Fighting Detection Based on Optical Flow Context Histogram

Yan CHEN[1,3] Ling ZHANG[2]
[1]Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China

Biyi LIN[3] Yong XU[1,*]
[2] Department of Biomedical Engineering, Zhejiang University, Hangzhou, China

Xiaobo REN[3]
[3]Shenzhen Sunwin Intelligent Corp. Shenzhen ,China

***Abstract*—This paper proposes a new feature, optical flow context histogram (OFCH) for detecting abnormal events, especially the fighting violence events from a live camera stream. The optical flow context histogram is a log-polar histogram system which combines the histogram of orientation and magnitude of optical flow together. The human action is represented by using the histogram sequence of orientation and magnitude of optical flow. PCA is adopted to reduce the dimension of the human action representation. Several machine learning methods, including random forest, support vector machine and Bayesnet are employed for sequence classification. The experiments were carried out on the video clips downloaded from the Internet. The results show that the proposed methods work well when using a fixed surveillance camera.**

Keywords-abnormal evernt detection; fighting detection; optical flow context histogram

I. INTRODUCTION

A large number of surveillance cameras have been installed in public areas, including shopping mall, rail stations, airport and so forth. Human monitoring all the video streams at the monitor center is impossible due to long time and tedious. An automatic detection on unusual event in real time is useful to prevent dangerous. This paper investigates fighting situations from surveillance camera. The aim of this work is to stop fighting violence situation in time and to prevent violence escalation.

Optical flow, describing coherent motion of moving objects [1], is a good feature for motion detection and tracking. It has been widely used for object tracking and motion segmentation.

In this research work, human actions are represented by the histogram sequence of orientation and magnitude of optical flow. We combined these two features of histograms into one histogram by using a log-polar system. The histogram of optical flow considers the magnitude and orientation together. No object tracking is needed for event detection and action recognition in our work.

The major contributions of this paper are listed as below.

1) The paper investigates the feasibility of identifying fighting situations. The ability to detect the violence breaking out is useful for stopping such event as soon as possible

2) The paper has proposed the a log-polar system to represent histogram of magnitude of optical flow and histogram of magnitude of optical flow. It combines the magnitude and orientation of optical flow in a single histogram representation.

The remaining parts of this paper are organized as follows. In section II, the related works are reviewed. In section III, the proposed fighting detection method is described. The experiments are illustrated in section IV. Section VI concludes this work and proposes some possible future work for this work.

II. RELATED WORKS

Adam's work [2] used either histogram of optical flow to represent an action, namely, histogram of orientation of optical flow or histogram of magnitude of optical flow. Their work may satisfy some unusual event, such as running in a shopping mall, anti-direction moving and so on. However, their work required to specify which kind of histogram was used for different applications, e.g. the histogram of magnitude of optical flow or the histogram of orientation of optical flow.

Chen and Aggarwal proposed a histogram of optical flow for recognizing actions from a far field of view in [3]. However, their work is similar with [2], which considered the orientation or magnitude of optical flow respectively. Another shortcoming with their work is that their work is based on the available human tracking.

Blunsden and Fisher proposed an approach to detect and classify pre-fighting by using a hierarchical Adaboost in [4]. They used Dallar's spatio-temporal features to represent an action sequence [5].

Cupillard et al. [6] investigated fighting within the domain of metro surveillance. They predefined the fighting template and classify a video sequence by matching to the predefined template. As we have known, fighting is an irregular action and there is no definition for fighting. Therefore, the template for fighting or non-fight was very hard to define.

Detta et al. [7] detected human violence by measuring the acceleration and jerk along the legs and arms orientations of a person. Their work required accurately foreground object segmentation and detection. For group violence, it is very hard to obtain a person's legs or arms correctly. Failure of

* corresponding author Prof. Yong XU, E-mail: laterfall2@yahoo.com.cn

978-0-7695-4606-3/11 $26.00 © 2011 IEEE
DOI 10.1109/IBICA.2011.28

IEEE computer society

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR. Downloaded on June 03,2020 at 16:27:26 UTC from IEEE Xplore. Restrictions apply.

accurate detection of legs and arms would lead to failure of detection of violence.

Chaudhry et al. proposed a method for action recognition by classifying Histogram of oriented optical flow[8]. They generalized Binet-Cauchy kernels to classify the histogram of optical flow.

Kim and Grauman [9] suggested a space-time Markov Random Field model to detect abnormal activities. They learned the activity patterns by capturing the distribution of optical flow with mixture probabilistic principal component analyzers.

III. PROPOSED APRROACH

A. *Feature Representation*

Using pure optical flow for action recognition is very difficult because the optical flow changes all the time. Histogram of oriented optical flow, reflecting a distribution of optical flows, can be used to for action recognition [8]. Furthermore, histogram of optical flow does not require foreground and background segmentation. Neither does it require object tracking.

A log-polar system is used for calculating the histogram of magnitude of optical flow and histogram of orientation of optical flow together. First, the optical flow feature is extracted from each frame. These feature points are distributed in the log-polar optical flow system based on the magnitude and orientation of optical flow together. Then, the histogram of optical flow is calculated based on the distribution of the optical flow points. Principle Component Analysis (PCA) can capture features, of the sample, that vary the most[10, 11], so it is adopted to identify the most important elements in the descriptor.

There are fixed number of optical flows calculated for a frame (400 optical feature points are calculated in our work). A log-polar coordination (r, θ) system is used for calculating the histogram of magnitude and oriented optical flow. The radius, r, represents the magnitude of optical flow, whose value ranges from 0 to M_{max}, the max magnitude of the optical flow. The magnitudes of optical flow are evenly divided into p magnitude sets. If the magnitude of optical flow is greater than the largest magnitude, then the largest magnitude value M_{max} is set for the magnitude of the optical flow. The angle, θ, valued from 0 to 360, represents the orientation of the optical flow. The orientation of optical flow is evenly divided into q sets. Fig. 1 shows an example of the histogram of optical flow (10 subsets for optical flow magnitude and 20 subsets for optical flow orientation). There are $p \times q$ sub-regions in total in the log-polar optical flow system (p subsets for optical flow magnitude and q subsets for optical flow orientation). An optical flow point must be located at one of the sub-regions of the log-polar optical flow system. For each sub-region ($r_1 < r <= r_2$ and $\theta_1 < \theta <= \theta_2$), the total number of optical flow located in this sub-region is computed as the histogram for this sub-region. Therefore, the of optical flow context histogram descriptor for a frame is represented by

$$\mathrm{F} = \left\{\mathrm{H}_{1-1}, \mathrm{H}_{1-2}, \dots, \mathrm{H}_{1-\mathrm{q}}, \mathrm{H}_{2-1}, \dots, \mathrm{H}_{i-j}, \dots, \mathrm{H}_{\mathrm{p}-\mathrm{q}}\right\} \quad (1),$$

where H_{i-j} represents the total optical flow point number in the sub-region whose optical flow magnitude ranges from $i \times \frac{M_{max}}{p}$ to $(i+1) \times \frac{M_{max}}{p}$ and optical flow orientation ranges from $j \times \frac{360}{pq}$ to $(j+1) \times \frac{360}{pq}$.

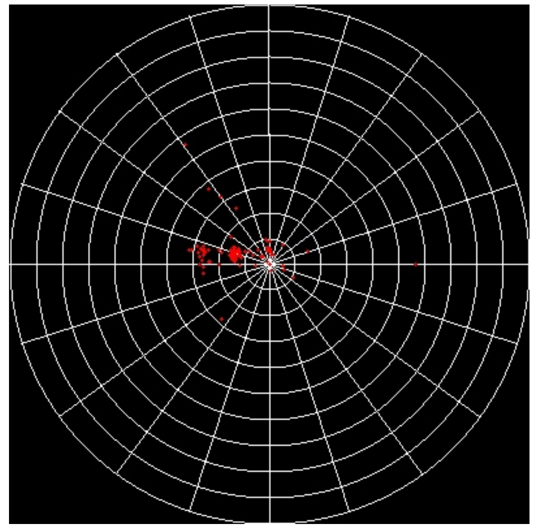

Figure.1: An example of optical flow context histogram working system for a frame.

The feature of an action sequence is represented by a sequence of optical flow context flow. It is denoted as

$$S = \{F_1, F_2, \dots, F_i, \dots, F_n\} \quad (2),$$

where F_i is the i^{th} frame of optical flow context histogram in the action sequence (represented by Eq.(1)) and n is the frame number for an action sequence. Figure 2 shows a histogram sequence of an non-fighting action sequence. Figure 3 shows a histogram sequence for a fighting action sequence. The distribution of optical flow in Fig 3 is more scatter than of in Fig. 2. The magnitude of optical flow in Fig.3 is larger than that of in Fig.2.

B. *Dimension reduction*

Supposing that the optical flow magnitude values are divided into 6 subsets and the optical flow orientation values are divided into 8 subsets, the histogram of optical flow of a frame is a 48-element vector. If a video speeds at 25frames per second, a three-second video has 48*25*3 elements. Training or recognizing on such large dimension of data is computational complex; therefore, dimension reduction is required to obtain more efficient features to represent the action sequence. Principle Component Analysis (PCA) [12] is adopted for feature dimension reduction. After computing the sequence of histogram of optical flow for action videos, PCA is used to find out the most important elements in the histogram of optical flow descriptor. According to our experiments, the first 5 elements in histogram of optical flow

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR. Downloaded on June 03,2020 at 16:27:26 UTC from IEEE Xplore. Restrictions apply.

selected by PCA represent 95% of the importance of the all histogram of optical flow elements. Therefore, only 5 elements projected by PCA are used for sequence training and classification.

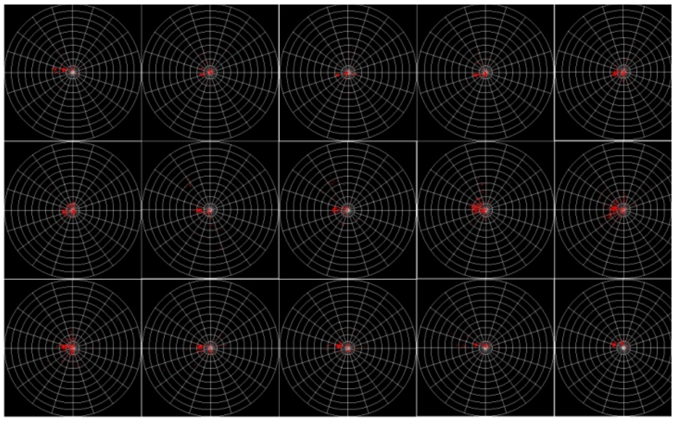

Figure 2: A histogram sequence of a non-fighting action sequence.

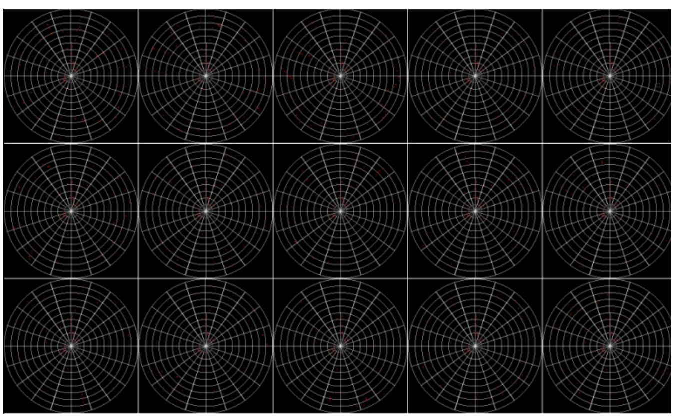

Figure 3: A histogram sequence of a fighting action sequence

C. System training and detection

Let us suppose that there are K fighting video clips and K non-fighting video clips were used for system training. At the system training stage, the sequence of the optical flow context histogram is computed for each action sequence. Then, PCA is used to reduce the dimension of the optical flow context histogram. The machine learning methods, including Support Vector Machine(SVM) [13], Bayesnet [14] and Random forests [15] are used in this research. In the training stage, the elements projected by PCA of the histogram sequence of each video clip are fed into the classifier, the classifiers learn from the training data , and generate vectors for determining a given action's category.

When given a probe action sequence, at first, its sequence of optical flow context histogram is calculated. Next, PCA is used to reduce the dimension of the histogram sequence. Finally, the reduced dimension of sequence of optical flow context histogram are fed into the classifier, the classifier determines the action sequence is a fighting or non-fighting action sequence based on the vector it generated at the training stage.

IV. EXPERIMENTS

Although there are some bench databases for abnormal events detection, however, to our best knowledge, there are no benchmark databases for fighting detection or recognition. To verify our proposed method, we downloaded the video clips from the Internet. 50 fighting video clips and 50 non-fighting video clips are used for our experiments. Most of the video clips is at around 25 frames per second and last around 3~5 seconds. If a video clips is longer than 5 seconds, the video clip was divided into 3-second clips, and calculated the optical flow context histogram separately. The experiments is implemented based on OPENCV libraries and it is running on Intel Cord IV CPU.

Because the video clip database is not large enough, five-cross folder verification method was used in our experiments. In our experiments, parameter p , M_{max} and parameter q are adjusted to find out the best results. According to our experiments, the parameter p ranges 10 to 15 achieves highest recognition results and lowest false alarm rate when M_{max} values from 800 to 1200. When M_{max} is higher than 1200 or lesser than 800, the accuracy decreased. The greater parameter q is, the higher accuracy is. However, when q is greater than 20, the accuracy of the proposed approach only increased a very small amount. Considering the computational complexity, parameters values are chosen as $M_{max} = 1000$, $p = 10$ and $q = 20$. Random forests achieved best result based on the above parameters. Table I shows the experiments results for fighting and non-fighting video clips using the above parameters and the classification method is Random forest; There are 3 non-fighting video clips recognized as fighting actions. We reviewed these 3 clips and found that these clips contain some sudden running around actions which generated high magnitude optical flow and scattered orientation of optical flow. Table II shows the experiments results using the above parameters and the classification method is support vector machine. Table III shows the experiments results using the above parameters and the classification method is Baysnet. Random forest achieved highest accuracy (accuracy rate is 96% and its error rate is 4%.)while Baysnet has lowest accuracy(accuracy rate is 86% and error rate is 14%).

TABLE I. RECOGNTION RATE AND FALSE ALARM RATE(RANDOM FOREST)

Predict / Observe	fighting	non-fighting
Fighting	49	1
Non-fighting	3	47

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR. Downloaded on June 03,2020 at 16:27:26 UTC from IEEE Xplore. Restrictions apply.

TABLE II. RECOGNTION RATE AND FALSE ALARM RATE(SVM)

Predict / Observe	fighting	non-fighting
Fighting	45	5
Non-fighting	5	45

TABLE III. RECOGNTION RATE AND FALSE ALARM RATE(BAYSNET)

Predict / Observe	fighting	non-fighting
Fighting	42	8
Non-fighting	6	44

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a feature representation called optical flow context histogram to detect abnormal events, especially the fighting violence events. The proposed method has been tested on the video clips and the experiments show that the result is encouraging.

This method works well when the camera is at a fixed point. However, the proposed method is not convincing when the camera is moving. This is because that the optical flow context histogram also reflects the moving of a camera and the proposed method cannot d distinguish the camera motion from the object motion.

The future work of this research is composing of two parts. One is building a benchmark fighting video database, the other is applying the proposed method to other abnormal event detection, such as anti-direction detection.

ACKNOWLEDGMENT

One of the authors, Xiaobo REN, was affiliated with Nanchang Hangkong University, Nanchang, P.R.China. when he took part in this research. The authors thank for Changyi Lu for his valuable discussions and comments on this research. The authors appreciate Jingyan SUN for her hard work collecting the video clips from the Internet for this research work.

REFERENCES

[1] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artificial intelligence, vol. 17, pp. 185-203, 1981.

[2] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," IEEE transactions on pattern analysis and machine intelligence, pp. 555-560, 2007.

[3]C.-C. Chen and J. K. Aggarwal, "Recognizing Human Action from a Far Field of View," presented at the IEEE Workshop on Motion and Video Computing (WMVC), Utah, USA, 2009.

[4] S. J. Blunsden and R. B. Fisher, "PRE-FIGHT DETECTION."

[5] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in IEEE International Workshop on VS-PETS, 2005, pp. 65-72.

[6] F. Cupillard, F. Br¨¦mond, and M. Thonnat, "Group behavior recognition with multiple cameras," in Sixth IEEE Workshop on Applications of Computer Vision, 2002. (WACV 2002). 2002, pp. 177-183.

[7] A. Datta, M. Shah, and N. Da Vitoria Lobo, "Person-on-person violence detection in video data," in Proceedings. 16th International Conference on Pattern Recognition(ICPR), 2002, pp. 433-438 vol. 1.

[8] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2009.

[9] J. Kim and K. Grauman, "Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates," in IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2009, pp. 2921-2928.

[10] Y. Xu, D. Zhang, and J. Y. Yang, "A feature extraction method for use with bimodal biometrics," Pattern recognition, vol. 43, pp. 1106-1115, 2010.

[11] Y. Xu, D. Zhang, J. Yang, and J. Y. Yang, "An approach for directly extracting features from matrix data and its application in face recognition," Neurocomputing, vol. 71, pp. 1857-1865, 2008.

[12] I. Jolliffe, "Principal component analysis," Encyclopedia of Statistics in Behavioral Science, 2002.

[13] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data mining and knowledge discovery, vol. 2, pp. 121-167, 1998.

[14] K. Murphy, "The bayes net toolbox for matlab," Computing science and statistics, vol. 33, pp. 1024-1034, 2001.

[15] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5-32, 2001

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR. Downloaded on June 03,2020 at 16:27:26 UTC from IEEE Xplore. Restrictions apply.