International Journal of Security and Its Applications
Vol.8, No.2 (2014), pp.315-322
http://dx.doi.org/10.14257/ijsia.2014.8.2.32

Moving Objects Representation for Object Based Surveillance Video Retrieval System

Jianping Han, Tian Tan, Longfei Chen and Daxing Zhang

College of Computer, Hangzhou Dianzi University, Hangzhou, China, 310018
hanjp@hdu.edu.cn

Abstract

Increasingly large amounts of surveillance video data have resulted in the critical need for indexing and retrieval from video databases. This paper addresses the problem of objects representation for surveillance video retrieval system. We extract dominant color histogram and edge direction histogram as the object's appearance model in which the improved histogram intersection algorithm is used to measure the similarity between two dominant color histograms. Furthermore, an agglomerative hierarchical based clustering method is proposed to select the most relevant and representative blobs for compact and effective object representation. Experimental results on real surveillance video sequences have proved the performance of our proposed approach.

Keywords: *Surveillance video retrieval, Object based indexing, Appearance model*

1. Introduction

As surveillance cameras are increasingly ubiquitous and producing huge amounts of video data, fast video retrieval system, which allows users to search their desired video clips efficiently, becomes more and more demanded. Object based indexing is crucial first step [1] in the processing chain in such systems. Moving objects are first segmented and tracked until it disappears, and the fundamental features of each tracked object are then extracted and indexed as metadata into the database [2], as shown in Figure 1. Objects in video surveillance are presented in the scene at a certain time. They are generally detected and tracked in a large number of frames. Therefore, an object is corresponding to a set of blobs and simply use of all these blobs for the object indexing and retrieval is redundant and ineffective because of the similarity between blobs [3].

This paper addresses the problem of objects representation for surveillance video retrieval. Various models and methods [1-7, 10-14, and 16-25] have been proposed for this purpose in recent years. In [4], the average MPEG-7 descriptors are computed during the object's life time. This method is not effective because the average descriptors can not accurately describe the objects with a large variety of object appearance. Calderara [5] uses a mixture of Gaussians to summarize the appearance of the object observed by a set of cameras. However, the mixture of Gaussians of objects is not reliably created and updated if the object detection and tracking are reliable. Ma [6] presents a representative blob detection method based on the hierarchical clustering in which the covariance matrix is extracted as the appearance model of object blobs. The limitation is that the covariance matrix is not able to well reflect the local changes of the object. One modification was proposed in [3], the authors have removed outliers that occur in a large number of frames using SVM.

There are two contributions in this paper. The first one is that we extract dominant color histogram and edge direction histogram as the object's appearance model in which the

ISSN: 1738-9976 IJSIA

improved histogram intersection algorithm is used to measure the similarity between two dominant color histograms. The second one concerns the clustering method to select the most relevant and representative blobs for compact and effective object representation.

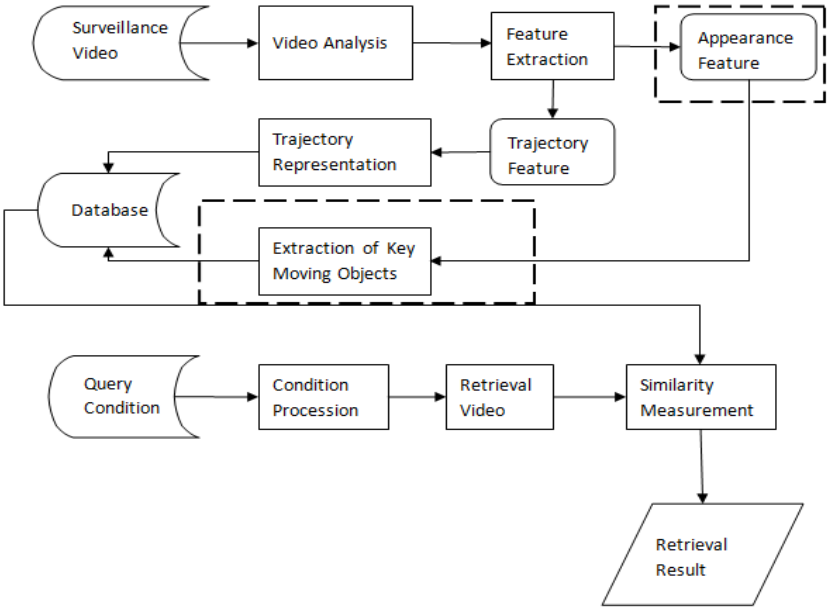


Figure 1. The Architecture of Video Retrieval System

The rest of this paper is organized as follows. Section 2 presents the details of extraction the appearance model of moving objects. Section 3 aims at selecting the representative blobs for tracked objects. Experimental results and the conclusion follow in Sections 4 and 5, respectively.

2. Appearance Model

As a moving object moves across the video scene, it is first segmented by a background subtraction approach and tracked using an appearance based algorithm. Then, we use the dominant color histogram [7] and the edge direction histogram [8] to represent the appearance of segmented blobs. Here, an object blob means a region determined by a bounding box in a frame where object is detected.

2.1. Dominant Color Histogram and its Similarity Measurement

In order to reduce the influences of the lighting conditions, we compute the dominant color histogram in the HSV color space. Each bin of the dominant color histogram is represented by a vector $v_i = <c_i, p_i>$, where c_i represents the average hue value of the i-th bin, and p_i represents the proportion of the i-th bin. Then the dominant color histogram *dchist* is constructed from the vectors v_i of the three bins with the largest p_i, as shown in Figure 2.

Commonly, the distance between the two dominant color histograms is defined as Eq. (1).

$$D_{dc}^2(dchist, dchist') = \sum_{i=1}^{3} p_i^2 + \sum_{j=1}^{3} p_j'^2 + \sum_{i=3}^{3}\sum_{j=1}^{3} 2a_{i,j} p_i p_j' \quad (1)$$

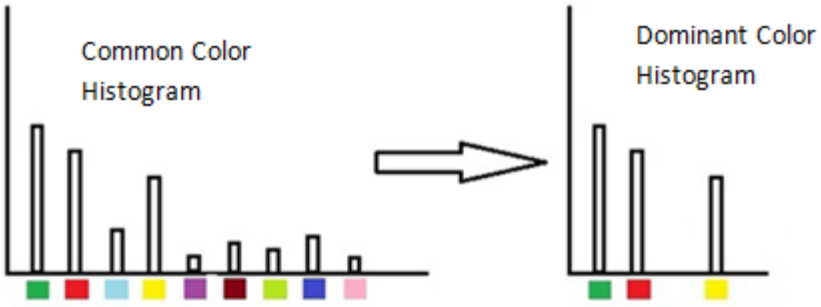


Figure 2. The Dominant Color Histogram

This method is simple and fast, but it does not match the perceptual similarity very well, and may cause incorrect ranks between similarity color distributions. Therefore, we use the histogram intersection algorithm [12] to measure the similarity between two dominant color histograms. However, due to the different dominant color histograms may be not in the same palette, the histogram intersection method can't be directly applied on the measurement of the dominant color histogram.

In this work, a common palette is first generated for the dominant color histograms. Let $dchist_1=\{<C_{1i},P_{1i}>,i=1,2,\ldots,N_1\}$, $dchist_2=\{<C_{2j},P_{2j}>,j=1,2,..,N_2\}$ be two dominant color histograms of image I_1 and I_2. The common palette is generated by searching the closest two colors between two palettes. If the minimum distance is shorter than the threshold T_m then the two colors will be merged using Eq. (2). This process continues until the minimum distance is larger than the threshold T_m.

$$c_{m(i,j)} = (p_{1i}c_{1i} + p_{2j}c_{2j}) / (p_{1i} + p_{2j}) \quad (2)$$

A common palette is composed of the merged colors and unmerged colors from the two palettes. The merged palette forms a common color space for the two histograms. Using the color space, we can redefine the histograms of $dchist_1$ and $dchist_2$. The redefined histogram $dchist_{1m}$ is given as Eq. (3).

$$dchist_{1m} = \{< c_{mj}, p_{1mj} >, j = 1,..., N_m\} \quad (3)$$

Where $P_{1mj} = \sum_{i=1}^{N_1} P_{1i}$, if the distance between c_{1i} and c_{mj} among c_m is minimum and shorter than the threshold T, otherwise $P_{1mj}=0$. The $dchist_2$ can also be redefined as $dchist_{2m}$ using this method.

Through the above steps, we get a common palette and two redefined histograms, and then we use the histogram intersection [14] to measure the similarity between the two histograms as Eq. (4).

$$D_{dc}(dchist_{1m}, dchist_{2m}) = 1 - \sum_{i=1}^{N_m} \min(p_{1mi}, p_{2mi}) \quad (4)$$

2.2. Edge Direction Histogram and its Similarity Measurement

The edge direction histogram is used to represent the shape feature of an object blob [15]. In order to describe both the global and local changes of the object, the edge direction

histogram is composed of three parts, the global histogram [16], semi-global histogram [17] and the local histogram [18]. Firstly, the image is divided into 4×4 blocks, as shown in Figurer 3 (a). At each block $b_{i,j}$ we count the directions of each pixel using the Sobel edge direction, and the vertical edges, horizontal edges, 45°edges and 135° edges are used to construct the edge direction histogram[19]. Every histogram has four bins, corresponding to the four directions. The edge direction histogram of $b_{i,j}$ at direction *dir* is constructed as follow:

$$ehist_{i,j,dir} = v_{i,j,dir} \sum_{(x,y)\in b_{i,j}} \delta_{dir}(e_{x,y}) \tag{5}$$

$$\delta_{dir}(e_{x,y}) = \begin{cases} 1, & \left|e_{x,y} > e\right| \ and \ e_{x,y} \in dir, \\ 0, & otherwise \end{cases} \tag{6}$$

Where $e_{x,y}$ is the sobel edge direction at point (x,y). $v_{i,j,dir}$ is the normalize factor, and e is a threshold that used to determine whether $e_{x,y}$ is a non-direction edge or not. The local histogram totally has 16×4=64 bins, and represented as $ehist^l$. According to the local histogram, we can get the global histogram $ehist^g$ by accumulating the local histogram. In order to compute the semi-global histogram, the local block is combined in the way shown in Figure 3(b, c, d). Then the semi-global histogram $ehist^s$ is constructed by accumulating the local block histogram in each semi-global block. The semi-global histogram totally has 13×4=52 bins. The distance between two edge direction histograms is defined by the following equation [20]:

$$D_{ed}(v_{ed}, v'_{ed}) = \sum_{i=1}^{64} \left|ehist_i^l - ehis\ t_i'^l\right| + \sum_{i=1}^{4} 5 \times \left|ehist_i^g - ehis\ t_i'^g\right| + \sum_{i=1}^{52} \left|ehist_i^s - ehis\ t_i'^s\right| \tag{7}$$

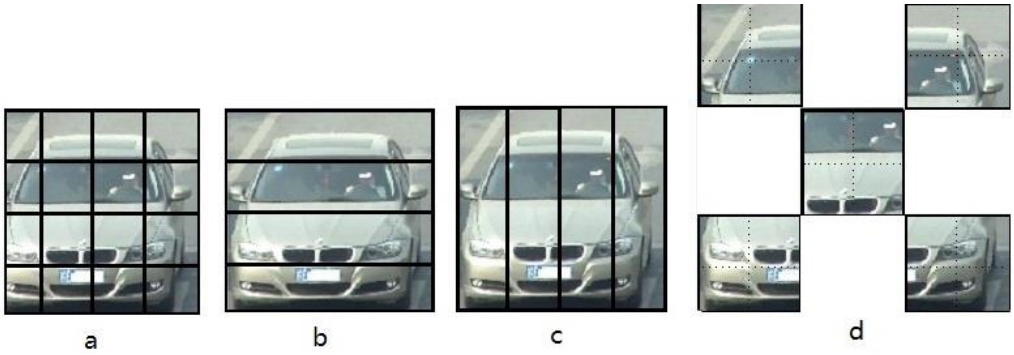


Figure 3. The Bounding Boxes of the Blocks for Construction of Edge Direction Histogram

3. Representative Blobs

Let $S^k = \{C_i^k, i = 1, 2, ..., n\}$ be the object blobs set of the k^{th} trajectory, where C_i^k represents the appearance model $\{dchist_i, ehist_i\}$ of the i^{th} blob in the k^{th} trajectory. The purpose of clustering is to get the compact subset $S^{r(k)} = \{C_i^{r(k)}, i = 1, 2, ..., m\}$ from S^k, where $m<<n$. This compact representation needs to contain all the typical appearances present within the set S^k. It should capture various appearances of blobs as it moves across the scene. Here, we use an

agglomerative hierarchical based clustering [22] driven by the similarity measurement defined by Eq. (4) and Eq. (7).

First, we use the agglomerative method to identify the number of the clusters. Initially, each C_i^k is placed in its own group, then we have *n* initial clusters, each of them only has one single blob. In order to merge the closest pairs of clusters, as for the proximity between clusters , we use average proximities as Eq. (8) and Eq. (9)

$$proximity = average \{\rho(C_i, C_j)\} \tag{8}$$

$$\rho(C_i, C_j) = 1 - \alpha_{dc} D_{dc}(dchist_i, dchist_j) - \alpha_{ed} D_{ed}(ehist_i, ehist_j) \tag{9}$$

Where α_{dc} and α_{ed} are the weighting coefficient corresponding to the distances of dominant color histogram and edge direction histogram, $0 \le \alpha_{dc} \le 1$, $0 \le \alpha_{ed} \le 1$. After the clustering of t^{th} layer, there are $\lceil n/(2^t) \rceil$ clusters left. And the iteration will be ended when there are only four clusters left.

Due to the errors of tracking or the presence of occluding, there are some exceptional clusters which should be discarded [23]. We find these exceptional clusters according to the number of blobs in each of them. If the number of the blobs in one cluster is less than T_c, then the cluster is discarded. Once the clustering is done, a representative blob is computed to describe the whole cluster. Each representative blob C_r can be got by the following equation [24].

$$r = \max_{j=1,2,3,\ldots,n, j\neq i} \sum_{i=1,2,3,\ldots,n} \rho(C_i, C_j) \tag{10}$$

4. Experimental Results

Here, we present experimental results on the proposed representative blobs selection approach. Two surveillance videos were used: one is a video sequence containing moving car, and the other is a video sequence which contains a person walking in a lateral direction and appearing in different poses. Both objects are presented in the scene for several seconds, and with a variety of appearance. Figure 4(a) and 5(a) shows original sequence of moving object blobs, which is determined by object segmentation and tracking algorithm [9]. In Figure 4(b) and 5(b), we depict the clustering results, and the corresponding representative blobs are given in Figure 4(c) and 5(c). It can be found that the representative blobs extracted by the proposed method can well reflect the variety of appearance aspect of each tracked objects.

The proposed approach is used in our surveillance video retrieval system. We apply the method described in Section 2 and 3 to the construction of models for tracked moving objects, and the appearance model of selected representative moving blobs are indexed as metadata into the database. Figure 6 shows a demonstration of user interface of the retrieval system. The query image is shown at the bottom-right, and the corresponding snapshots of retrieved objects are shown in the left, which are displayed in the order of top-down and then left-right according to the similarity. Users can select any retrieved object and playback the video clip which contains the object.

5. Conclusion

In this paper, we have presented a method to deal with the problem of objects representation for surveillance video retrieval. Dominant color histogram and edge direction histogram is extracted as the object's appearance model in which the improved histogram intersection algorithm is used to measure the similarity between two dominant color histograms. Furthermore, we have also proposed a method to select the most relevant and representative blobs for compact and effective object representation. This work was partly supported NSF of China (Grant No 61272391).

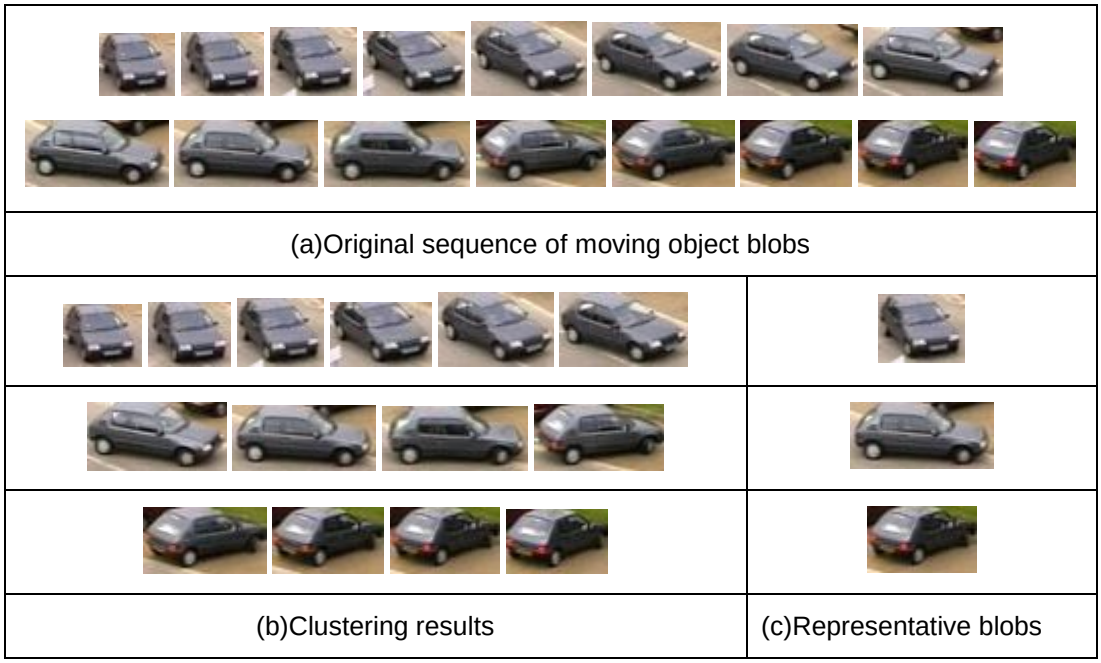
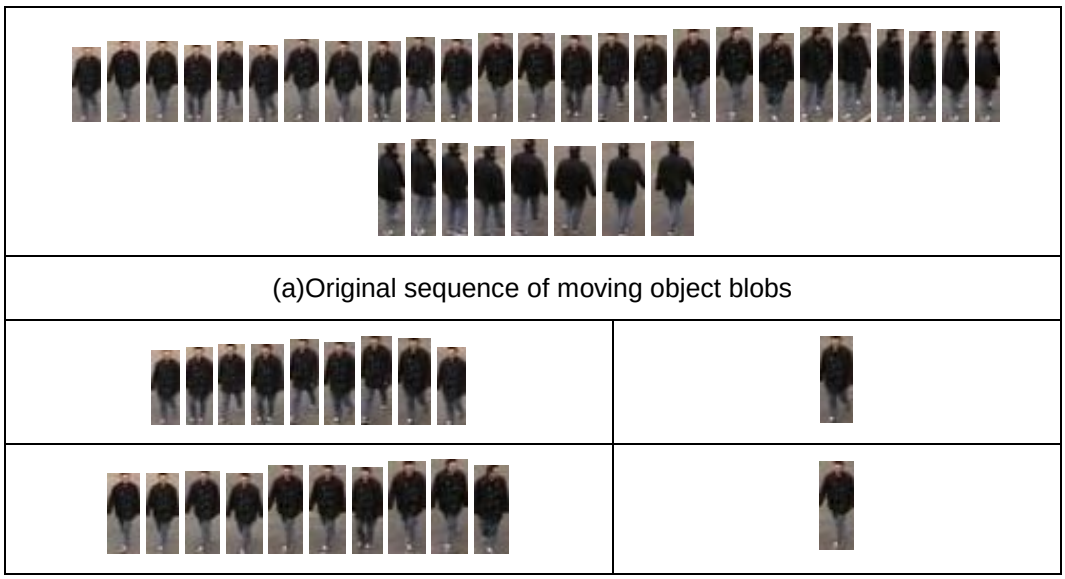
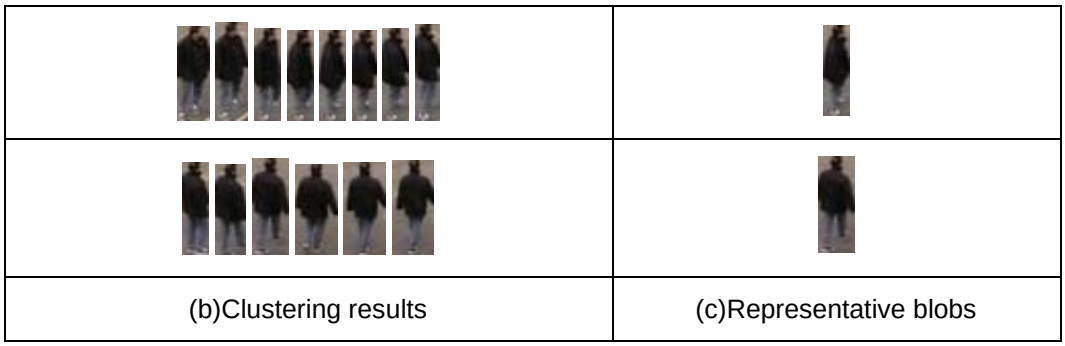


Figure 4. Selection Results of Representative Blobs for a Moving Car





(b)Clustering results | (c)Representative blobs

Figure 5. Selection Results of Representative Blobs for a Walking Man

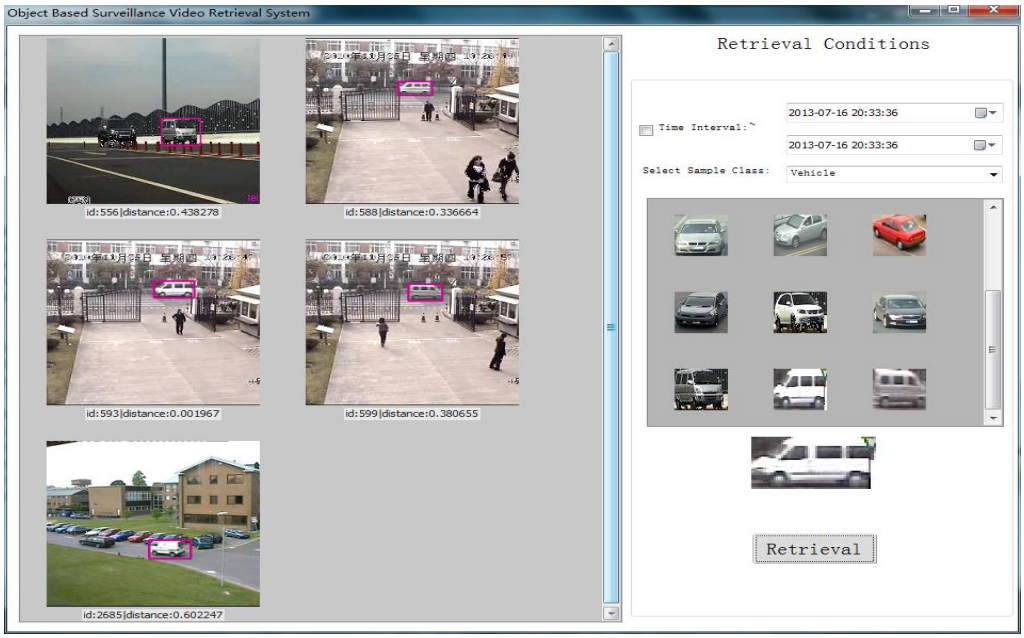


Figure 6. User Interface of the Surveillance Video Retrieval System

References

[1] L. M. Brown, "Example-Based Color Vehicle Retrieval for Surveillance", IEEE Seventh International Conference on Advanced Video and Signal Based Surveillance, Boston, USA, **(2010)** August 29-September 1.

[2] W. Hu, D. Xie, Z. Fu, W. Zeng and S. Maybank, "Semantic-Based Surveillance Video Retrieval", IEEE Transactions on Image Processing, vol. 16, no. 4, **(2007)**.

[3] T. L. Le, M. Thonnat, A. Boucher and F. Bremond, "Surveillance Video Indexing And Retrieval Using Object Features And Semantic Events", International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, no. 7, **(2009)**.

[4] J. S. C. Yuk, K. Y. K. Wong, R. H. Y. Chung, K. P. Chow, F. Y. L. Chin and K. S. H. Tsang, "Object-Based Surveillance Video Retrieval System with Real-Time Indexing Methodology", International Conference on Image Analysis and Recognition, Montreal, Canada, **(2007)** August 22-24.

[5] S. Calderara, R. Cucchiara and A. Prati, "Multimedia Surveillance: Content-based Retrieval with Multicamera People Tracking", ACM International Workshop on Video Surveillance & Sensor Networks, Santa Barbara, CA, USA, **(2006)** October 27.

[6] Y. Ma, B. Miller and I. Cohen, "Video Sequence Querying Using Clustering of Objects' Appearance Models", International Symposium on Visual Computing, Lake Tahoe, NV, USA, **(2007)** November 26-28.

[7] N. Yang, W. Chang, C. Kuo and T. Li, "A Fast MPEG-7 Dominant Color Extraction with New Similarity Measure for Image Retrieval", Journal of Visual Communication and Image Representation, vol. 19, no. 2, **(2008)**.
[8] X. Gao, "Image categorization: Graph edit distance+ edge direction histogram", Pattern Recognition, vol. 41, no. 10, **(2008)**.
[9] J. Han, M. Zhang and D. Zhang, "Background Modeling Fusing Local and Global Cues for Temporally Irregular Dynamic Textures", Advanced Science Letters, vol. 7, **(2012)**.
[10] X. Chen and C. Zhang, "An Interactive Semantic Video Mining and Retrieval platform–Application in Transportation Surveillance Video for Incident Detection", Sixth International Conference on Data Mining, Hong Kong, China, **(2006)** December 18-22.
[11] R. Feris, S. Pankanti and B. Siddiquie, "Learning Detectors from Large Datasets for Object Retrieval in Video Surveillance", IEEE International Conference on Multimedia and Expo, Melbourne, Australia, **(2012)** July 9-13.
[12] C. Kim and J. Hwang, "Object-Based Video Abstraction for Video Surveillance Systems", IEEE Transactions on Circuits System for Video Technology, vol. 12, no. 12, **(2002)**.
[13] A. B. Hampapur, L. Feris, R. Senior, A. C. F Shu, Y. Tian, Y. Zhai and L. Max, "Searching surveillance video", IEEE Conference on Advanced Video and Signal Based Surveillance, London, United Kingdom, **(2007)** September 5-7.
[14] T. Le, A. Boucheryz, M. Thonnatx and F. Bremond, "Surveillance video retrieval: what we have already done?", Third International Conference on Communications and Electronics, Nha Trang, Vietnam, **(2010)** August 11-13.
[15] L. Wang, W. Hu and T. Tan, "Recent developments in human motion analysis", Pattern recognition, vol. 36, no. 5, **(2003)**.
[16] C.-F. Shu, A. Hampapur and M. Lu, "IBM smart surveillance system (S3): A open and extensible framework for event based surveillance", IEEE conf. Advanced Video and Signal Based Surveillance, Cerno, Italy, **(2005)** September 15-16.
[17] J. Li, Z. Wang and B. Zhang, "The Interactive Video Retrieval System in SMARTV 2009", CIVR 2009 - Proceedings of the ACM International Conference on Image and Video Retrieval, Santorini, Fira, Greece, **(2009)** July 8-10.
[18] C. Hau Chan and G. J. F. Jones, "An Affect-Based Video Retrieval System with Open Vocabulary Querying", 8th International Workshop on Adaptive Multimedia Retrieval, Linz, Austria, **(2010)** August 17-18.
[19] A. Ekin, A. M. Tekalp and R. Mehrotra, "Automatic Soccer Video Analysis and Summarization", IEEE Transactions on Image Processing, vol. 12, no. 7, **(2003)**.
[20] J. R. Smith and S-f C., "VisualSEEk: a fully automated content-based image query system", MULTIMEDIA '96 Proceedings of the fourth ACM international conference on Multimedia, Boston, USA, **(1996)** August 9-18.
[21] M. K A. Smith, "An object-based approach for digital video retrieval", Proceedings of 2004 International Conference on Information Technology: Coding and Computing, Los Alamitos, Canada, **(2004)** April 5-7.
[22] R. Jianxin Wu and J. M., "Beyond the Euclidean distance: Creating effective visual codebooks using the Histogram Intersection Kernel", Kyoto, Japan, **(2009)** September 29-October 2.
[23] T. Eun Choe and M. W. L., "Semantic Video Event Search for Surveillance Video", 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, **(2011)** November 6-13.
[24] A. N. Shehzad Khalid, "Motion Trajectory Clustering for Video Retrieval Using Spatio-temporal Approximations", Visual Information and Information Systems Lecture Notes in Computer Science, vol. 3736, no. 10, **(2006)**.
[25] B. V. Patel, "B B M. Content Based Video Retrieval Systems", International Journal of Ubi Comp., vol. 3, no. 2, **(2012)**.

Author

Jianping Han, he received the Ph.D. degree from the Department of Computer Science and Engineering, Zhejiang University, China, in 2010. Currently, he is a Associate Professor at the School of Computer Science, Hangzhou Dianzi University. His research interests include visual surveillance, pattern recognition and computer vision.

