International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.1 (2016), pp.289-302
http://dx.doi.org/10.14257/ijsip.2016.9.1.28

Abnormal Crowd Motion Behaviour Detection based on SIFT Flow

Dongping Zhang, Kaihang Xu, Huailiang Peng, Ye Shen

College of Information Engineering China Jiliang University, Hangzhou, 310018, China,
silenttree_zju@cjlu.edu.cn

Abstract

This paper focuses on the detection of the abnormal motion behaviour recognition of the crowd, and proposes an innovation method which is consist of three steps, i.e. SIFT flow + weighted orientation histogram + Hidden Markov Model(HMM). Analogous to optical flow, which is used to get the motion information of the pixels from two adjacent frames, SIFT flow is of higher precision. Next, we build up a a weighted orientation histogram as a statistical measurement for the SIFT flow features from the first step. Finally, the derived histogram is taken as the input for HMM in preparation for the detection of abnormal crowd motion. Experimental results show that compared to the existing method, our proposed one can detect the abnormal motion behaviour more effectively.

Keywords*: SIFT flow, coarse to fine, dense scene correspondence, HMM, crowd motion behaviour detection, weighted orientation histogram.*

1. Introduction

In recent years, the crowd behaviour detection in intelligent video surveillance (IVS) recognition has gained popularity and has been proven effective for many application scenarios. Currently, the video surveillance system as important urban monitoring tools has been widely used in the public safety, financial security, transportation, and other areas [1]. Such as become a very active research focus, by computer visual analysis technology, automatic identification of the crowd scenes video surveillance and interactive behaviour. It has changed the traditional video surveillance undue reliance on artificial characteristics, to be able to automate the detection of target objects in complex scenes, and the target object behaviour analysis and a description of the true sense of the intelligent monitoring.

Several methods (e.g. [2], [3], [4]) that are used to detect abnormal behaviour in crowd scene have been proposed in recent research. The Existing crowd behaviour recognition methods can be divided into description-based method and statistical-based method. The description- based method can keep the behaviour of space and time structure. It treats movement of the crowd as a sub-movement which satisfies certain relations. Thus the recognition of the behaviour is completed with the search of sub-behaviour which satisfies the definition (e.g. In [5] it proposed an Event Recognition Language (ERL) scheme, in which they first divide the event into primitive events, single-line event, multi-line events and then combine the time-determine method, and finally ERL is built by setting the sentence structure, data type, and related parameters). In contrast, the statistical-based method is superior to the description-based counterpart in that it has been proven more suitable for behaviour detection even in the noisy case, as long as there are sufficient training data. In addition, the more complex the behaviour is, the more sequence data is needed, which makes the statistical-based method very easy to apply to the complex behaviour(e.g. In

ISSN: 2005-4254 IJSIP

[6] it regard the crowd behaviour as multi-line events. Using random finite state representation of the single event, the state can identify by the Bayesian analysis methods through the path of motion and the shape. However, this approach is over-dependent on low-level test results, and cannot avoid the errors of detection or identification for the low-level sub-behaviour.

In previous research, complicated motion feature representations (e.g.[9-12]) have been developed to deal with the variations of object shapes and appearances from two adjacent frames in the crowd scene sequences. Of which the optical flow [15] technology is applied to obtain the motion features for the abnormal crowd motion behaviour detection. However, the motion feature of optical flow contains poor information due to the low dimensions. In a recent study, the SIFT flow technology (e.g.[13,14]) has been proven superior to the optical flow and it is a beneficial technique for image alignment as well as prediction for the motion field. The motion feature obtained upon pixel-wise from SIFT flow has 128-dimensional vector instead of RGB or gradient in optical flow [15] can more effectively represent the motion information for each pixel and robust to lighting changes and large displacement.

Hence, taking advantage of the statistcal-based method and SIFT flow technology, we propose an innovative method for the abnormal crowd motion behaviour detection, which consist of three steps. i.e. SIFT flow + weighted orientation histogram [16] + Hidden Markov Model (HMM) [7, 8] . In the first step, SIFT flow is taken for obtaining the pixels' motion information (velocity and orientation) from the two adjacent frames in the video sequences; in the second step, based on the pixels' motion information, we quantize the orientation of a circle, i.e. The angle 2π into 12 levels, meanwhile we weigh these orientations in light of their corresponding velocities, to form the weighted orientation histogram as a statistical measurement for the SIFT flow; In the third step, we take the derived histogram as the input for the HMM model. In such a model, mixtures of Gaussians are assumed, and our task here is to estimate their parameters iteratively using EM algorithm[17]. For the test, these obtained optimal parameters are used directly for the abnormal crowd motion behaviour detection. The training the testing processes stated above are illustrated as the flowchart in Fig.1., in which the black arrows stand for training while the red arrows represent testing.

The remainder of this paper is organized as follows. In section 2, we describe a feature extraction method based on SIFT flow technology. Section 3 covers the details of hidden Markov model. Section 4 shows the experimental results and finally we conclude the presentation in section 5.

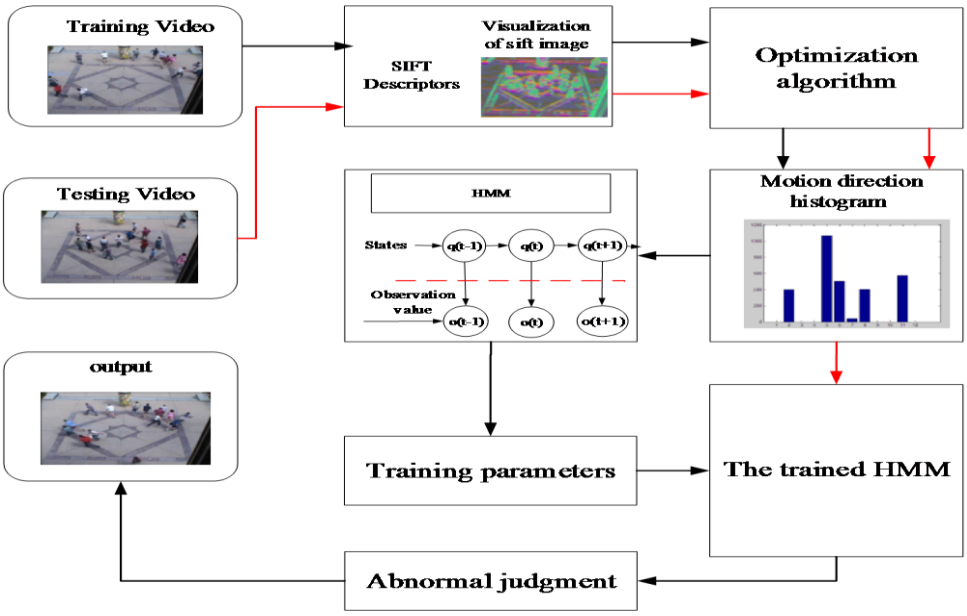


Figure 1. The Flowchart of the Proposed Method for Training (Black Arrows) and Testing (Red Arrows)

2. SIFT Flow Technology

In order to obtain meaningful correspondences from two adjacent frames in a video sequence, this study presents a novel method which combines motion flied $w(p) = (\mu(p), v(p))$ with SIFT features extracted from the pixel p.

2.1 Dense SIFT Descriptors

By calculating SIFT local feature descriptors of each pixel in the frame, we can obtain a series of robust background characterization from this frame, and ensure the consistency of the features of adjacent frames. Dense SIFT descriptor calculation process is divided into the following steps:

a) Derivative values and Derivative direction .

Each pixel in the grayscale can be calculated with the following formula for its derivative values l and derivatives direction θ.

$$l(x, y) = \sqrt{dx^2 + dy^2} \tag{1}$$

$$\theta(x, y) = \tan^{-1} \frac{dy}{dx} \tag{2}$$

Where, dx, dy are the derivatives direction of x, y, respectively, which can be calculated by Sobel gradient operator.

b) Pixel area histogram features.

SIFT is a descriptor to describe the neighborhood gradient information. For each pixel in an image, we select its $b*b$ neighbor pixels as the local area, and calculate the derivative values $l(x, y)$ and derivatives direction $\theta(x, y)$ in the neighborhood. Then, we divide the neighborhood $b*b$ into a $c*c$ cell array. The weighted derivative histogram is calculated in each block by the size of $b/c * b/c$, quantizing

the orientation into α bins. The weighted value is given by Gaussian function. The generating principle of SIFT local feature descriptor is shown in Fig. 2.

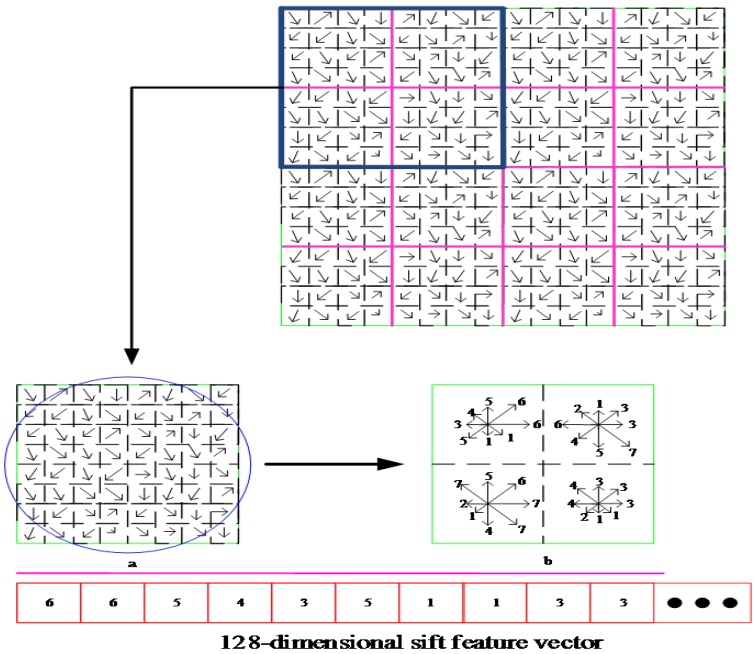


Figure 2. The Calculation Principle of SIFT Descriptor

Fig. 2 shows the generating principle of SFIT local feature descriptor when $a=8, b=8, c=2$, of which (a) shows the neighbor area gradient of the candidate pixels, (b) shows the image gradient of the $b*b$ neighbor area, (c) shows the direction histogram of the $c*c$ cell. In 8* 8 neighbor area of the image, we calculate the gradient value and gradient direction and through the weighted Gaussian window. The neighbor area of candidates pixels are then divided into 4×4 blocks (purple line), and in each block it computes one derivative direction histogram (quantitative for 8 directions). The following line below (b) and (c) is the 128 dimensional SIFT feature vector aligned by the 8-direction histogram of these 16 blocks (i.e. $8*16=128$).

2.2 SIFT Flow Flied

Inspired by optical flow [15], [13] introduces an objective function $E(w)$, which is used to build a dense correspondence from a pair of SIFT images (s_1, s_2). p is the every pixel in an image

$$
\begin{aligned}
E(w) = &\sum_{P} \min(\|s_1(p) - s_2(p + w(p))\|_1, t) && \text{(a)}\\
&+ \sum_{p} \eta(|\mu(p)| + |v(p)|) && (b)\\
&+ \sum_{(p,q)\in\varepsilon} \min(\alpha|\mu(p) - \mu(q)|, d) && \text{(c)}\\
&+ \min(\alpha|v(p) - v(q)|, d) && \text{(d)}
\end{aligned}
\tag{3}
$$

Eq.(3) contains 4 terms, of which the data term in (a) constrains the dense SIFT feature descriptors to be matched via the flow vector $w(p) = (\mu(p), v(p))$. The small displacement term in (b) constrains the flow vectors to be as small as possible, and the smoothness term in (c) and (d) constrains the adjacent pixels to have similar displacement. The parameter ε represents the 4-neighborhood system. To match

outliers and flow discontinuities, the truncated L1 norm is used in both the data term and the smoothness term with thresholds t and d, respectively(In this paper, t and d are chosen as 0.01 and 0.1 respectively).

2.3 Optimization of SIFT Flow Flied

A dual-layer loopy belief propagation has been used as the base algorithm to optimize the objective function by [13]. However, using the dual-layer loopy belief propagation directly has poor performance with regard to the dimension of the image. So in this paper, a coarse-to-fine SIFT flow matching method is used to build the correspondence from two adjacent frames, which significantly improved the matching efficiency.

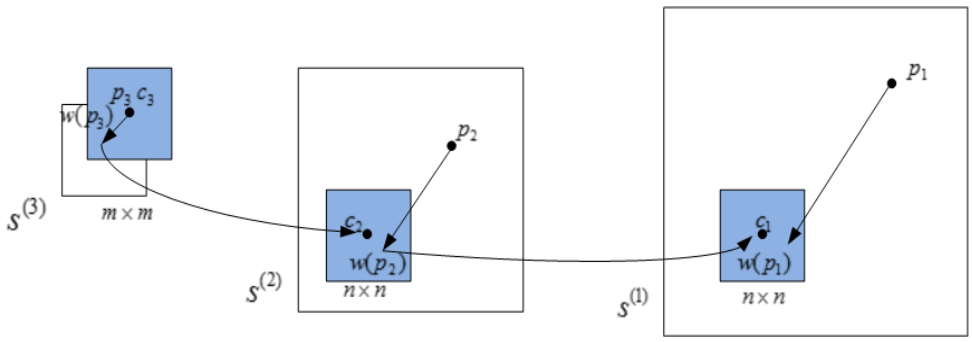


Figure 3. The process of coarse-to-fine SIFT flow matching method on a pyramid. The blue square is the searching window for p_k at each pyramid level k. For simplicity, only one image is shown here, where p_k is on image s_1, and c_k and $w(p_k)$ are on image s_2. See text for details.

The main idea of coarse-to-fine SIFT flow matching method is to roughly estimate the flow at a coarse level of image grid, then gradually propagate and refine the flow from coarse to fine [13]. The procedure is illustrated in Fig. 3. In the following paragraphs, s_1 and s_2 are both represented by s. A SIFT pyramid $\{s^{(k)}\}$ is established before matching, $s^{(k)}$ is obtained by smoothed and down-sampled from $s^{(k+1)}$. At the level k of pyramid, p_k is the coordinate of the pixel to match. c_k is the offset or centroid of the searching window, and $w(p_k)$ is the best match. At the top pyramid level $s^{(3)}$, p_3 is the centroid of the searching window with size $m \times m$. Then BP converges is used at this level. After BP converges, we only need to search the position corresponding to flow vector $w(p_3)$ when we search at the next (finer) level. Once the flow vector $w(p_1)$ is estimated, the procedure iterates from $s^{(3)}$ to $s^{(1)}$ stops.

2.4 Weighted Orientation Histogram

In this paper, the weighted orientation histogram [16] is used as an observation vector, which is built from the motion vectors of each frame. The observation vector has 12 dimensions on behalf of 12 directions. We get the angle of every motion vectors using Eq. (4), i.e.

$$angle = \frac{n\pi}{2} + arctan\left(\frac{u}{v}\right) \tag{4}$$

Here μ, ν are obtained from the previous subsections. Then we divide the angle 2π into 12 levels, and each level has an increase of $\pi/6$. The direction probability distribution of each frame can be directly estimated from the direction. We compute

the angle of every motion vector of a certain frame and then put the total number of motion vectors with the same change in the angle range into the corresponding dimension of the observation vector. Considering the magnitude of the vector, we weigh the orientation histogram, namely, set a threshold value θ_s, if the magnitude θ is between $(e-1)*\theta_s$ and $e*\theta_s$, the angle number of the vector is set to e. For example, θ_s is set to 0.1, if the magnitude θ=0.56, the angle number of the vector is set to 6. The observation vector of a frame is obtained using this method, and the same processing is performed to all frames which we will train and test. The calculated observation vector indicates the direction distribution and direction tendency of the frame. To have an intuitive representation of observation vector we draw a weighted orientation histogram of the vector field. The eight weighted orientation histograms are displayed in Fig. 4.

3. Hidden Markov Model

HMM (hidden Markov model) is a kind of double embedded stochastic process which is composed of two stochastic processes: One is the visible output observation sequence, and another is invisible state transition sequence that can be inferred by the output observation sequence.

In order to express the spatio-temporal variation of SIFT flow, an HMM combined with mixture of Gaussians (MOGHMM) is utilized. A video sequence can be modeled with states and observation values by the HMM show as Fig.5.

We define the HMM parameter set $\lambda = \{\pi, A, B\}$ with N states and with each state of M Gaussians. The observation sequence of length T is defined as $O = O_1, O_2, ..., O_T$. In our case, each observation corresponding to one frame in the sequence is a 12-dimensional vector, i.e. the $t-th$ frame is $O_T = \{d_1, d_2 \quad ,\ldots, d_{12}\}$. The value of each dimension is the total number of pixels whose motion directions are in its corresponding fixed interval (as described in the previous section). The model parameters determined by the Expectation-Maximization (EM) algorithm are $\lambda = (\pi_i, a_{ij}, c_{im}, \mu_{im}, \quad , \mu_{im}, U_{im})$, where π_i is the prior probability for state $i(i = 1,\ldots, N)$, a_{ij} is the state transition matrix $(i = 1,\ldots, N; j = 1,\ldots, N)$, c_{im} is the mixture coefficient, μ_{im} is the mean vector or and U_{im} is the full covariance matrix for Gaussian $m(m = 1,\ldots, M)$ in state i.

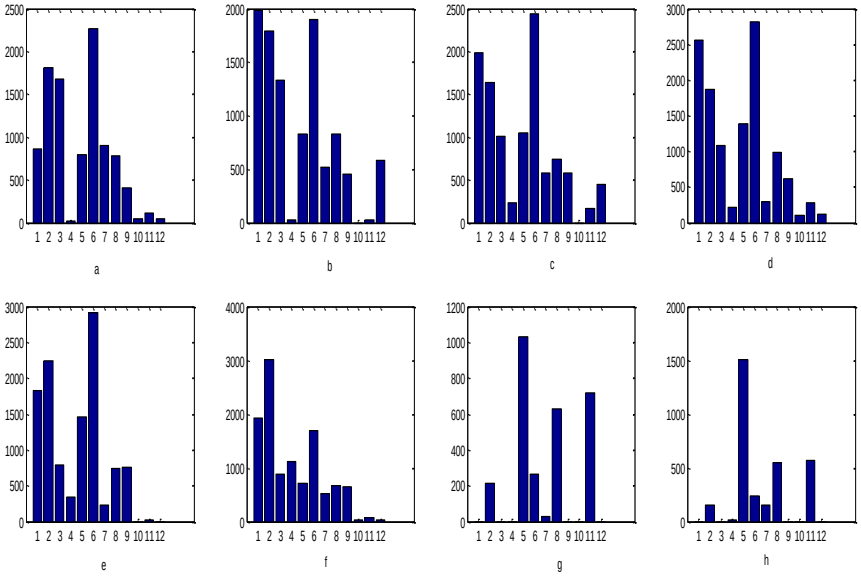


Figure 4. The distribution histograms of weighted orientation for the pixels in different frames. (a) ~ (f) show the distribution in the frame that a group of people walking around on the lawn normally, and (g) ~ (h) show the performance of the frame that the crowd suddenly spread around due to panic.

The probability distribution function of a vector O generated from state j is a finite mixture of the form

$$b_j(O) = \sum_{m=1}^{M} c_{jm} \chi(O, \mu_{jm} | U_{jm}),\ 1 \le j \le N \tag{5}$$

Where χ is a Gaussian density, with mean vector μ_{jm} and covariance matrix U_{jm}.

$\gamma_t(i,k)$ is the probability of being in state i at time t with the k th mixture component accounting for O_t, i.e.

$$\gamma_t(\mathrm{j},k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)} \right] \left[\frac{c_{jk}\chi(O_t, \mu_{jk}, U_{jk})}{\sum_{m=1}^{M} c_{jm}\chi(O_t, \mu_{jm}, U_{jm})} \right] \tag{6}$$

Where α and β are the forward and backward variables.

The update equations for the EM procedure [7] are:

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j,k)} \tag{7}$$

$$\overline{\mu}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot O_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{8}$$

$$\overline{U}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot (O_t - \mu_{jk})(O_t - \mu_{jk})'}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{9}$$

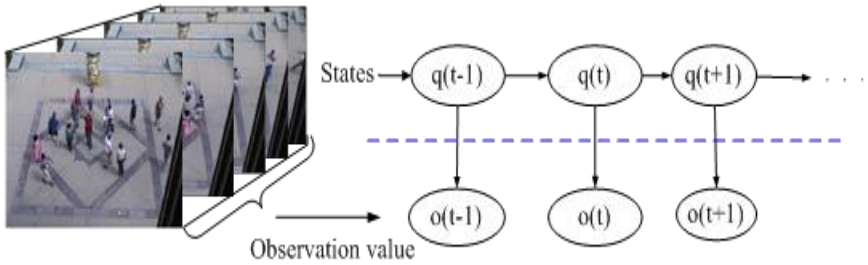


Figure 5. A video Sequence Modeled by HMM

4. Experimental Result

In order to evaluate the efficiency of our proposed method, we do experiments on two video sequence datasets, i.e. the UMN dataset [18] and the videos shot by ourselves using Sony HVR-V1C camera. Both of them include some panic events as the abnormal case in the crowd scene(e.g. the crowd suddenly changes their motion state from walk to scatter in different directions). The length of these two video datasets is taken by 1000 frames, with the frame size 320*240 and 360*240, respectively. We divide each video into 100 segments with 10 consecutive frames for each segment. So the HMM defined in Section 3 is trained with 100 observation sequences of T=10 vectors (each of the vectors has 12 dimensions).

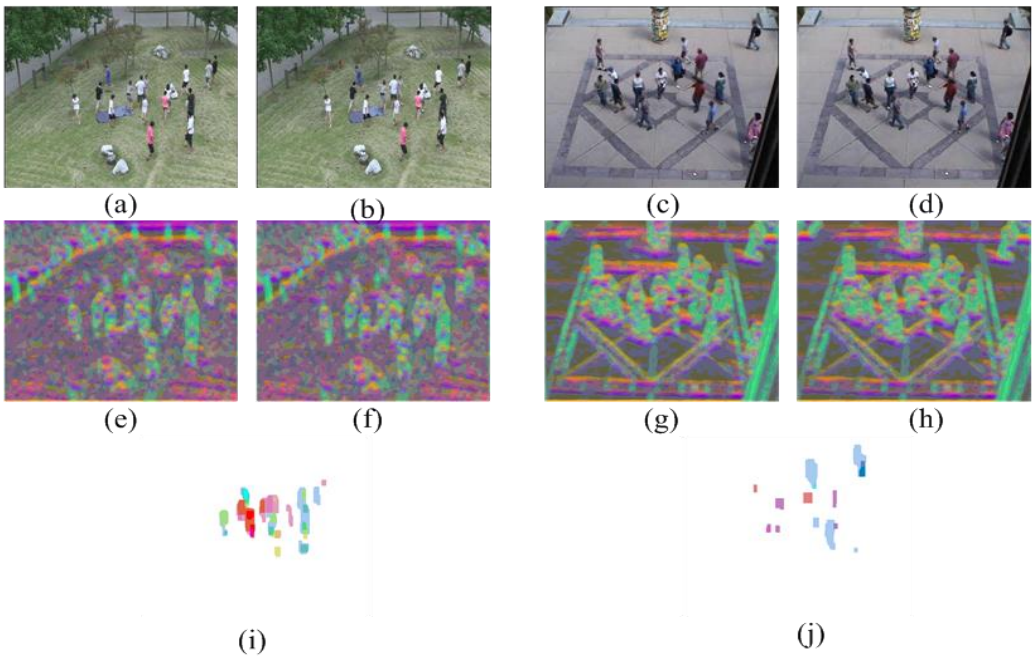


Fig. 6. Visualization of SIFT image and SIFT flow field. (a), (b) and (c), (d) are the two adjacent frames from the video sequence of OUR dataset and UMN dataset respectively. (e), (f) and (g), (h) are the two SIFT images correspond to (a), (b) and (c), (d); (i) and (j) are the SIFT flow field of these two adjacent frames(i.e.(e), (f) and (g), (h)), respectively.

4.1 SIFT Flow Processing

The purpose of this step is obtaining the SIFT flow features. In order to visualize SIFT images, the top three principal components of SIFT descriptors are computed and mapped to the principal components of the RGB space. As shown in Fig.6, (a), (b) and (c), (d) are the original two adjacent frames from the video sequence of OUR dataset and UMN dataset, respectively; We put the 128-dimensions SIFT descriptor to a 3D subspace, and visualize the SIFT image in Fig. 6 (e) , (f) and (g), (h). Via a coarse-to-fine SIFT flow matching method, we are able to obtain the field of the two adjacent frames for both of the two datasets, as shown in Fig. 6 (i) and (j), respectively.

In this coarse-to-fine process, it implies that those pixels that have similar color may share similar local image structures, please see Fig. 6 (i) and (j). Note that such process is only for visualization. In the SIFT flow, however, the entire 128 dimensions of each pixel is used for matching.

Now we conduct the pixel-wise SIFT descriptors for every two adjacent frame, as shown in Fig. 7. Our next task is to build dense correspondence to match these descriptors, obtaining the SIFT flow feature for further process.

4.2 HMM Processing

The input of HMM is a numerical sequence instead of a single value, because the analysis of HMM is a dynamic process [16]. So it is needed to combine the observed value of symbols into a sequence of observations.

The steps of crowd abnormal behaviour modeling with HMM are as follows:

(1) Evaluation questions: After obtaining the observation sequence $O = O_1, O_2, ..., O_T$ and parameters $\lambda = \{\pi, A, B\}$, we utilize the forward-backward algorithms to calculate events happening probability $P(O/\lambda)$ of the observation sequence under this model.

(2) Decoding problem: Using the Viterbi algorithm, the corresponding hidden state sequence $S = \{q_1, q_2, \cdots q_t\}$ is selected and it explains the observation sequence O reasonably, aiming to reveal the hidden part of the model through finding the optimal state sequence under the optimization criteria.

(3) Learning problems: Using the Baum-Welch algorithm which is one of the typical EM algorithms, we adjust the parameters $\lambda = \{\pi, A, B\}$ such that $P(O/\lambda)$ achieves the max value.

Crowd abnormal behaviour identification task is to analyze and establish the HMM through the image sequence, and conduct training and recognition.

4.3 Abnormal Behaviour Detection

As explained in details above, since each HMM is built based on one observation sequence in the training set, we must combine them to determine the abnormal behaviour by a single score of HMM for each observation sequence. We perform the experiment on the observation's likelihood, to determine in which HMM the abnormal event happens. The observation from HMM built by the test video segment is labeled abnormal if:

$$\log(P(O|\lambda)) < Th \tag{10}$$

The detection threshold Th is determined by the minimum value of the log-likelihood that appears in the normal training set. In Fig.8, the anomaly is detected when the log-likelihood sharply drops less than Th =-537 and Th =-583 in the two datasets respectively.

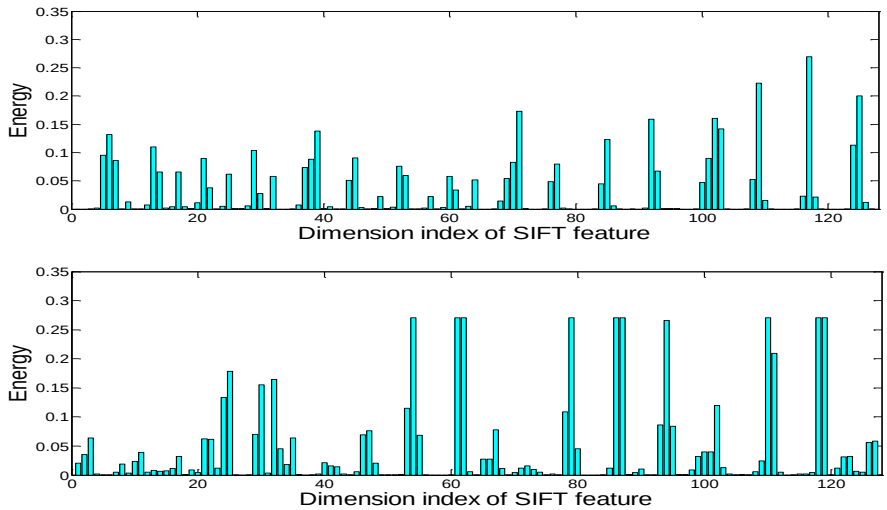


Figure 7. 128-dimension SIFT Features from One of the Pixels in Different Frame

The number of states and Gaussians are selected empirically by an HMM structure with a better likelihood of the training set among different configuration of number of states and Gaussians per state. Finally, 4 states and 3 Gaussians are used in the HMM in this paper, and the HMM parameters are received by the Baum–Welch algorithm after several iterations. The log-likelihood curves of both datasets are presented in Fig.8, where the red and blue curves correspond to the normal behaviour segment and abnormal behaviour segment, respectively. The log-likelihood curve has a clear and quick drop when the scattering event happened, and the long lasting log-likelihood drops when the abnormal events happens. Here, we choose the minimum value of log-likelihood that appears in the normal training set as the detection

threshold. Scattering event happens in the 58-th segment in the UMN test video. Another scenario on campus in OUR dataset is tested by the same procedure, and anomaly happens in the 45-th segment.

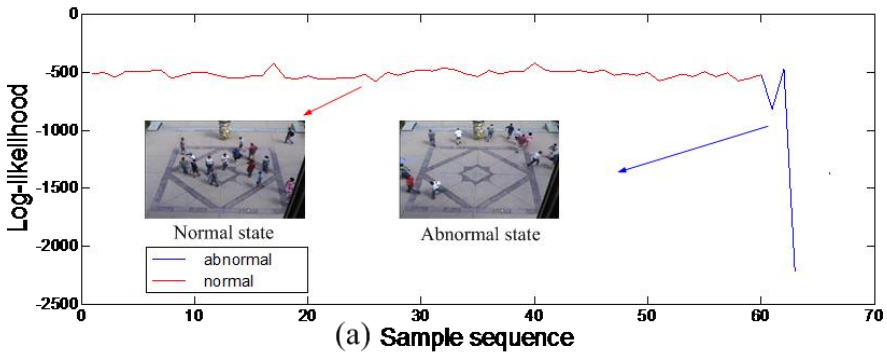
Our method is tested on these two datasets, too. To demonstrate the efficiency of our method, we re-implement the optical flow method [19] for comparison. Table 1 shows the performance results of our proposed algorithm and the optical flow method, and we can see our method achieves a higher accuracy rate.

Method	accuracy rate	
	UMN dataset	OUR dataset
Optical flow	84.2%	81.3%
Our method	89.3%	83.5%

Table 1. Performance Comparison of Different Method on the UMN Dataset and OUR Dataset

5. Conclusions

In this paper, we propose a new method for crowd motion behaviour detection and it could achieve better performance than the traditional optical flow method. In our method, we utilize a novel 3-step method for crowd motion behaviour detection task. First ,the SIFT flow technology is utilized for detecting the motion information from the two adjacent frames in the video sequences, then the weighted orientation histogram which plays the role of a statistical measurement for the SIFT flow is taken as the input for the HMM model. After training the assumed mixtures of Gaussians in the HMM by EM iterative algorithm, the optimal parameters can be directly used for test. The experiments on two different scenarios show that our method is able to detect the abnormal crowd motion more effectively. In future work, we will divide the image into blocks to capture small local motions and involve more motion characteristics such as velocity and acceleration in the abnormality detection of a crowd scene.



(a) Sample sequence

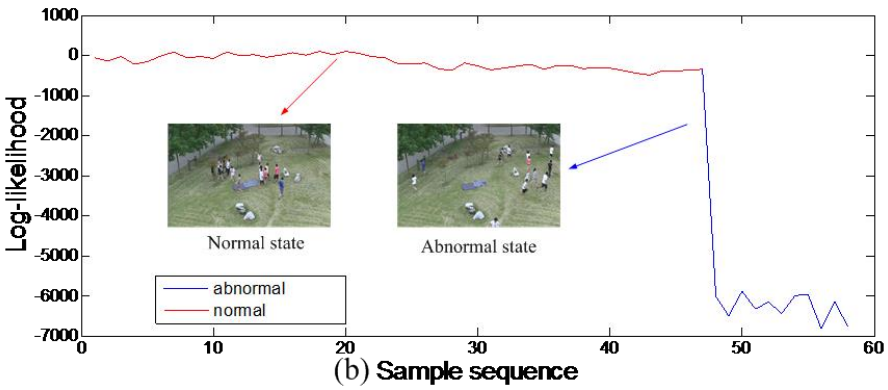


Figure 8. UMN Dataset (a) and OUR Dataset (b). Log-likelihood Results with 4 States and 3 Gaussians

Acknowledgments

This work was supported by Zhejiang Provincial Science & Technology Research Program (Grant No.2013C33046，2014C33075) and Zhejiang Provincial NSF (Grant No.LY15F020021, LY13H180011).

References

[1] McPhail C, Wohlstein R T. Using film to analyze pedestrian behaviour, J. Sociological Methods & Research, (1982), vol.10, no.3, p. 347-375.
[2] Wang Y, Mori G. human action recognition by semi-latent topic models, J. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2009), vol.31, no.10, p. 1762-1764.
[3] Monti F, Regazzoni C S. Human action recognition using the motion of interest points. Proceedings of International Conference on Image processing, (2010), p. 709-712.
[4] Mehran, R., A. Oyama, M. Shah. Abnormal crowd behaviour detection using social force model. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2009), p. 935-942.
[5] Nevatia R, Zhao T, Hongeng S. Hierarchical language-based representation of events in video streams. Proceedings of IEEE Workshop on Event Mining, (2003), vol.4, p. 39-46.
[6] Hongeng S, Nevatia R, Bremond F. Video based event recognition: activity representation and probablisitic recognition methods, J. Computer Vision and Image Understanding, (2004), vol. 96, p. 129-162.
[7] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition, J. Proceedings of the IEEE, (1989), vol. 77, no. 2, p.257-286.
[8] M. Piccardi, O. Perez. Hidden markov models with kernel density estimation of emission probabilities and their use in activity recognition. IEEE Conference on Computer Vision and Pattern Recognition, (2007), p. 1-8.
[9] S. Belongie, J. Malik, and J. Puzicha, Shape Context: A New Descriptor for Shape Matching and Object Recognition. Proc. Conf. Advances in Neural Information Processing Systems, (2000),p.831-837
[10] Berg A., T. Berg, J. Malik, Shape Matching and Object Recognition Using Low Distortion Correspondence. Proc. IEEE Conf. Computer Vision and Pattern Recognition, (2005), vol.1, p.26-33.
[11] P. Felzenszwalb, D. Huttenlocher. Pictorial Structures for Object Recognition, J. Computer Vision, (2005), vol. 61, no. 1, p. 55-79.
[12] R. Szeliski, Image Alignment and Stitching: A Tutorial. Foundations and Trends in Computer Graphics and Computer Vision, (2006), vol. 2, no. 1, p. 1-10.
[13] C. Liu, J. Yuen, Antonio Torralba. SIFT flow: dense correspondence across difference scenes, J. IEEE PAMI, (2011), vol. 33, no.5, p. 978-994.
[14] Mahamed Dahmane , Jean Meunier. SIFT-flow registration for facial expression analysis using Gabor wavelets. International Conference on Information Science, Signal Processing and their Applications, (2012),p. 175-180
[15] T. Brox, A. Bruhn, N. Papenberg, J. Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Warping. Proc. European Conf. Computer Vision, (2004), p. 25-36.
[16] Dongping Zhang, Yafei Lu, Xinghao Jiang, Huailiang Peng, Abnormal Crowd Motion Detection with Hidden Markov Model, J. IJACT: International Journal of Advancements in Computing Technology, (2013), Vol. 5, No. 8, p. 889 -897.
[17] T K Moon. The expectation-maximization algorithm, J. Signal processing magazine, IEEE, (1996), vol.13.no.6, p. 47-60.

[18] http://mha.cs.umn.edu/movies/crowdactivity-all.avi.
[19] N Li, Z Zhang. Abnormal Crowd Behaviour Detection Using Topological Methods. Software Engineering, Artificial Intelligence, Networking and Parallel Distributed Computing (SNPD), 12th ACIS International Conference on. IEEE, (2011): 13-18.

Authors

Dongping Zhang, He was born in 1970. He received the Ph.D in Information & Communication Engineering from Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2006. Since 2006, He is an associate professor at College of Information Engineering, China Jiliang University in Hangzhou. His research interests include image processing and pattern recognition, computer vision and video surveillance.

KaiHang Xu, He was born in 1991, and graduated at the China Jiliang University in Hangzhou in 2013. Currently he is a MSc. student at the College of Information Engineering of China Jiliang University. His research interests are oriented to image processing and image compression for security and multimedia applied imaging systems.

HuaiLiang Peng, He was born in1988, and graduated at the China Jiliang University in Hangzhou in 2011. Currently he is a MSc. student at the College of Information Engineering of China Jiliang University. His research interests are oriented to image processing and image compression for security and multimedia applied imaging systems.

Ye Shen, He was born in Changzhou, China, on September 7, 1978. He received the PhD degree in Biomedical Engineering from Zhejiang University, China, in 2014. He is currently an Associate Professor at the Department of Signal and Information Processing, China Jiliang University , China. His current reseach interests include intelligen signal analysis and processing, machine learning.

