Real time backpack detection in visual surveillance based On verticals contour analysis

Ulziisaikhan Lkhagvasuren^{a,*}, Stephen Karungaru^a, Kenji Terada^a

^aTokushima University, 2 - 1 Minami Josanjima, Tokushima, 770-8506, Japan.

*ulzii.ict@gmail.com

Abstract

In this paper, we aim to propose a method to detect backpacks and sling bags which pedestrians are carrying in visual surveillance. A real time video analysis and processing on passive recording provided from surveillance camera could give more valuable information for a variety of applications. We detect backpacks without protrusion supposition or any training data. We propose an algorithm to find expectable vertical lines inside or near a human body contours using some fuzzy rules. To improve detection accuracy, area of carried object is compared with next several frames. It influences our detection result greatly. As a result, we detect carried objects including backpack and sling bag successfully. We run several experiments with previously designed VIRAT ground video dataset and result shows that our method is promising.

Keywords: vertical contour, backpack detection, sling bag detection

1. Introduction

Potential classification of specific object in crowded area is a demanding application. In recent years, terrorism activity is increasing in crowded areas such as stadiums, hall etc. Fig 1 shows convicts of Boston marathon bombing in 2013, where some attributes of convicts could be 'carrying backpack', 'wearing hat' and 'dark clothes'. Government and intelligence bureaus are focusing on public security. Surveillance cameras are mainstream and such static video streams can be used as an observation source. An object classification, a detection and a tracking on stream can provide more important and valuable information.

Fig1. A combination of security camera images of Boston marathon bombing convicts carrying backpack.

Detection of carrying object is challenging in case its size is too small or its color is similar with the clothes. Because, the people can carry any kind of objects in different ways, it makes recognition more complicated. Moreover, its carrying location is variable. For example, a sling bag can be carried on shoulder or can be held by arms. A backpack can be carried on both or one of the shoulders which make it difficult to define geometrical characteristic of backpack. In this research, we aim to detect carrying object such as backpack, sling bags using real time surveillance video stream.

1.1 Related works

Until now different methods of carrying object detection have been proposed. One of them ⁽¹⁾ defines geometrical properties of backpack and sling bag strap with Hough transform and edge detection method. In this method, geometrical properties are separately calculated. The contour is approximated with a polygon with accuracy proportional to the contour perimeter. Re-implementation of ⁽²⁾ Markov random field with a map of prior probabilities for carried objects and spatial continuity assumption, from which

segmentation of carried object using MAP solution is completed ⁽³⁾. A motion based recognition approach investigated is considered as well ⁽⁴⁾. Here, after detecting pedestrians from video frames, they will be classified into categories, walking or object-carrying based on spatiotemporal analysis of the obtained binary silhouettes. To accomplish it, temporal correspondence free analysis have been made on binary frames with periodic motion of pedestrians. Moreover, neural network trained on negative and positive dataset ⁽⁵⁾ has been investigated. An extract features from silhouettes and neural network are trained on set of images. A human carrying status examined ⁽⁶⁾ by using general tensor discriminant analysis, which coherently incorporates the Gabor based human gait appearance model.

2. Proposed Method

As mentioned in introduction, our system is designed to track and classify backpack in street surveillance. It consists of different modules that communicate with each other. An operation flow of our method is shown in Fig2.

2.1 Pedestrian detection

In last years, notable progress has been made for detecting and tracking human in computer vision field. Since we use surveillance camera stream, we implement background subtraction. In foreground image, erosion and dilation are used to convert grayscale image and equalized histogram.

Fig2. System low-level architecture

Contours of each blob are groups to mark out an approximating rectangle.

In addition, Gaussian probability is calculated on aspect ratio feature. In crowded area sometimes, blob (person close to each other) is enormous.

Therefore, we extract human features by using Local Binary Patter LBP ⁽⁷⁾ and Histogram-of-Oriented-Gradient HOG ⁽⁸⁾. In order to improve detection result, we exclude people appearing at a distance or are far from camera. In calculation, an approximation is used to define the minimum possible size of person by using normal distribution.

2.2 Bag detection

Capturing the connection of the object with human posture is a challenging task. Carrying object may occupy a small part of body or is not comparable in height with the human.

In this paper, we consider a solution for detecting bags and backpacks from surveillance cameras. Integration of object features and pedestrian contour is achieved based on foreground subtraction. As a result of several kinds of experiments, we have observed that carrying objects create visible vertical lines in the human body contour or near with human body. Depending on the length of these lines and the distance from human body, a vote is done to define the carrying item Fig. 3.

A Sobel operator is applied to X coordinates to find all verticals. The next operation is to find vertical lines on the binary image by intersecting (bitwise) foreground mask and Sobel X operator result. An ally vertical lines are represented by four element vectors (x_1, y_1, x_2, y_2) , where (x_1, y_1) and (x_2, y_2) are the ending points of each detected line segments.

Then we can calculate the points of these lines that are located furthest from contour. If one of the dots is far from contours by around 10 pixels (parameter should tune video size) or is too close to the bottom part, those lines are excluded, because it is usually legs of pedestrian.

Fig3. Backpack rear and frontal detection based on strap and rear appearance verticals

Fig4. Sling bag rear and frontal detection

After this calculation, we get one line, which probably is a part of bag or bag strap. With those coordinates, we can approximate bag position, fig. 4.

Depending on the human movement, the appearance of the bag can be changed. However, this method might miss some types of carried objects such as roller bag, luggage, since compactness is one of the feature, it still can be voted if objects make straight contour with fist.

Therefore, it is required to compare detected bag results with the next frames to improve result.

3. Experimental result

We performed several experiments with VIRAT ground video dataset ⁽⁹⁾. It is more advanced dataset design than conventional datasets, because its realism and natural scene based operation makes it excellent. This dataset works on data collected from human motion based scenes with undefined backgrounds.

In our experiment, some scenes are captured in university campus area, where most people are carrying backpacks or sling bags. Each video recorded 23 fps at 1280x720 pixel resolution with 679 kbps bitrate.

The system implemented have done by C++ and runs on an Intel i7 3.4GHz machine.

Fig 5 shows an video player and bag carrier and region of interest frames.

Fig5. Detected bags appearing on right side of screen

Table 1. Detected results with compare iteration

	Pedestrian count	No bag	Bag carrier	Undetected	Mis - detection
	Eye count	82	139	-	-
Bag frame compared iteration count	3	99	122	22	5
	2	75	146	19	26
	1	48	173	16	50

Depending on video quality and pedestrian count, each bag frame detection takes an average 38ms – 70ms. To compare bag regions with following three frames, average calculation takes around 3.2s – 5.3s while it takes 1.1s – 1.3s for two frames comparison.

A Table 1 shows our experimental results. All pedestrians on VIRAT dataset are counted by manual and then the detection is implemented with differently adjusted comparing duration.

In each iteration count, both of the pedestrian and bag are counted automatically. However, analyze with 1 frame might be effective, but misdetection is high.

In case of 2 frames comparison, the misdetection is higher as well, but undetected bags reduced slightly. When sequence of 3 frames are used, detection for some bags is increased and the misdetection is decreased slightly.

In such manner, comparing number of frames results increase of undetected bags and delay on system operation. Experimental result shows that comparing bag frame with following next two frames is efficient in terms of less misdetections. Approximate success ratio of carried object detection is 80.5% while single frame detection is 53% of success ratio. Upon investigation on false detection, we found several false detection shown Fig7 and Fig8.

Fig7. Undetected backpack and sling bags

Fig8. Misdetected results

Most of undetected bags are caused by color contrast as shown Fig7 (a) and (b) where color contrast with background or color of the clothes itself causes difficulties in vision detection. Also, in some rare case like Fig7 (c), pedestrian's long hair hides backpack, because long hair lies down the shoulder covers bag strap or bags and it makes it impossible to vote as a bag. Sometimes, even it is hard to discriminate by human. In Fig7 (d), pedestrian occluded strap by one hand. If the bag strap is too short or it's held by, it's not sufficient to vote that it's bag sling.

Fig8 shows some examples of misdetection results. In Fig8 (a) and (b), because pedestrians are holding hands, it makes vertical straight line within human contour. In Fig8 (c), pedestrian is wearing clothes with long vertical textures.

4. Conclusion

We have investigated real time intelligent surveillance camera video analytics for carried object detection. System could achieve 80.5 percent success ratio and it can be more fine-tuned by every view angle and distance of pedestrian path. Another advantage of the method is that detection speed is very fast for real time computation.

Our method can be applied in public spaces like stadiums, hall entrances or metro stations to provide more accurate information for public security.

Ideally, it is common use to have a dataset of multiple images featuring people carrying a bag, but our method does not have such a constraint, therefore detecting any carried object. Future work will focus on optimization and computational speed of the core algorithms.

References

- (1) Teck Wee Chua, Karianto Leman, Hee Lin Wang and Nam Trung Pham, "Sling Bag and Backpack Detection for Human Appearance Semantic in Vision System", International Conference on Intelligent Robots and Systems (IROS) 2013
- (2) Haritaoglu, I., Cutler, R., Harwood L. and Davis,

- "Backpack: detection of people carrying objects using silhouettes. In: Proc. Int. Conf. on Computer Vision (ICCV)", vol. 1, pp. 102–107 (1999)
- (3) Dima Damen and David Hogg, "Detecting Carried Objects in Short Video Sequences", ECCV 2008, Part III, LNCS 5304, pp. 154–167, 2008
- (4) Chiraz BenAbdelkader and Larry Davis, "Detection of People Carrying Objects : a Motion-based Recognition Approach", Computer Vision Laboratory University of Maryland, 2013
- (5) A. Branca, M. Leo, G. Attolico, and A. Distanti, "Detection on objects carried by people", Proc. of IEEE International Conference on Image Processing (ICIP), 2002, pp. 317–320.
- (6) Dacheng Tao, Xuelong Li, S.J. Maybank and Xindong Wu, "Human Carrying Status in Visual Surveillance", Computer Vision and Pattern Recognition on Computer Society Conference 2006
- (7) T. Ojala and M. Pietikainen, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", Pattern Anal. Machine Intell., vol. 24, pp. 971–987, 2002 -4
- (8) N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005 -5
- (9) Sangmin Oh, Anthony Hoogs, Amitha Perera and Naresh Cuntoor "A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video", Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2011.-6