Electronic Letters on Computer Vision and Image Analysis 16(2):33-36, 2018

Contributions to the Problem of Fight Detection in Video

Ismael Serrano-Gracia*

* *VISILAB, E. T. S. I. Industriales, University of Castilla-La Mancha, Spain*
Supervisors: Oscar Deniz and Gloria Bueno**

Received 31th Jul 2017; accepted 9th Mar 2018

Abstract

While action detection has become an important line of research in computer vision, the detection of particular events such as violence, aggression or fights, has been relatively less studied. These tasks may be extremely useful in several video surveillance scenarios such as psychiatric wards, prisons or even in camera smartphones. The clear practical applications have led to a surge of interest in developing violence detectors.

1 Introduction

In recent years, the task of human action recognition from video has been tackled with computer vision and machine learning techniques. Experimental results have been obtained for recognition of actions such as walking, jogging, pointing or hand waving using traditional features. However, action detection has been denoted comparatively less effort. Violence detection is a task that can be leveraged in real-life applications. While there is a large number of studied datasets for action recognition, important datasets with violent actions (fights) were not available until [1].

This PhD thesis is online available on *https://ruidera.uclm.es/xmlui/handle/10578/12481* [2].

2 Proposed Methods

Firstly, a detailed review of state-of-the-art research in violence detection is provided. A categorization of violence detection methodologies is proposed. Violence detection methods are divided into seven categories according to the main algorithm used. These reviewed categories are: spatial and spatio-temporal descriptors, optical flow, trajectories, deep learning, use of audio and others. Moreover, publicly available violence datasets are also reviewed.

Secondly, a novel method for detecting fights is proposed [3, 4]. Blobs of movement are first detected and then different features are used to characterize them. The proposed method makes no assumptions on number of individuals (it can be also used to detect vandalism). Experiments show that the method does not outperform the best methods considered. However, it is much faster while still maintaining useful accuracies ranging from 70% to near 98% depending on the dataset.

Correspondence to: $< isma\_150@hotmail.com / ismael.serrano@uclm.es >$

Recommended for acceptance by Anjan Dutta and Carles Sánchez
https://doi.org/10.5565/rev/elcvia.1135
ELCVIA ISSN: 1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

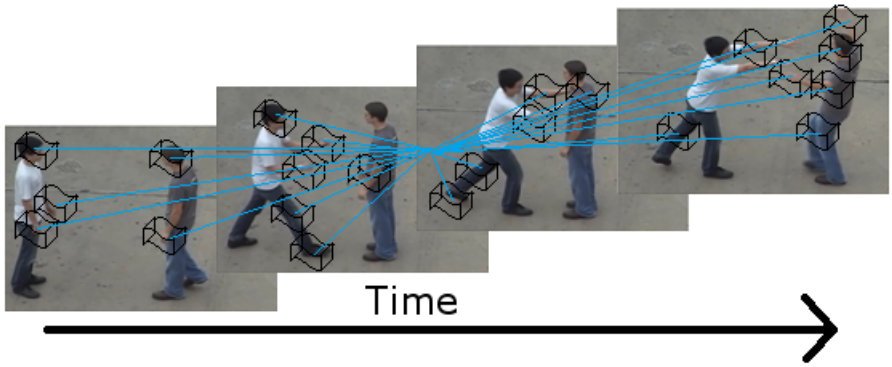

Figure 1: A pushing sequence with some STEC trajectories represented. Each STEC trajectory points to the action center

Thirdly, inspired by results in human perception of other's actions, another novel method for detecting fights is proposed [5], using acceleration patterns as the main feature. These acceleration patterns are highly informative in the task of fight recognition. In general, acceleration can be inferred from tracked point trajectories. However, note that extreme acceleration implies image blur, which makes tracking less precise or even impossible in this case. Motion blur entails a shift in image content towards low frequencies. Such behavior allows to build an efficient acceleration estimator for video. The method was assessed with five other methods using three different datasets. Accuracy improvements of up to 8% with respect to state-of-the-art action recognition techniques were achieved in the most challenging dataset (two of the three datasets).

In addition, the Spatio-Temporal Elastic Cuboid (STEC) Trajectories [6] method for fight detection is proposed. This method is based on the use of blob movements to create trajectories that capture and model the different motions that are specific to a fight. The proposed method is robust to the specific shapes and positions of the individuals. In this novel descriptor (STEC) trajectories are always centered around tracked parts (see Fig. 1). The descriptor only focuses on moving parts whereas the classic cuboids are on a fixed position. The second step, an adaptation of Hough Forests classifier ([7], [8]) is proposed to classify STEC trajectories. In contrast, the classical BoW [9] model assumes independence between spatio-temporal "words" and does not make use of the rich spatio-temporal relationships inherent in actions. Hough Forests leverage this important information. Although it is not the fastest one, the trade-off between computational time and accuracy is clearly better. This method gives an accuracy between 98% and 99.5% in the compared datasets.

Finally, inspired by psychophysics experiments suggesting that motion features may be more important for this specific task a novel hybrid approach is proposed. The rich temporal-voting information from Hough Forests classifier is used, which is fed with BRISK descriptor that capture motion and appearance from a video sequence. Finally, a much simpler 2D Convolutional Neural Network (CNN) is fed with the "handcrafted" images, see Fig. 2. The method demonstrated superiority over both handcrafted feature methods and the previous 3D CNNs.

3 Conclusion

In this PhD thesis, a detailed review of state-of-the-art violence datasets was carried out. A dataset division was presented and compared. In addition, four novel methods for detecting fights in video were introduced.

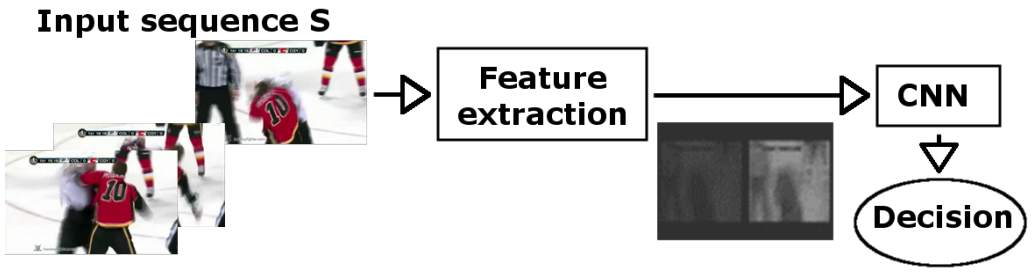

Figure 2: General diagram of the hybrid proposed method

References

[1] E. B. Nievas, O. D. Suarez, G. B. García, R. Sukthankar, Violence detection in video using computer vision techniques, in: International conference on Computer analysis of images and patterns, Springer, 2011, pp. 332–339.

[2] I. Serrano, "Contributions to the Problem of Fight Detection in Video", Ph.D. Thesis, University of Castilla-La Mancha, Spain, [Online]. Available: https://ruidera.uclm.es/xmlui/handle/10578/12481 (2016).

[3] O. Deniz, I. Serrano, G. Bueno, T.-K. Kim, "Fast violence detection in video", in: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on, IEEE, 2014, pp. 478–485.

[4] I. Serrano, O. Deniz, G. Bueno, "VISILAB at MediaEval 2013: Fight Detection", in: Working Notes Proceedings of the MediaEval 2013 Workshop.

[5] I. Serrano, O. Deniz, G. Bueno, T.-K. Kim, "Fast fight detection", PloS ONE 10 (4).

[6] I. Serrano, O. Deniz, G. Bueno, G. Garcia-Hernando, T.-K. Kim, Spatio-temporal elastic cuboid trajectories for efficient fight recognition using hough forests, Machine Vision and Applications 29 (2) (2018) 207–217.

[7] D. Waltisberg, A. Yao, J. Gall, L. Van Gool, "Variations of a Hough-voting action recognition system", in: Recognizing Patterns in Signals, Speech, Images and Videos, Springer, 2010, pp. 306–312.

[8] G. Garcia-Hernando, H. J. Chang, I. Serrano, O. Deniz, T.-K. Kim, Transition hough forest for trajectory-based action recognition, in: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, IEEE, 2016, pp. 1–8.

[9] J. C. Niebles, H. Wang, L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words", International Journal of Computer Vision 79 (3) (2008) 299–318.