity and gradient magnitude image so that texture and gradient information are combined which means more data on the human body for each training sample can be achieved than HOG method and SLBP method. Moreover, non-redundant patterns are adopted on SLBP method, allowing better discrimination. Compared with SLBP method, no additional cost of the feature dimensions NRGSLBP method is necessary and the calculation complexity is considerably smaller than that of other methods. Furthermore, better than GLBP and SLBP methods, NRGSLBP method could not only be robust to the relative contrast between the background and foreground but also arrange similar pixel into connected histogram bins.
Experimental results on both INRIA and Daimler datasets show that the detection rate of our proposed method outperforms those of other methods such as HOG, SLBP, HOT, GLBP, and B-LTP methods. It could achieve a 95.5% detection rate at $10^{-4}$ FPPW and a 98.2% detection rate at $10^{-3}$ FPPW on INRIA dataset and a 79.4% detection rate at $10^{-4}$ FPPW and an 89.3% detection rate at $10^{-3}$ FPPW on Daimler dataset using linear SVM.

In chapter 4 [Multi-scale Extension and Acceleration Methods], a multi-scale framework is proposed for human detection as an extension of previous proposed B-LTP method and NRGSLBP method. So two multi-scale based extensions, multi-scale bidirectional local template patterns method (MBLTP) and multi-scale non-redundant gradient semantic local binary patterns method (MNRGSLBP) are proposed.

For multi-scale framework, it not only integrates the textural and gradient information according to the single-scale method but also calculates information for additional feature vectors by adjusting the scale of the training samples. These additional feature vectors contain multi-scale information on the samples, which can make the feature more discriminative than its original form.

Experimental results on INRIA dataset show that the detection rate of our proposed MBLTP method outperforms those of other multi-based methods such as the multi-level histogram of oriented gradient method (multi-level HOG) (S. Maji, CVPR 2008), multi scale block histogram of template method (MB-HOT) (S. Tang, ICIP 2010), multi-resolution histogram of oriented gradient method (multi-resolution HOG) (Z. Wei, ICCV 2007) and HOG-LBP method (X. Wang, ICCV 2009). Our MBLTP method achieved a detection rate of 98.27% at $10^{-5}$ FPPW and 99.1% at $10^{-4}$ FPPW using linear SVM, which could achieve the best performance compared with other methods.

In order to make our algorithms suitable for real-time system, graphics process unit (GPU) is also applied for acceleration, the original B-LTP method and MBLTP method were adopted for this experiment. For the experimental environment, a quad-core Intel Core i7 is used for CPU calculations, while a NVIDIA (GeForce GTX 680) graphics card with 1536 cores is used for GPU calculations. In the feature extraction process, if the block size of the feature is $16 \times 16$ (256), six thread blocks with 256 threads are utilized for maximum 1536 parallel computations in the system. For an image size of $640 \times 480$, CPU takes about 13.3 seconds, while GPU takes about 29.49 milliseconds using B-LTP method. By applying the GPU rather than the CPU, the overall calculation time can be about 450 times shorter, which makes the system suitable for real-time.

In chapter 5 [Conclusion], we conclude the contributions of this dissertation.