ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/291957148

Dangerous human event understanding using human-object interaction model

Conference Paper · September 2015

DOI: 10.1109/ICSPCC.2015.7338786

CITATIONS
2

READS
138

4 authors, including:

Zhaozhuo Xu
Rice University
10 PUBLICATIONS **73** CITATIONS
SEE PROFILE

Xinjue Hu
Beijing University of Posts and Telecommunications & University of Ottawa
7 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Fangling Pu
Wuhan University
44 PUBLICATIONS **293** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Transmission and Processing for Light Field Image/Video View project

All content following this page was uploaded by Zhaozhuo Xu on 23 August 2016.

The user has requested enhancement of the downloaded file.

Dangerous Human Event Understanding using Human-Object Interaction Model

Zhaozhuo Xu IEEE student member, Yuan Tian, Xinjue Hu, Fangling Pu IEEE member
School of Electronic Information
Wuhan University
Wuhan, China
xuzhaozhuo@whu.edu.cn

***Abstract*—Detection of complex human events in videos and images is a challenging problem of computer vision. The difficulty lies in constructing effective connection between human activities and specific events. In this paper we focus on dangerous human events, especially when people with handheld weapons are presented in images. By introducing Human-Object Interaction model, we are able to establish methods and systems to recognize events that are dangerous. In our approach, the process of event understanding is based on identifying dangerous objects in possible areas predicted by human body parts. The accuracy of dangerous human events understanding is improved when human body parts estimation is combined with objects detection. Utilizing a developed dangerous human events data set, we show our model and system outperform conventional event classification approaches in efficiency.**

***Index Terms*—Human Event Classification, Human-Object Interaction, human pose estimation**

I. INTRODUCTION

Computer vision based complex human events classification, because of its application in surveillance systems, is receiving more and more attention. Human events can be thought of as compositions of divergent movements and tendencies of human bodies. Existing approaches focus on social roles determination in large crowds and pedestrian event detection for driver assistance systems [1-3].Useful classification model like Scale Invariant Feature Transform-(SIFT) combined with Support Vector Machine (SVM) [12] has been successfully implemented on event recognition. In this paper, we discuss a special situation of human events that has dangerous tendencies. As terrorist attacks including recent one to Charlie Hebdo continue to raise panic in public. Recognizing dangerous human events and giving early warning are desired.

With an aim of understanding dangerous human event, the use of existed methods in human activities understanding has shown to boost detection rate. In recent years, developed approaches such as Deformable Parts Model (DPM) [4-6] and poselets [7] treat human actions as groups of human body parts. In part based models, human body parts are a set of locations whose geometric arrangements are gained by a Gaussian distribution [8] or a set of "springs" which connect the body parts [9]. When approaches such as wavelet-like features [10] or locally normalized histograms of gradients [11] were employed, the performance of action recognition was advanced. Building upon the previous models, poselets, an algorithm for detecting people using the 3D and 2D annotations of human key points to represent body movements was developed and receive high accuracy when applied. When all methods above are applied to dangerous human events, it seems that the performance will be promising.

However, unlike other human events, dangerous human events have closely connection with special objects. Most of dangerous human events consist of handheld weapons such as guns and swords. Obviously the body movement recognition is not enough for accurate dangerous event understanding. As shown in Fig.1, without knowing that the men is holding a threat weapons, it is not easy to estimate dangerous tendencies in those pictures. Established methods in human pose recognition like DPM and poselets cannot detect dangerous human event accurately because the key role of objects in human events classification is neglected.

Fig. 1. Dangerous human event recognition using HOI approaches.

ICSPCC2015
978-1-4799-8920-1/15/$31.00 ©2015 IEEE

Therefore, Human-Object Interaction (HOI) [13-15] is considered an applicable method to provide comparatively effective solutions that can be introduced to understand dangerous human events. Existing approaches in detecting HOI from still images include modeling the mutual context between objects and human poses in activities [16], recognizing it via exemplar based modelling [17-19] and other inspiring approaches [20-22]. These models have an extraordinary effect on classification of people playing sports, performing with instruments [15], drinking and smoking [23]. However, they are seldom utilized in human dangerous event recognition.

In our work, we establish an algorithm by constructing HOI model indicating connections of suspicious objects and specific body parts, which treats objects as the extension of human's hips. When HOI method is introduced, the detection of dangerous human events is done by determination of dangerous objects (knives, guns, and etc) in predicting bound drawn by certain direction and distance based on the location of hips. Our approach increase the accuracy in dangerous event detection from various image sources.

The paper is organized as follows: In Section 2, the HOI model dangerous human event is described. The detecting systems for dangerous human event are introduced in Section 3. The computation results and their analysis are presented in Section 4. Summarization is in Section 5

II. HUMAN-OBJECT INTERACTION MODEL FOR DANGEROUS HUMAN EVENTS

In this section, our objective is constructing relationships between body parts and handheld objects in order to provide accurate prediction of holding objects. Our initial idea focused on specific connections between hands and holding weapons. As human poses are highly articulated and parts of bodies are self-occluded, we have to avoid parts that are difficult to locate and easily influenced by the position of human in pictures. And there is no doubt that hands are the toughest parts for computer to recognize in images. The training samples will accumulate rapidly in recognition because hands have various shapes shown in pictures.

Realizing the problems above, here, we propose a model that use hips instead of hands to predict possible areas of handheld objects according to the mechanism of the human body. As Fig. 2 shows, our idea divide the space into 9 disjoint regions based on the nature of human body. After identifying dangerous objects in areas that are predicated with reference to hips, dangerous human event understanding can be simplified as object detection in images fragments.

With everything gathering together, we are able to represent our model as:

$$\Psi(O,X) = \sum_{l=1}^{N_e} \sum_{i=1}^{N_b} \sum_{j=1}^{N_d} 1_{(O^i = o_j)} \cdot \gamma_{i,j}^{T} \cdot h(x_E^l) \quad (1)$$

Where $\Psi(O,X)$ models the judgment system based on the classification in the extensive area of human's hips. O and X are the objects and hips of human body located by Poselets. N_b is the number of object areas determined by hips. O^i represents the detected objects in predicted area. N_d is the number of weapons trained to detect in pictures. $1_{(O^i = o_j)} = 1$ if dangerous objects are recognized in the idea area. N_e is the number of the hips detected. $\gamma_{i,j}^{T}$ stands for the estimating rules extended from hips in still images. Finally $h(x_E^l)$ is the function aims at finding comparative position of hips to shoulders in images to determine whether someone's hands are up or down.

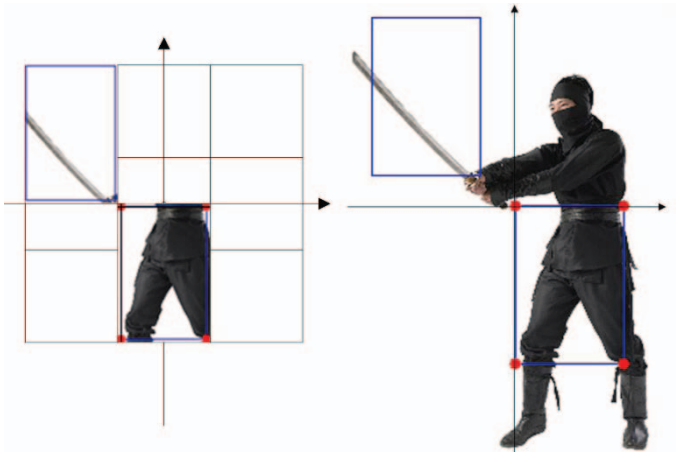

Fig. 2. HOI model for violent events

III. DANGEROUS HUMAN EVENT DETECTION

An overview of our detecting system is shown in Fig. 3. We first get the location of hips using Poselets and draw a bound that mark the area that hips cover. Then established Human-Object Interaction model are applied to forecast the possible areas for objects. After that we are able to identify whether the object is dangerous by objects classifier. Finally, the dangerous human events can be understood.

A. Hips recognition based on poselets

The first step in our detecting systems for dangerous human events is to locate where the hips are in effective ways. Nowadays human body parts are usually described in an innovative interpretation called Poselets. This description allow us to portray the same body parts by persuasive criterion [24, 25]. Based on the models used to obtain the features in still images like Histograms of oriented gradients (HOG), it is reasonable for us to segment human bodies into several parts and get features of parts we needed. In this paper we focus on special poses in order to get bounds that contain hips. Detection of other body parts are also included because they helps in hip recognition.

With a group of training samples of human body parts, we have access to the normalized 3D coordinates of the key-point of the examples by clustering features representing body parts. We denote the annotation of human hips as x_E^l, where x_E^l is a vector got from locations of body parts. The locations are transferred from 2-dimentional positions into 3-dimentional positions [25]. At first we normalize the 3-D positions to [-1, 1]. Then hierarchical clustering is applied to obtain sets of body parts.

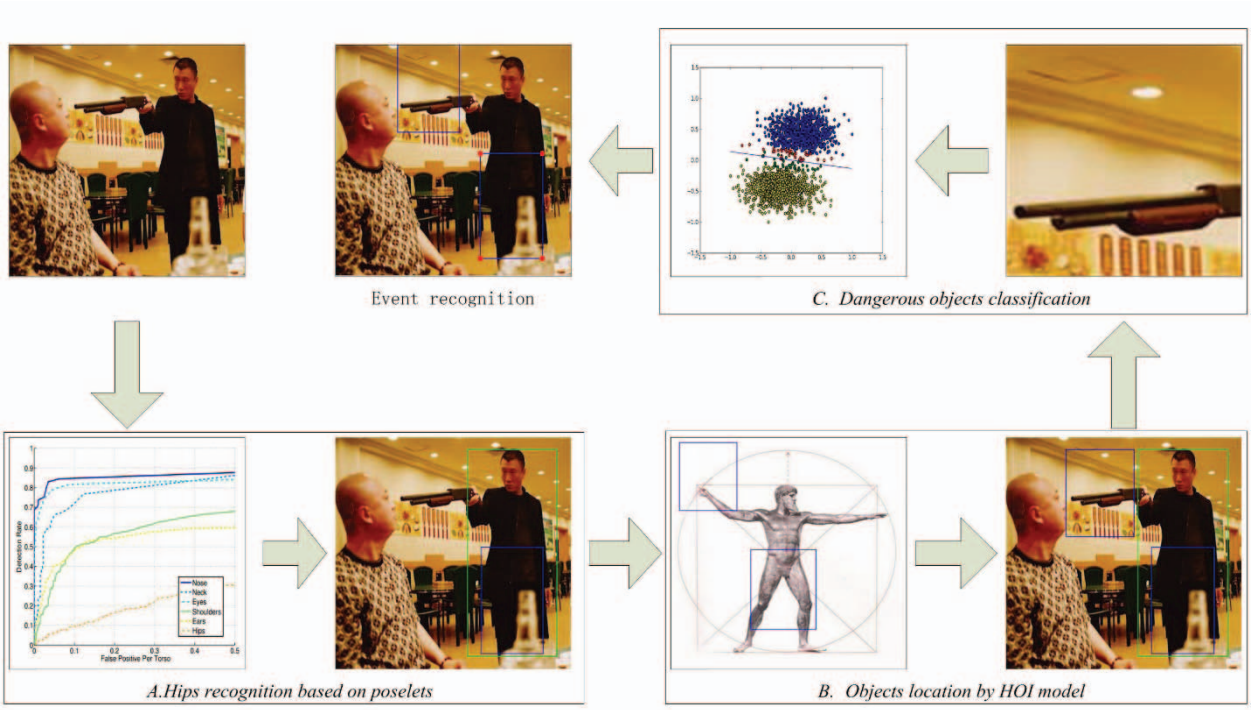

Fig. 3. Overview of our detecting systems

B. Objects location by HOI model

In our work, the possible location of objects in dangerous human events are obtained by established HOI models. When people are detected in images, it is easy to get the locations of hips based on the body proportion and practice of body movement in previous sections. The area in oblique upward direction of hips on both sides can be drawn depending on our HOI model. As shown in Fig. 4, our model performs well in accurately locating hand-held guns and swords. The bounds drawn in pictures narrow down the places we are searching for and benefits the following classification of handheld objects.

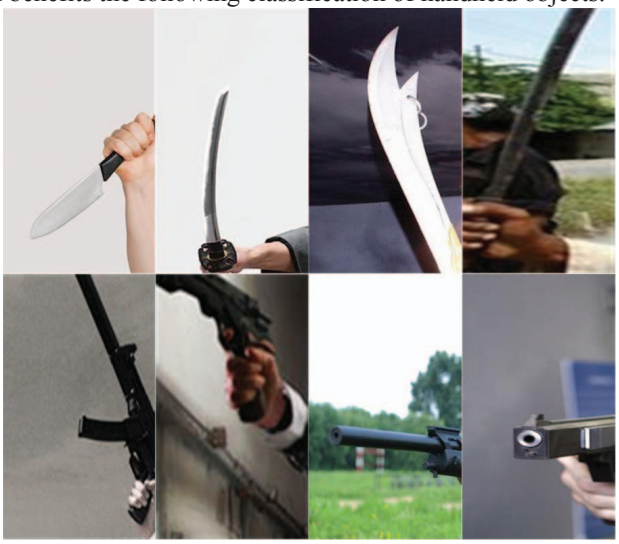

Fig. 4. Effect of HOI model for objects location

C. Dangerous objects classification

In this section, our goal is to identify whether the handheld objects are dangerous using useful objects classification algorithms. When predicted locations of threat objects are obtained in previous procedure, human events can be judged dangerous if objects detected in the drawn bounds are threat. Therefore, the object classification determine the final results of our systems.

SVM classification has become a very prevailing kernel based classification approaches in objects classification because it can provide high detecting rate. In our work the objects recognized in predicted area are divided into two categories which represent if they are dangerous. A binary classifier is trained for both categories. All of the training samples in both category are gained by cutting out handheld objects from dangerous and not dangerous images to make sure its accuracy.

In our work, the HOG features of handheld objects are regarded as the training and testing input of linear kernel based SVM classifier. When the target of the classification is determined, proper selection of training samples will help us to comparatively accurate recognition of violent events.

IV. EXPERIMENTS

A. The PHW data set

Among all the data sets in computer vision, images collected for human and objects interaction are relatively rare compared with the abundant collections of scenes [26] and objects [27, 28]. Nevertheless data sets for human and objects interaction such as People-playing-musical-instruments data set and sports data set, have been applied widely in comparing

advantages of models developed to detect human actions. With the purpose to build a data set belongs to violent event detection, we therefore collected a new data set named People-Holding-Weapons. Each group consists of ~50 PHW+ images (humans holding weapons) and ~50 PHW-images (humans not holding weapons). As Fig.5 indicates, pictures in PHW are highly diverse and cluttered.

In this data set we are going to help computer figuring out whether a person is holding a weapon with his hands up or down at first, then we can distinguish different objects by PHW+ and PHW- samples.

B. Results

In this experiment, we evaluate accuracy of our methods and existing image classification algorithm on dangerous human events classification on PHW data set. In order to make comparison, traditional image classification approach which get SIFT features and adopted them to SVMs classifiers is introduced in our paper.

Our goal is to differentiate dangerous human events using HOI models to narrow down the detecting areas according to human poses and understand dangerous human event by constructing connections between body movements and threat objects. Linear kernel SVM classification are applied in both approaches to make sure the outcome is not influenced by various classification methods.

TABLE I. TABLE TYPE STYLES

Approaches	**Dangerous**		
	Training samples	***Testing samples***	***Accuracy***
Proposed HOI method	30	60	74%
SIFT+SVM	30	60	52%

As TABLE I shows, our approaches has superior results over classic methods in dangerous event classification. Results are gained by representative training and testing samples.

In our work, the HOI model for violent events and detecting systems is implemented by Matlab. The system can be applied to recognize dangerous human events in divergent situations including bank robbery and terrorist attacks. As shown in Fig. 5, the dangerous pedestrian events can be recognized in past threat attacks by our approaches.

V. CONCLUSION

In this paper, human-object interaction (HOI) is introduced to recognize dangerous human events. The dangerous human events can be identified by detecting dangerous objects in areas predicted according to the positions of hips. Combining the classified objects and movement of hips, we can understand dangerous human event. Also we have set up the PHW data set for further research. However, some special dangerous human events remain to be recognized in better ways. More training samples are also required. In future works, we hopes to expand learning data for our PHW data set and apply our models in large scale detection.

Fig. 5. Experiment in well-known terrorist attacks from street and classroom cameras.

ACKNOWLEDGMENT

This wok was supported by National High Technology Research and Development Scheme (863 Project) of China under Grant No. 2013AA122301.

Gratitude to Professor Chu He from Wuhan university, who guide us in our research and helps us a great deal in constructing original models and approaches to achieve our goals.

REFERENCES

[1] Wohler C, Kressler U, Anlauf J K. Pedestrian recognition by classification of image sequences: global approaches vs. local spatiotemporal processing. In:Proceedings of 15th International Conference on Pattern Recognition. Barcelona, Spain. IEEE, 2000.2:540-544.

[2] Curio C, Edelbrunner J, Kalinke T, Tzomakas C, Werner von Seelen. Walking pedestrian recognition. IEEE Transactions on Intelligent Transportation Systems, 2000, 1(3): 155-163.

[3] Bertozzi M, Broggi A, Fascioli A, Graf T. Meinecke M M Pedestrian detection for driver assistance using multiresolution infrared vision.IEEE Transactions on Vehicular Technology, 2004, 53(6): 1666-1678.

[4] A Discriminatively Trained, Multiscale, Deformable Part Model

[5] Y. Amit and A. Trouve, "POP: Patchwork of parts models for object recognition," International Journal of Computer Vision, vol. 75,no. 2, pp. 267–282, 2007

[6] D. Crandall, P. Felzenszwalb, and D. Huttenlocher, “Spatial priors for part-based recognition using statistical models,” in IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[7] L. Bourdev and J. Malik, “Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations,” Proc. 12th IEEE Int’l Conf. Computer Vision, 2009.

[8] M. Weber, M. Welling, and P. Perona, “Towards automatic discovery of object categories,” in IEEE Conference on Computer Vision and Pattern Recognition, 2000.

[9] M. Fischler and R. Elschlager, “The representation and matching of pictorial structures,” IEEE Transactions on Computer, vol. 22,no. 1, 1973.

[10] C. Papageorgiou, M. Oren, and T. Poggio, “A general framework for object detection,” in IEEE International Conference on Computer Vision, 1998.

[11] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[12] SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. H. Zhang, A. C. Berg, M. Maire, and J. Malik. In Proc. CVPR, 2006.

[13] C. S. A. Prest and J. Malik. Weakly supervised learning of interactions between humans and objects. TPAMI, 34(3):601–614, 2012.

[14] C. Desai, D. Ramanan, and C. Fowlkes. Discriminative models for static human-object interactions. In Workshop on Structured Models in Computer Vision, 2010

[15] B. Yao and L. Fei-Fei. Grouplet: A structured image representation for recognizing human and object interactions.CVPR, 2010

[16] B. Yao and L. Fei-Fei. Recognizing human-object interactions in still images by modeling the mutual context of objects and human poses. TPAMI, 34(9):1691–1703, 2012.

[17] B. Yao, A. Khosla, and L. Fei-Fei. Combining randomization and discrimination for fine-grained image categorization. In CVPR, 2011.

[18] V. Delaitre, I. Laptev, and J. Sivic. Recognizing human actions in still images: a study of bag-of-features and partbased representations. In Proc. BMVC, 2010.

[19] G. Sharma, F. Jurie, and C. Schmid. Discriminative spatial saliency for image classification. In CVPR, 2012.

[20] M. Andriluka, S. Roth, and B. Schiele, “Pictorial Structures Revisited: People Detection and Articulated Pose Estimation,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[21] B. Sapp, A. Toshev, and B. Taskar, “Cascade Models for Articulated Pose Estimation,” Proc. European Conf. Computer Vision, 2010.

[22] A. Gupta, A. Kembhavi, and L. Davis, “Observing Human-Object Interactions: Using Spatial and Functional Compatibility for Recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 10, pp. 1775-1789, Oct. 2009.

[23] I. Laptev and P. Perez, “Retrieving Actions in Movies,” Proc. IEEE Int’l Conf. Computer Vision, 2007.

[24] S. Maji, L. Bourdev, and J. Malik. Action recognition from a distributed representation of pose and appearance. CVPR,2011

[25] V. Delaitre, I. Laptev, and J. Sivic. Recognizing human actions in still images: a study of bag-of-features and partbased representations. In Proc. BMVC, 2010. 1

[26] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the shape envelope. Int. J. Comput. Vision, 2001.

[27] A. Gupta, A. Kembhavi, and L. Davis. Observing human object interactions: Using spatial and functional compatibility for recognition. TPAMI, 31(10):1775–1789, 2009.

[28] M. Everingham, L. van Gool, C. Williams, J. Winn, and A. Zisserman. The PASCAL voc 2008 Results. 1, 2

View publication stats