International Journal of Future Generation Communication and Networking
Vol. 9, No. 5 (2016), pp. 143-152
http://dx.doi.org/10.14257/ijfgcn.2016.9.5.14

Detecting Moving Object via Projection of Forward-Backward Frame Difference

Zhuo Bian[1] and LiangliangWang[2]

[1]School of Architecture, Harbin Institute of Technology, Harbin 150001, China
[1]Art Academy of Northeast Agriculture University, Harbin 150001, China
[2]State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China
[1]bianzhuo1219@126.com, [2]yueyangmeng@163.com

Abstract

Frame difference is a quick dissimilarity based segmentation approach for object detection, unfortunately, it gets trapped in over-segmented when the pixels of interest over time overlap each other. This paper presents a rather fast visual object detection approach capable of approximating the location of moving object under heavy background noise or big overlap caused by negative similarity. Specifically, frame forward-backward difference concept is proposed to extract object features in current frame through fusion of pixel-based current-previous and current-following frame difference. Based on this, we formulate object localization applying the statistics of horizontal-vertical projection of the fused difference. Therefore, our object detection can be regarded as a direct thresholding process which guarantees high efficiency while holds good accuracy performance. We evaluate our method on Weizmann human action dataset and some traffic videos for both single and multiple objects detection which demonstrates its applicability and prospect.

Keywords*: Object detection, target localization, frame forward-backward difference, horizontal-vertical projection statistics*

1. Introduction

Object detection is a basic but critical topic allied closely with image segmentation, object tracking and recognition in computer vision, all of which have considerable potential demand in. the field of video surveillance, human-computer interaction, virtual reality, robotics, intelligent transportation system and others [1-3].

Generally and briefly, the process of object detection can be broken down into two stages: first, segment the salient foreground of the image by analyzing interest features and second, outline the boundary of the most probable parts. It is no wonder that a large number of attempts have been made to exploit what features can be used to represent the interest area from noninterest parts which plays the key role, despite central challenges from scenario complexity, scale variation, occlusion, illumination changing and others.

Existing techniques can be discussed from two aspects according to the acting subject for study which goes to a single independent image and video frames comprising of image mutual relevance over time additionally. The major purpose of object detection in a single image is to find out the spatial dissimilarity between every pixel or pixel group to generate a distinguishable symbolic description of image, which is often difficult and consuming to accomplish. SIFT (Scale-Invariant Feature Transform) [4], HOG (Histogram of Oriented Gradient) [5], SURF (Speeded-Up Robust Features) [6], FAST (Features from Accelerated Segment Test) [7] are widely studied spatial detectors in object detection as their merit on accuracy, unfortunately, the detectors are often time or computation consuming, also, the correlation of objects is ignored, which indicates a

ISSN: 2233-7857 IJFGCN

learning or matching process is often needed, for instance, in [4], BBF (Best-Bin-First) [8] algorithm is implemented to match the SIFT points for localization and SVM (Support Vector Machines) is employed in [5] for classifying HOG detectors.

With respect to the video frames based object detection, it's simpler to efficiently achieve detection by virtue of combination of temporal features whose prerequisite is there is relative motion, meanwhile, techniques based on spatial dissimilarity are also of great significance. Background subtraction, frame difference and optical flow techniques are typical strategies we'd like to coverer. Background subtraction is a most discussed method to get interest region via current image subtracts the background estimated based on the statistics of at least some known status, in [9], a video coding method combining GMM (Gaussian Mixture Model) [10] is proposed by decoding an overhead flag marked in possible foreground areas. Frame difference approach is another popular strategy to distinguish whether a pixel belongs to foreground through evaluation of difference between two or three frames using the set threshold, as a good example, Wang *et al.*, [11] achieve moving object detection by difference thresholding considering the temporal information in continuous frames. Optical flow based methods depend on image optical flow computation which reflects object's relative motion, Wang and Snoussi [12] employ Horn-Schunck [13] optical flow computation method to construct a histogram, based on which SVM is applied for final classification and localization. It is not surprising, these spatial-temporal techniques are often limited to constrained scenario because with no exception, they are either sensitive to brightness variation, shadow occlusion or trapped in poor precision although big progress has been made, as well, they have great overlap with techniques of detection in single image.

The main goal of this paper is to create an applicable detecting scheme enables computers to observe target and label the interest area in video frames exceedingly quick. Our work is distinguished by detection and localization via adopting a forward-backward difference projection statistics scheme which is available for real-time application with a comparatively high reliability in different scenarios. Different with conventional frame difference method, which often leads to poor accuracy consists of foreground holes if the brightness intensity disparity between two or three frames is not obvious, and especially in the case of there is slight movement of a rigid object, the forward-backward difference introduced in our work is a fusion process consists of forward and backward difference which has good robustness performance allows slight rigid object movement with big brightness intensity overlap, object shape or scale transformation. To summarize, first, frame difference between current-forward and current-backward focused on brightness intensity domain is computed respectively, after which, the interest region in current frame is further determined by finding out the fusion of foreground pixels between the two differences. Then, the fused foreground brightness features are projected in both horizontal and vertical direction for statistics. Thus, target can be detected and localized via deciding the intersection of projection transition points.

2. Methodology

We present our moving object detection approach based on horizontal-vertical projection statistics of frame forward-backward difference of brightness intensity in this section.

2.1. Forward-backward Difference

Suppose there are three frames $\{F_{t-1}, F_t, F_{t+1}\}$ sized by $m \times n$ in a video sequence, and the only prerequisite is there are even small movement disparities frame by frame in one-way direction under small illumination changing environment. Let$\{I(x, y, i) : i = t-1, t, t+1\}$ denotes the brightness intensity of a pixel point (x; y) at previous, current or following time on gray scale. For each pixel, frame difference between F_t

and F_{t-1} $is\Delta\frac{-1}{x,y,t}$=$1(x;y;t)$ -$I(x; y; t-1)$, and moreover, forward difference is defined as:

$$\mathrm{D}\frac{-1}{x,y,t} = \begin{cases} \Delta\frac{-1}{x,y,t}, \Delta\frac{-1}{x,y,t} > 0 \\ 0, \quad \Delta\frac{-1}{x,y,t} \leq 0 \end{cases} \tag{1}$$

While frame difference between F_t and F_{t+1} $is\Delta\frac{+1}{x,y,t}$ = $I(x; y; t)$-$I(x; y; t+1)$, and backward difference is:

$$\mathrm{D}\frac{+1}{x,y,t} = \begin{cases} \Delta\frac{+1}{x,y,t}, \Delta\frac{+1}{x,y,t} > 0 \\ 0, \quad \Delta\frac{+1}{x,y,t} \leq 0 \end{cases} \tag{2}$$

2.2. Projection of Forward-backward Difference

Since the intensity varies only where the pixel moves, which means the foreground pixel, so non-zero locations of $D\frac{+-}{x,y,t}$ and $D\frac{+-}{x,y,t}$ are moving foreground pixels perhaps with contour losing as the brightness intensity overlap. Furthermore, forward-backward difference by fusion of foreground pixels is formulated to estimate the whole moving object without location features losing in F_t. Therefore, the segmented frame F_t is represented as a matrix: $(d_{xy})m\times n$, where:

$$d_{xy} = \begin{cases} 1 \;\; D\frac{-1}{x,y,t} + D\frac{+1}{x,y,t} \geq T \\ 0 \;\; D\frac{-1}{x,y,t} + D\frac{+1}{x,y,t} < T \end{cases} \tag{3}$$

T is the threshold that can be adjusted according to the practical scenario.

So, no matter whether there is overlap, if there is relative motion between three frames, the edge of objects can be obtained by statistics of projection of forward-backward difference. Our projection in horizontal direction is defined as:

$$P_x = \sum_{y-1}^{n} d_{xy} \tag{4}$$

While projection in vertical direction is defined as:

$$P_y = \sum_{y-1}^{m} d_{xy} \tag{5}$$

Finally, after P_x and P_y are computed, the interest area of moving object Area is defined as:

$$\text{Area}= \{(x,y)|P_x > 0, Py > 0\} \tag{6}$$

2.3. Single Object Detection

In order to present our approach more directly, let a rectangle represents a moving object only along horizontal direction (x axis) no matter whether the brightness intensity and size of object are same in F_{t-1}, Ft, and F_{t+1}. As mentioned before, the movement is one-way frame by frame, so we can imagine that if object's left and right edge locations in F_{t-1} (H_s^-, H_f^- ;), F_t (H_s;H_f)and F_{t+1} (H_s^+, H_f^+) satisfy:

$$(H_s^- - H_f^-) + (H_s^+ - H_f^+) \geq (H_f - H_s) \tag{7}$$

Which means the movement of the object in the three frames does not have overlap or the fusion of forward and backward difference (shading in rectangles) overlap covers the original range in Ft, which also means the relative movement is very large, the projection of forward-backward difference in x direction would produce two transition points indicating the left (H_s) and right edge (H_f) in horizontal direction as illustrated in Figure 1, since the fusion of the forward and backward difference values defined in our method between the two edges are continuous to some extent. Besides, the projection value at each position along horizontal direction (x axis) is determined by the total statistics of intensity difference including both forward and backward along vertical direction. In addition, the unknown but distinguishable part after difference in rectangle is denoted via blank in our figures and the background part is ignored although their value may be not completely zero.

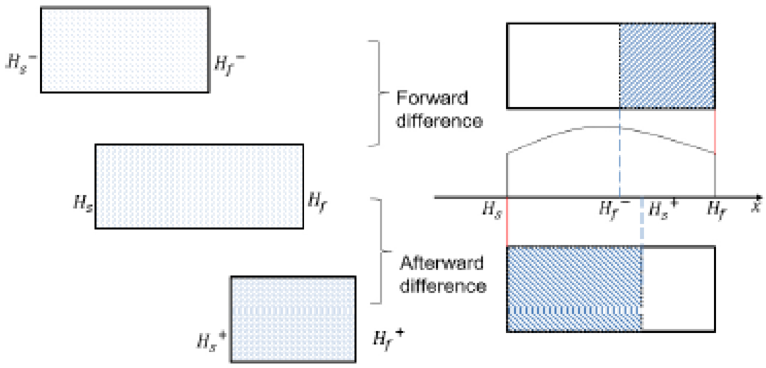

Figure 1. Projection of Forward-backward Difference in Horizontal Direction of Large Relative Motion with Object Size Variation

Otherwise, if the left and right edge locations of object satisfy:

$$(Hf - H_f^-) + (H_s^+ - Hs) < (H_f - H_s) \tag{8}$$

Which means object movement is small as illustrated in Figure 2, the projection of forward-backward difference generates four obvious transition points, projection of the unknown but distinguishable part between H_s^+ and H_f^- f is represented by a conic curve. Also, we can estimate the object's left and right edge location in F_t by finding out the smallest and largest ones.

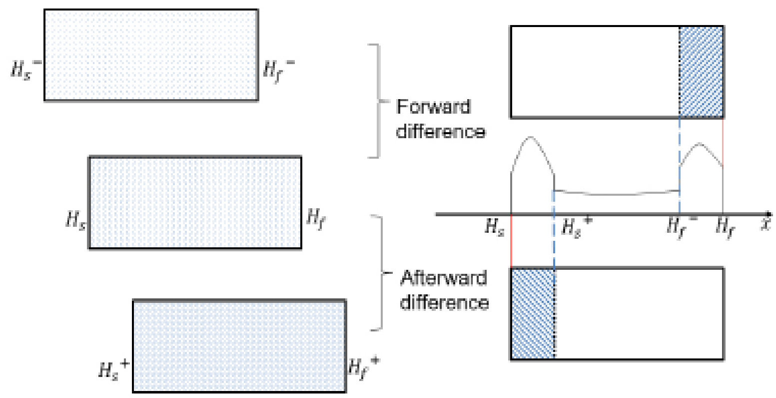

Figure 2. Projection of Forward-Backward Difference in Horizontal Direction of Small Relative Motion without Object Size Variation

Along y axis direction, it's obvious that the projection of forward-backward difference describes all the locations on the edge as there is no relative movement in the direction, so two edge transition points would be produced whose values are decided by the range of relative motion in x direction.

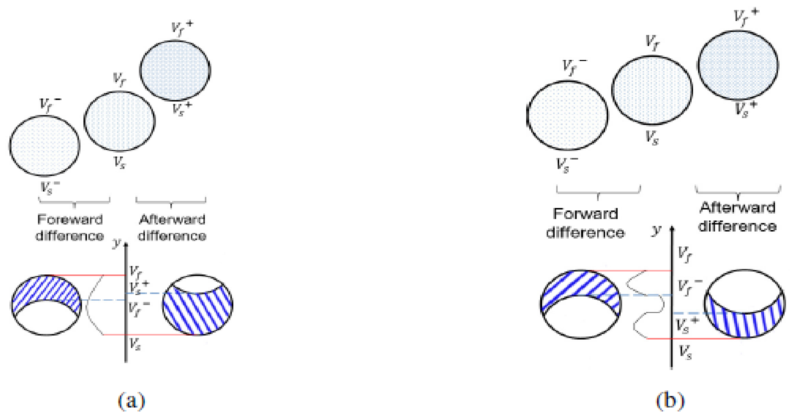

Figure 3. Projection of Forward-afterward Difference in Vertical Direction of Large (a)/ Small (b) Relative Motion without Object Size Variation

Similarly, if there is only movement along the vertical direction (y axis), circles are used for denoting objects, and if object's top and bottom edge locations in three one-way frames $V_{t-1}(V_s^- - V_f^-)$, V_{t+1} $(Vs; Vf)$ and $V_{t+1}(V_s^+ - V_f^+)$ satisfy:

$$(Vf - V_f^-) + (V_s^+ - Vs) < (Vf - Vs) \tag{9}$$

As illustrated in Figure 3a, two transition points along vertical direction can be used for vertical detection and localization, or else, satisfy:

$$(Vf - V_f^-) + (V_s^+ - Vs) < (Vf - Vs) \tag{10}$$

As illustrated in Figure 3b, the smallest and largest transition points of four can be used for vertical localization.

As well, suggested by Figure 3, as the characteristics of object shape, the forward and afterward difference may overlap between V_f^- and V_s^+, we regard this form of overlap as positive overlap since it has no impact on interest edge locations.

Equally, along x axis direction, the projection of forward-afterward difference can describe all the locations on the edge, and two edge transition points would be produced whose values are decided by the range of relative motion in y direction.

As to multiple objects detection, suppose there are two objects in three one-way frames only moves in x direction, edge locations and their relationships are the points what we concern.

If both the two objects move slowly, from previous mentioned work, we can infer that the projection of forward-backward difference includes eight transition points, four of which suggest the locations of object edge. If both two objects move fast, four transition points would be generated after projection of forward-backward difference as illustrated in Figure 4b, while every localize an object edge. Another case is one object moves fast while the other moves slowly, it's obvious that six transition points would be obtained, based on which four points can be selected for edge localization.

During our multiple objects detection discussion, we suppose that the two objects have the same ranges vertically though our scheme is applicable for otherwise. Also, it's easy to conclude that the projection in y axis direction would characterized by two edge transition points.

In all, it's recommended to utilize projection of forward-backward difference scheme to detect multiple objects under condition that the brightness intensity on the edge is not to be cut by its neighbors. In spite of cases of multiple objects which only move along vertical direction or random direction are not detailed in this section, we definitely can conclude that our approach has potential prospect for three or more complicated objects detection problems.

3. Experimental Results

Our algorithm was tested using some sequences from Weizmann action dataset in addition to ChangeDetection.NET dataset 2014.

Weizmann dataset contains 90 videos separated into 10 actions (walking, running, jumping, galloping sideways, bending, one-hand waving, two-hands waving, jumping in place, jumping jack, and skipping), performed by 9 single persons, the background is relatively simple. A more detailed illustration can be found in [14], and some examples are given in Figure 4.

In our work, we pick up sequences of galloping sideways with size 180 × 144, both the thresholds are set to be 20 and a median filter pre-processing process with a 3 × 3 window is implemented. We show our projection and detection results depicted on the images after difference (Figure 5a) and on original images (Figure 5b), respectively. Clearly, the detection accuracy is comparatively high. It is evident that that our algorithm outperforms the conventional frame difference approach for detection purpose as the edges of interest are retained and it is also easy to imagine that the proposed approach is especially suited for single object detection.

Figure 4. Example Frames from Weizmann Dataset

Moreover, a 1700-frame sequence under scenario of highway comprising one and multiple vehicles is selected from ChangeDetection.NET dataset 2014 [15] for testing the robustness of our algorithm as shown in Figure 5(c). Here, the background is slightly changing, shadow and occlusion exist, thus, median filter with a 2× 2 window is carried out and the threshold is set to be 20, too. Figure 5(d) shows the result. Since the scenario is complicated where negative overlap appears very often, only projection of forward-backward is given as instance for interest area detection, further processing is needed for accurate localization of each object and will be discussed in our future work.

Figure 6 shows example results of continuous motion detection, which indicates our algorithm is rather efficient and very suited for continuous motion detection issue as the boundaries of motion is completely detected frame by frame.

Figure 7 gives walking detection results using our algorithm on Weizmann dataset in many scenarios. The projection in horizontal and vertical directions are depicted respectively while the interest motion area is given, from which we can conclude that our algorithm has a good robustness performance.

Figure 8 illustrates examples of jumping detection using frames from Weizmann dataset in different scenarios. It is worth noting that our method is applicable for all other motion detection issues as long as there is movement between frames.

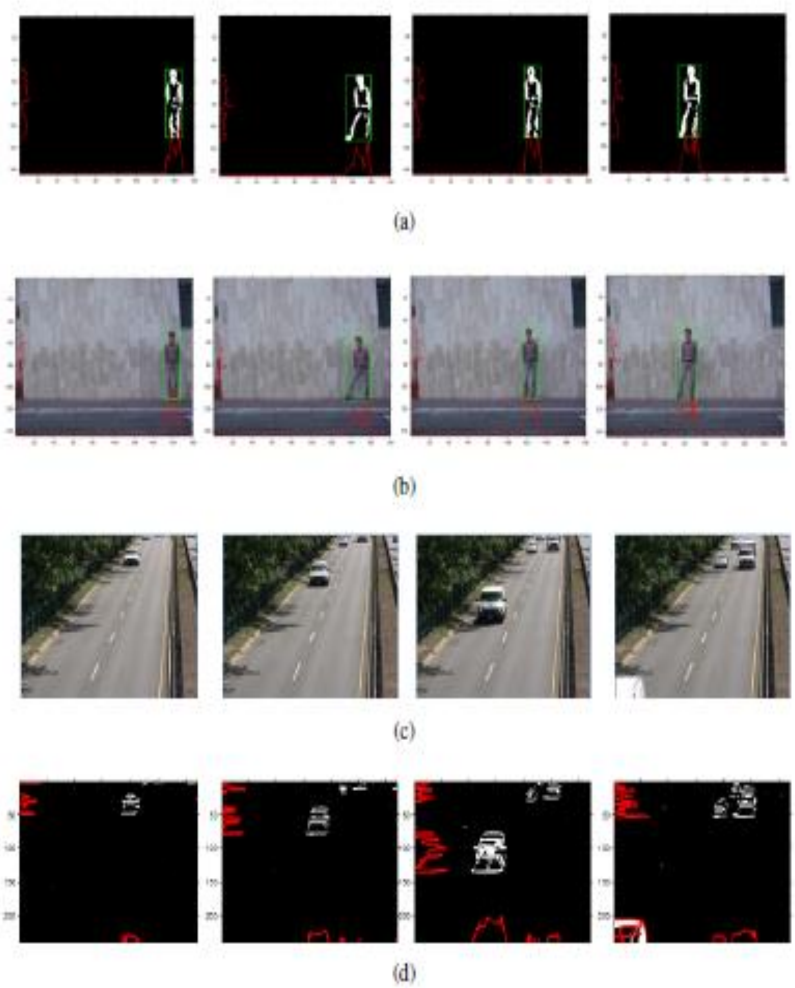

Figure 5. Example Results of Projection of Forward-backward Difference and Localization Depicted on Images after Difference (a)/ and on Original Images (b) on Weizmann Dataset, on Images After Difference (c)/ and on Original Images (d) on Changedetection.NET Dataset 2014

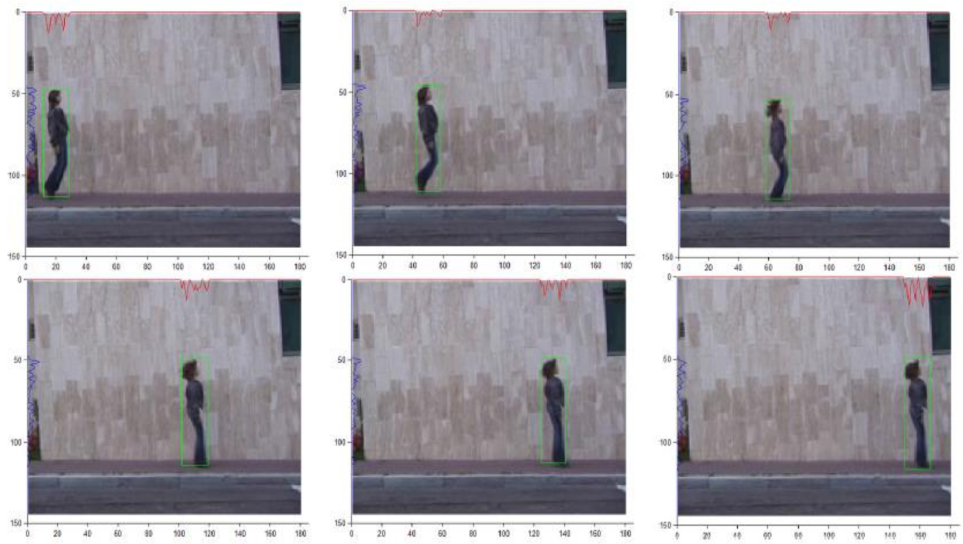

Figure 6. Example Results of Continuous Motion Detection

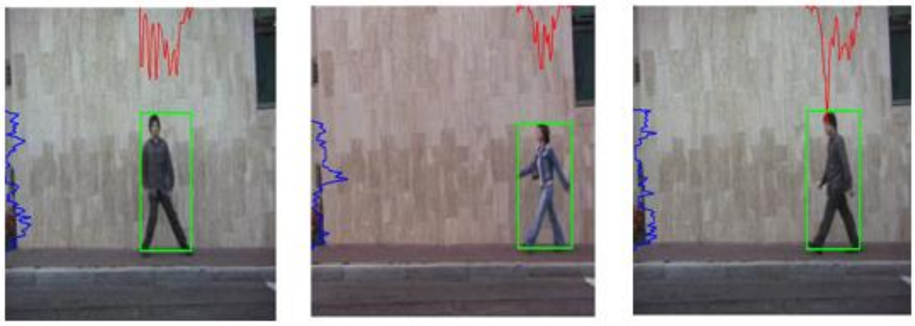

Figure 7. Example Results of Walking Detection

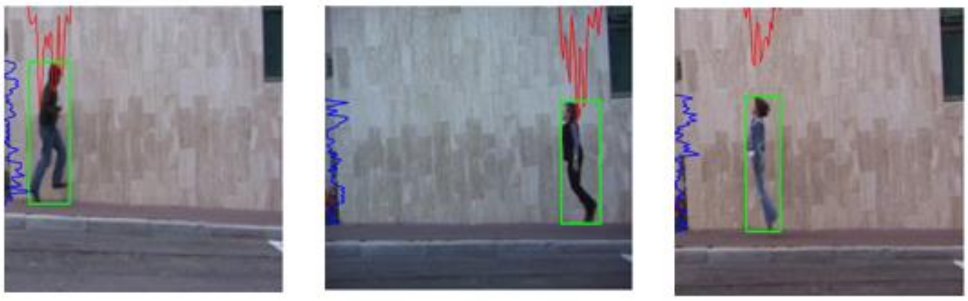

Figure 8. Example Results of Jumping Detection

4. Conclusion

This paper proposes a very efficient frame-based approach for moving object detection. Compared to previous frame difference methods, our forward-backward frame difference method can deal with the limitation of edge gap produced by overlap between connecting frames and is successful in maintaining the entire contour of object for objection task, based on which, interest area can be outlined very fast by statistics of segmented pixels projection horizontally and vertically, whose computational load is low. The tolerance of our algorithm to background noise and its applicability is validated theoretically and proved to be satisfying practically, especially for single object detecting while useful information is provided for further specific multiple objects detection. As well, the advantage of our approach covers the robustness of object size or object brightness intensity variation and the ability to identify small movement, which has wide prospect though it is not the complete solution for all detection cases.

References

[1] C. Wang and S. Zhu, A design of FPGA-based system for image processing, Review of Computer Engineer Studies, vol. 2, **(2015)**, pp. 25-30.

[2] Q. Zhong, J Zhao and C. Tong, "Motion planning of humanoid robot based on embedded vision, Review of Computer Engineer Studies", vol. 2, **(2015)**, pp. 35-38.

[3] J. Tu, "A novel building boundary extraction method for high-resolution aerial image", Review of Computer Engineer Studies, vol. 1, **(2014)**, pp. 19-22.

[4] D. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, **(2004)**, pp. 91–110.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", Proceedings in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, **(2005)** June 20-26.

[6] T. T. H. Bay and L. V. Gool, "Surf: Speeded up robust features", Proceedings in European Conference on Computer Vision, Graz, Austria, **(2006)** May 7-13.

[7] R. P. E. Rosten and T. Drummond, "Faster and better: a machine learning approach to corner detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, **(2010)**, pp. 105–119.

[8] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbor search in high-dimensional spaces", Proceedings in IEEE Conference on Computer Vision and Pattern Recognition, **(1997)** June 17-19, San Juan, USA

[9] R. H. Z. Huang and Z. Wang, Faster and better: a machine learning approach to corner detection, IEEE Signal Processing Letters, vol. 20, pp. 1058–1061, **(2013)**

[10] C. Stauffer and W. E. L. Grimson, Adaptive background mixture models for real-time tracking, Proceedings in IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, USA, **(1999)** June 23-25.

[11] J. X. Z. Wang, K. Liao and Q. Zhang, "Moving object detection based on temporal information", IEEE Signal Processing Letters, vol. 21, **(2014)**, pp. 1403–1407.

[12] T. Wang and H. Snoussi, "Detection of abnormal visual events via global optical flow orientation histogram", IEEE Transactions on Information Forensics and Security, vol. 9, **(2014)**, pp. 988–998.

[13] B. K. Horn and B. G. Schunck, "Determining optical flow", Artificial Intelligence., vol. 17, **(1981)**, pp. 185–203.

[14] L. Gorelick, M. Blank, E. Schechtman, M. Irani and R. Basri, "Actions as space-time shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, **(2007)**, pp. 2247–2253.

[15] Y. Wang, P. Jodoin, F. Porikli, J. Konrad, Y. Benezeth and P. Ishwar, "Cdnet 2014: An expanded change detection benchmark dataset", Proceedings in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, USA, **(2014)** June 24-27.

Author

Zhuo Bian, she received MS of Design Management in 2006 at Harbin Institute of Technology. Now she is a lecture in Art Academy of Northeast Agriculture University, Harbin, China. Her current research interest includes visual communication design and human-computer interaction design.