doi:10.3772/j.issn.1006-6748.2016.02.011

Abnormal activity detection for surveillance video synopsis①

Zhu Xiaobin (祝晓斌)②*, Wang Qian*, Li Haisheng*, Guo Xiaoxia**, Xi Yan**, Shen Yang**
(*School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, P. R. China)
(**Academy of Broadcasting Science, SAPPRFT, Beijing 100866, P. R. China)

Abstract

Video synopsis is an effective and innovative way to produce short video abstraction for huge video archives, while keeping the dynamic characteristic of activities in the original video. Abnormal activity, as the critical event, is always the main concern in video surveillance context. However, in traditional video synopsis, all the normal and abnormal activities are condensed together equally, which can make the synopsis video confused and worthless. In addition, the traditional video synopsis methods always neglect redundancy in the content domain. To solve the above-mentioned issues, a novel video synopsis method is proposed based on abnormal activity detection and key observation selection. In the proposed algorithm, activities are classified into normal and abnormal ones based on the sparse reconstruction cost from an atomically learned activity dictionary. And key observation selection using the minimum description length principle is conducted for eliminating content redundancy in normal activity. Experiments conducted in publicly available datasets demonstrate that the proposed approach can effectively generate satisfying synopsis videos.

Key words: abnormal activity detection, key observation selection, sparse coding, minimum description length (MDL), video synopsis

0 Introduction

Security applications have great demands on efficient technologies for fast video browsing, retrieval or analysis, facing endlessly produced surveillance videos. Therefore, how to obtain short and comprehensive video abstractions becomes an urgent task in research domain. Video synopsis is an effective method, which makes the abstraction video greatly shorter than the original one by displaying the activities from different periods simultaneously.

Video synopsis can eliminate redundancy in the spatial-temporal domain, and generate short video abstraction. However, the existing synopsis methods[1,2] still have the following limitations: They always tend to summarize all types of activities from input videos. In video surveillance context, people mainly concern with particular activities, especially abnormal activities. The traditional synopsis video will include lots of activities people are not really interested in; They always concentrate on eliminating redundancy in the spatial-temporal domain, while neglecting the redundancy in the content domain. Too many observations for activities can make the synopsis videos chaotic and less understandable.

To address the above issues, a novel video synopsis approach is proposed based on abnormal activity detection and key observation selection. In the proposed algorithm, based on sparse coding framework, activities are classified into two types, namely abnormal type and normal type, based on which two synopsis videos are generated separately. Synopsis with abnormal activities is usually the main concern in surveillance context, while synopsis with normal activities is a complementary video. Because adjacent observations in an activity are always similar in action and appearance, thus key observation selection is adopted to eliminate content redundancy in synopsis video for normal activities. Section 1 overviews the related works. Section 2 elaborates the methodology of this work. The experiments are given in Section 3, and this study is concluded in Section 4.

① Supported by the National Natural Science Foundation of China (No. 61402023), Beijing Technology and Business University Youth Fund (No. QNJJ2014-23) and Beijing Natural Science Foundation (No. 4162019).

② To whom correspondence should be addressed. E-mail: buddysoft@sina.com

Received on Oct. 29, 2015

1 Related work

Video abstraction can be broadly divided into three categories, namely videosummarization, video skimming, and video synopsis. Video summarization techniques try to provide a summary by creating shorter video remaining descriptivesections of the original video. Typically, these techniques adopt static representations such as key-frames[3]. Although summarization based on key-frame could greatly compress theoriginal video, it loses not only the dynamic nature of video but also meaningful video contents.

Video skimming[4,5] aims to extract informative video segments from the original video to obtain a condensed summary video. Ref. [5] adopted long-term and short-term audiovisual tempo analyses to detect valuable substories of a video and combined them for video skimming. The skimming video is generated based on the selected scene periods. Although, video skimming method can generate relatively more coherent and expressive summary video than those key-frame based ones. However, people will tend to spend large amounts of time browsing video segments with little information.

Video synopsis methods break the previous framework through rearranging spatial-temporal location of foreground objects to generate a new efficient summary video, while keeping the dynamic nature of the original video[1,6-11]. In Ref. [1], activities were represented by space-time tubes. Then, energy function comprised of collision cost, activity cost and temporal consistency cost, etc., is minimized using annealing, MRF, or greedy optimization method, yielding an abbreviated synopsis video for fast browsing. Ref. [6] formulated the synopsis video generation problem as a maximum posterior probability (MAP) estimation problem, where video objects are chronologically rearranged in realtime without pre-computing the complete trajectory of activities. In Ref. [9], Feng, et al. adopted an online content-aware approach to achieve efficient video condensation. In the above mentioned methods, all the activities of the input video are equally treated and summarized together. The generated synopsis video tends to comprise lots of activities people are not really interested in. In addition, great redundancy in the content domain is always neglected in the above video synopsis methods, which leads to collisions, and negative impaction on subjective effect of video synopsis. In Ref. [10], space-time worms were correlated with an user-specified query to identify actions of interest, which were then condensed by optimizing their temporal shift, allowing simultaneous display of multiple instances of relevant activity. Motivated by the works in Ref. [10] and Ref. [11], a novel framework is proposed for video synopsis, which can overcome the above limitations of the current methods.

2 The proposed algorithm

2.1 Framework

The proposed framework is shown in Fig. 1. Firstly, background subtraction followed by a graph-based tracking[12-14] is adopted to extract moving objectactivities. Then, sparse reconstruction cost is adopted to classify normal and abnormal activities based on dense trajectory. For normal activities, key observation selection is used to eliminate content redundancy with the minimum description length (MDL) principle. Finally, two synopsis videos are generated for normal and abnormal activities, respectively.

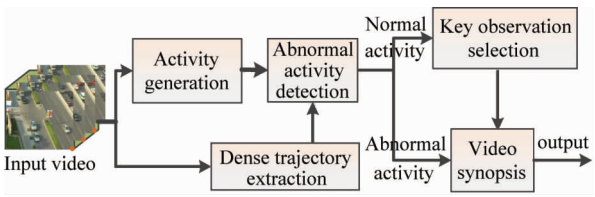

Fig. 1 The proposed framework

2.2 Abnormal activity detection

After the background/foreground segmentation, a multiple hypothesis tracker is used to track blobs[15]. Then, foreground blobs are grouped into targets by employing similar techniques as in Ref. [16]. Each activity is represented by a sequence of object masks in those frames. The space-time sequences of an object is deemed as a tube, and the central point of object is taken as motion trajectory. Motion trajectories of objects are widely used in abnormal activity detection. Sparse coding framework is suitable for model high-dimensional samples. Normal samples tend to generate sparse reconstruction coefficients with a small reconstruction cost, while abnormal one is dissimilar to any of the normal basis, thus generates a dense representation with a large reconstruction cost. Recent works[17] showed the power of sparse coding in detecting abnormal activities (events). In the proposed algorithm, abnormal activity is conducted based on the sparse coding framework.

2.2.1 Feature extraction

In Ref. [18], typical trajectories were modelled with hierarchical clustering method for identifying abnormal behavior. One limitation of trajectory-based ap-

proaches is that the detection performance greatly relies on the accuracy of foreground object extraction and tracking. The other is that single trajectory cannot well describe the overall motion of the corresponding object across the scene[19]. Dense trajectory is extensively applied to a variety of tasks, e. g., action recognition[20], and abnormal event detection[21], etc. It is capable of well describing object activities, even in complex and crowded scenes. So, dense trajectories are extracted using particle advection[22] for abnormal activity detection in the proposed algorithm. In Fig. 2(a), the dotted lines denote the dense trajectories belonging to one object (encircled in the tube) across the scene.

To describe the motion of an object, the multi-scale histogram of optical flow (MHOF) is adopted as the feature descriptor[17], as shown in Fig. 2(b). The noise motion is firstly filtered with extremely large amplitude. MHOF has $K = 64$ bins including four scales, for more precisely preserve motion direction information and motion energy information. The first scale uses the first 16 bins to denote 16 directions with motion magnitude $r \leqslant T_1$, the second scale uses the next 16 bins with motion magnitude $T_1 \leqslant r \leqslant T_2$, the third scale uses the next 16 with motion magnitude $T_2 \leqslant r \leqslant T_3$, and the fourth scale uses the final 16 with motion magnitude $T_3 \leqslant r$.

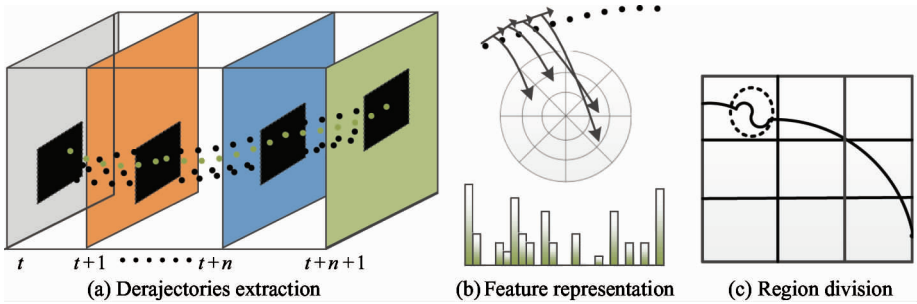

Fig. 2 The proposedabnormal activity detection algorithm

2.2.2 Dictionary learning

Because the local trajectory may be abnormal, while global trajectory is normal, as shown in Fig. 2(c), the video is splited into several sub-regions and extract MHOF features in each region to train dictionary[23,24] for sparsely representing each feature. Given a training set of feature pool as $\boldsymbol{B} = [b_1, b_2, \cdots, b_k] \in \boldsymbol{R}^{m \times N}$, where each column vector $\boldsymbol{b}_i \in \boldsymbol{R}^m$ denotes a feature vector with m-dimension and N denotes the total number of feature vectors. To find a set of basis $\boldsymbol{D}$ (it can be initialized by k-means) and a matrix of mixing coefficients $\boldsymbol{W}$, $\boldsymbol{B}$ can be reconstructed by the weighted sum of computed basis $\boldsymbol{D}$ well. More formally, this problem can be formulated as

$$\min_{D,W} \| \boldsymbol{B} - \boldsymbol{DW} \|_F^2 + \boldsymbol{\lambda} \| \boldsymbol{W} \|_{2,1} \quad (1)$$

The efficient sparse coding algorithm is utilized as in Ref. [23]. The objective function is not convex in terms of all the variables jointly. Therefore, it is unrealistic to expect an algorithm to easily find the global optimal solution. An alternating optimization method is adopted to solve it.

2.2.3 Sparse reconstruction cost

With the dictionary dat hand, a test sample y can be classified as normal activity or not. As mentioned above, the feature of a normal sample can be constructed by only a few number of bases in the dictionary $\boldsymbol{D}$, while an abnormal sample cannot. So, the sparse representation problem can be formulated as

$$\boldsymbol{w}^* = \min_w \| y - \boldsymbol{D}w \|_F^2 + \boldsymbol{\lambda} \| w \|_1 \quad (2)$$

This can be solved by the gradient based method described in Ref. [23]. The $l_{2,1}$ -norm is adopted for W during dictionary learning. Here, l_1 -norm is adopted for $\boldsymbol{w}$. The $l_{2,1}$ -norm is a general version of the l_1 -norm in nature. Since if $\boldsymbol{w}$ is a one dimension vector, then $\| \boldsymbol{w} \|_{2,1} = \| \boldsymbol{w} \|_1$. After that, the optimal reconstruction weight vector $\boldsymbol{w}^*$ is got, the sparsity reconstruction cost (SRC)[17] can be computed as

$$S = \| \boldsymbol{y} - \boldsymbol{Dw}^* \|_F^2 + \boldsymbol{\lambda} \| \boldsymbol{w}^* \|_1 \quad (3)$$

And the test sample $\boldsymbol{y}$ will be detected as an abnormal activity, if the following criterion is satisfied:

$$S > \varepsilon \quad (4)$$

where ε is a parameter that is set by cross-validation. It determines the sensitivity of classifying abnormal activity.

2.3 Key observation selection

Video synopsis method[1] provides an effective way for fast browsing activities by spatial-temporal rearranging them into a greatly condensed video. In typical scenarios, the activities always consist of numerous observations, resulting in collision and degradation of

subjective effect in synopsis video. In addition, adjacent observations may be very similar in action and appearance. In light of these factors, a few representative observations are used, namely key observations, to depict the original behavior of normal behavior, which can greatly eliminate the redundancy in the content domain, and promote the efficiency of video synopsis. In Ref. [11], k-means clustering method was adopted to select a pre-defined number of key actions. However, the number of key actions can not be fixed for different objects, even in the same scenario.

Different from Ref. [25], the key observations is extracted from every object instead of input video. The observations which have significant action are selected as the key ones, according to the proposed criteria. In the proposed algorithm, the difference between the sampled activities and the original activities (representativeness) is tried, while to is used minimized as small number of observations as possible (compressibility), as shown in Fig. 3 (the points denote key observations, and the dotted line denotes the sampled trajectory). However, the representativeness and the compressibility are contradictory to each other. For example, if all the observations of the trajectory are chosen as the key ones, then the represent ativeness is maximized. In contrast, if only the starting and ending observations of the trajectory are chosen as the key ones, the representativeness is minimized, but the compressibility is maximized. Take the representativeness and compressibility into consideration, a data-driven method is adopted to select key observations by transforming it into an MDL optimization problem[26] for exploring an optimal selection.

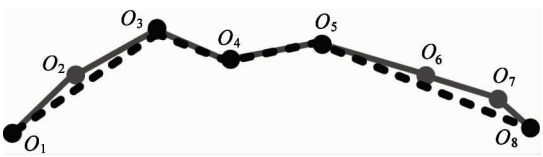

Fig. 3 An example of key observation selection

The description length (DL) in the proposed algorithm is computed as: $L(D,S) = L(D \mid S) + L(S)$, where S is the learned key observation selection solution, and D is the input trajectory (consisted of observations). $L(D \mid S)$ is the number of bits required for encoding the data with the help of the key observation selection, while $L(S)$ stands for that to encode the selection solution. The optimal key observation selection solution S is the one that minimizes $L(D, S)$ (MDL), namely MDL. $L(S)$ and $L(D \mid S)$ is computed as follows:

$$L(S) = -\log\frac{N_k}{N} \tag{5}$$

$$L(D \mid S) = -\log\frac{\sum_{i=1}^{N_{k-1}}\left(\sum_{j=co_{i+1}^{l}}^{co_{i}^{l}-1} len(O_jO_{j+1}) - len(O_i^lO_{i+1}^l)\right)}{R_{\max}} \tag{6}$$

where $R_{\max} = \sum_{i=1}^{N-1} len(O_iO_{i+1}) - len(O_1O_N)$, N_k is the number of key observations, N is the total number of observations belonging to the trajectory, O_i denotes the ith observation, O'_i denotes the ith selected key observation. The distance function *len* denotes the length of a line segment of two observations, for considering the speed factor. $R_{\max}$ denotes the minimum representativeness, maximum different the sampled trajectory and original trajectory, when only start and end observations are selected as key ones. $R_{\max}$ is used to normalize representativeness between 0 to 1. Fig. 3 is an example to demonstrate the function of key observation selection. Then, $L(S) = -\log(5/8)$ can be got by Eq. (5), and $L(D \mid S) = -\log((len(O_1O_2) + len(O_2O_3) - len(O_1O_2) + len(O_5O_6) + len(O_6O_7) + len(O_7O_8) - len(O_5O_8))/R_{\max})$ by Eq. (6).

As mentioned above, it is needed to search the optimal key observation selection scheme that minimizes the DL. However, it is an NP-hard problem. Therefore, an approximate method is adopted by choosing a local optimum. As shown in Fig. 4, if $DL(O_kO_{k+3} < DL(O_kO_{k+4})$, then O_{k+3} is deemed as the proceeded key observation, proceed with the former key observation Ok. The detailed algorithm is summarized in Algorithm 1.

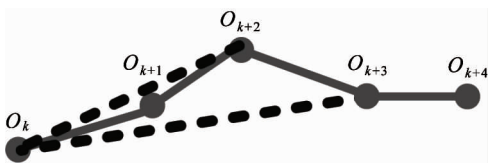

Fig. 4 An example of the approximate algorithm

2.4 Synopsis video generation

Video synopsis can be seen as an energy minimization optimization problem, and the energy includes the cost of objects collision, the cost of objects time inconsistency, the cost of objects lost, and so on. Following Ref. [1], the concepts of collision E_c, time consistency cost E_t, and compression rate cost E_l are introduced for generating lossless synopsis video in the proposed algorithm. The synopsis of abnormal activities and normal activities are conducted respectively using the same energy function, generating two videos for fast

browsing. The energy function can be formulated as follows:

$$E = \operatorname{argmin}_{\forall M} E(M) \tag{7}$$

$$E(M) = \sum_{b_n, b'_n \in \boldsymbol{B}} (\alpha E_t(b_n, b'_n) + \beta E_c(b_n, b'_n) + E_l(b_n, b'_n)) \tag{8}$$

where $\boldsymbol{B}$ is the whole tube set, b_n and b'_n are two tubes mapped into synopsis video. $E_t(b_n, b'_n)$ is the time inconsistency cost, for preserving the chronological order of objects. $E_c(b_n, b'_n)$ is the collision cost, penalizing for the spatial-temporal overlaps among objects. $E_l(b_n, b'_n)$ is the compression rate cost, penalizing for the long synopsis video. α and β are two empirical parameters set by the user according to their relative importance. Reducing the weights of the collision cost, e.g., will result in a dense video where objects may overlap. Increasing this weight will result in a sparse video where objects do not overlap and less activity is presented. The minimization of energy function is addressed by simulated annealing. After achieving the optimal arrangement of tube set, the tubes are stitched into the background image using poisson editing to generate final synopsis video.

Algorithm 1: Selecting key observations in trajectory based on MDL.

Input: N, the number of observations
Output: trajectory O' consist of key observations
Data: $\{O_s, O_{s+1}, \cdots, O_e\}$
$O'_1 = O_s$;
$p = 2$;
for {k = 1; k < N − 1; k + +} **do**
  $j = 1$
for {;;} **do**
  if DL($O_k O_{k+j}$ < DL($O_k O_{k+j+1}$)) **then**
    $O'_p = t_{k+1}$, p ++;
    continue;
  else
    j ++;
  end if
 end for
 $k = k + j - 1$;
end for

3 Experimental evaluation

In order to evaluate the performance of the proposed abnormal activity detection based surveillance video synopsis method, experiments are conducted on three real world testing videos, captured by the equipment in outdoor scenes. The first dataset (D1) targets at pedestrian activity, which consists of 31,530 frames, and the representative images are shown in Fig. 5(a). The second dataset (D2) targets at vehicle surveillance of street scenario, which consists of 43,685 frames, and the representative images are shown in Fig. 5(b). The third dataset (D3) targets at vehicle surveillance of street scenario, which consists of 39,556 frames, and the representative images are shown in Fig. 5(c). All video are resized to resolution 352 × 288, 15 FPS. The first 8minutes are selected for training dictionary.

(a) D1 (b) D2 (c) D3

Fig. 5 The representative images

To demonstrate the benefit of key observation selection in video synopsis, the proposed synopsis method (denoted as Proposed) is compared with traditional method without abnormal activity detection and key observation selection[1] (denoted as Method 1), cluster-based synopsis[2] (denoted as Method 2), and another key observation selection based synopsis method without abnormal detection[27] (denoted as Method 3). The detailed results are displayed in Table 1, Table 2 and Table 3, corresponding to D1, D2 and D3 respectively. From Table 1, it can be concluded that our method obtains 4.7% compression rate, while causing 3,784 energy loss. Method 1 obtains 7.4% compression rate, while causing 5,008 energy loss. Method 2 obtains 6.7% compression rate, while causing 4,445 energy loss. And Method 3 obtains 5.9% compression rate, while causing 4,281 energy loss. From Table 2, it can be concluded that the proposed method obtains 2.3% compression rate, while causing 2,841 energy loss. Method 1 obtains 7.2% compression rate, while causing 3,641 energy loss. Method 2 obtains 6.8% compression rate, while causing 3,577 energy loss. And Method 3 obtains 6.0% compression rate, while causing 3,302 energy loss. From Table 3, it can be conclude that the proposed method obtains 11.5% compression rate, while causing 5,682 energy loss. Method 1 obtains 15.3% compression rate, while causing 7,752 energy loss. Method 2 obtains 14.9% compression rate, while causing 7,172 energy loss. And Method 3 obtains 13.8% compression rate, while causing 6,465 energy loss. The energy loss are mainly caused by object collision and chronological mis-order, which can heavily degrade the quality of synopsis video. Ob-

viously, the proposed method achieves lower energy loss, while preserving a high compression rate.

Table 1 Detailed lost information in synopsis for Dataset 1

Dataset 1	Energy lost		Frame number	
	E_t	E_c	Original	Synopsis
The proposed method	575	3,209	3,1530	1,476
Method 1	978	4,030	3,1530	2,328
Method 2	786	3,659	3,1530	2,134
Method 3	698	3,583	3,1530	1,876

Table 2 Detailed lost information in synopsis for Dataset 2

Dataset 2	Energy lost		Frame number	
	E_t	E_c	Original	Synopsis
The proposed method	378	2,463	43,685	2,267
Method 1	476	3,165	43,685	3,164
Method 2	501	3,076	43,685	2,987
Method 3	426	2,876	43,685	2,640

Table 3 Detailed lost information in synopsis for Dataset 3

Dataset 3	Energy lost		Frame number	
	E_t	E_c	Original	Synopsis
The proposed method	430	5,252	39,556	4,548
Method 1	487	7,265	39,556	6,041
Method 2	476	6,696	39,556	5,879
Method 3	445	6,020	39,556	5,465

4 Conclusion

In this work, a novel video synopsis is proposed based on abnormal activity detection and key observation selection. In the proposed algorithm, the activities are classified into normal and abnormal ones based on sparse coding framework. For normal activities, key observation selection using MDL principle is conducted for eliminating content redundancy. Experimental results on publicly available datasets demonstrate the effectiveness of the proposed approach.

References

[1] Pritch Y, Rav-Acha A l, Gutman A, et al. Webcam synopsis: Peeking around the world. In: Proceedings of the IEEE International Conference on Computer Vision, Janeiro, Brazil, 2007. 1-8

[2] Pritch Y, Ratovitch S, Hendel A, et al. Clustered synopsis of surveillance video. In: Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, 2009. 195-200

[3] Hanjalic A, Zhang H, Vecchi M P. An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 1999, 9(8): 1280-1289

[4] Smith M A, Kanade T. Video skimming and characterization through the combination of image and language under standing. In: Proceedings of the Content-BasedAccess of Image and Video Databases, Bombay, India, 1998. 61-70

[5] Li Y, Lee S, Yeh C, et al. Techniques for movie content analysis and skimming tutorial and overview on video abstraction techniques. *IEEE Signal Processing Magazine*, 2006, 23(2):79-89

[6] Huang C, Chung P, Yang D H, et al. Maximum a posteriori probability estimation for online surveillance video synopsis. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014, 24(8): 1417-1429

[7] Zhang X Y, Wang S. Bidirectional active learning, a two-way exploration into unlabeled and labeled dataset. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 28(12):3034-3044

[8] Zhang X Y. Interactive patent classification based on multi-classifierfusion and active learning. *Neurocomputing*, 2014, 127(1): 200-205

[9] Feng S, Lei Z, Yi D, et al. Online content-aware video condensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, USA, 2012. 2082-2087

[10] Rodriguez M. Cram:Compact representation of actions in movies. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010. 3328-3335

[11] Tian Z, Xue J, Lan X, et al. Key object-based static video summarization. In: Proceedings of the ACM Multimedia, Arizona, USA, 2011. 1301-1304

[12] Taj M, Maggio E, Cavallaro A. Multi-feature graph-based object tracking. In: Proceedings of the 1st International Evaluation Conference on Classification of Events, Activities and Relationships, Southampton, England, 2006. 190-199

[13] Su S J, Nian X H, Pan H. Trajectory tracking and formation control based on consensus in high-dimensional multi-agent systems. *High Technology Letters*, 2012, 18(3): 326-332

[14] Hu Z T, Fu C L. A novel multi-sensor multiple model particle filter with correlated noises for maneuvering target tracking. *High Technology Letters*, 2014, 20(4): 355-362

[15] Arulampalam S, Maskell S, Gordon N, et al. A tutorial on particlefilters for on line non-linear/non-gaussian-bayesian tracking. *IEEE Transactions on Signal Processing*, 2002, 50(2):174-188

[16] McKenna S, Jabri S, Duric Z, et al. Trackinggroups of people. *Computer Vision and Image Understanding*, 2000, 80(1):42-56

[17] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado, USA, 2011. 3449-3456

[18] Hu W, Xiao X, Fu Z, et al. A system forlearning statistical motion pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (*TPAMI*), 2006, 28(9): 1450-1464

[19] Meng F F, Qu Z S, Zeng Q S, et al. Video objects behavior analyzing based on motion history image. *High Technology Letters*, 2009, 15(3): 319-324

[20] Wang H, Klaser A, Schmid C, et al. Action recognition by dense trajectories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado, USA, 2011. 3169-3176

[21] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009. 935-942

[22] Sand P, Teller S. Particle video: Long-range motion estimation usingpoint trajectories. *International Journal of Computer Vision*, 2008, 80(1): 72-91

[23] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms. In: Proceedings of the Advances in Neural Information Processing Systems, Vancouver, Canada, 2007. 3169-3176

[24] Zhang C J, Liu J, Tian Q C, et al. Image classification bynon-negative sparse coding, low-rank and sparse decomposition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado, USA, 2011. 1673-1680

[25] Kim C, Wang J. An integrated scheme for object-based video abstraction. In: Proceedings of the ACM Multimedia, Los Angeles, USA, 2000. 303-311

[26] Lee J, Han J. Trajectory clustering: A partitionandgroup framework. In: Proceedings of the ACM Special Interest Group on Management of Data, Beijing, China, 2007. 593-604

[27] Zhu X, Liu J, Wang J, et al. Key observation selection-based effective video synopsis for camera network. *Multimedia Vision and Application*, 2013, 25(1):145-157

Zhu Xiaobin, born in 1982. He received his Ph. D in 2013 from Institute of Automation, Chinese Academy of Sciences. His research interests include machine learning, pattern recognition, video and image analysis, etc.