A Surveillance Video Condensation System Based on the Spatial and Temporal Rearrangement of Moving Object Trajectories

Chin-Shyurng Fahn
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
csfahn@mail.ntust.edu.tw

Meng-Luen Wu
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
D10015015@mail.ntust.edu.tw

Chun-Chang Liu
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
M9615913@mail.ntust.edu.tw

***Abstracts*— In this paper, a video condensation system is proposed to reduce the total length of surveillance videos without missing any information of moving objects. To begin with, we carry out foreground segmentation, and exploit tracking techniques such as connected component labeling to extract the trajectory of a moving object. Each trajectory plays the role of a basic unit for video condensation. The sequence of trajectories are sorted and rearranged to constitute a condensed video to reduce the length of the original one. Different methods based on spatial distribution, shortest path, and moving direction are respectively employed for sorting the trajectories. In addition to reducing the video length, the movement harmonic among moving objects is an important issue, since a condensed video must be easy to watch; otherwise, it will be embarrassed for users. In experiments, we test our system under three kinds of scenarios, such as campus, highway, and square. The results reveal that our system is able to produce a condensed video with short video length, as well as objects with harmonic moving directions.**

Keywords*— *Video condensation, surveillance video, spatial and temporal rearrangement, moving object trajectory, movement harmonic, abnormal event detection.

I. INTRODUCTION

Along with the widespread of digital cameras and the progress of surveillance systems, artificial-intelligence-based forensic systems successfully help solve lots of criminal cases in recent years. However, most of the surveillance systems are restricted to handle their defined events, whereas undefined cases would not be detected. It is not possible that all criminal cases are defined in advance, so human-aided surveillance video examinations are still needed even with the state-of-the-art technology. Nevertheless, a surveillance video is tediously long, full examination is time-consuming. Fast-forwarding the video would save monitor time, but it often leads users fail to declare critical events.

Currently, most intelligent surveillance systems adopt an event detection approach [1]. In such systems, however, the abnormal events must be pre-defined, which means that the abnormal activities must be observed and defined beforehand, and create a knowledge database of event examples from which the features are extracted. The aforementioned systems cannot detect abnormal events by self-working without these preparations. Besides, redefining the knowledge base is required in different environments. Another method is designed to automatically detect abnormal objects in interesting regions by analyzing a video and extracting information from the video [2]. As far as the above reasons are concerned, an artificial-intelligence-based method is expected to be developed by use of self-learning techniques [3] to detect abnormal objects in different environments automatically.

Moreover, in some of the cases, the purpose for watching surveillance video is to find known criminals, but their walking paths are usually similar to other pedestrians in videos, which causes some intelligent surveillance systems make no use [4]. In such a case, the videos must be examined manually [5]. The drawback of manual examination is not only inefficient but also that human's concentration time is limited [6]. Because of the above facts, in this paper, a video condensation system is proposed, which can promote security, and reduces labor consumption in all surveillance environments.

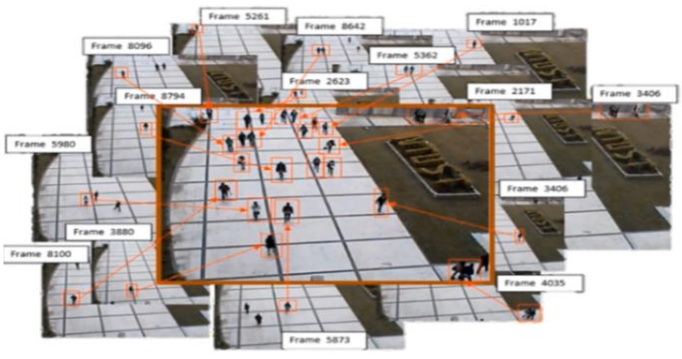


Figure 1 A frame of a condensed video where the moving objects are synthesized from different video frames.

Figure 1 illustrates a goal of our proposed condensation system. In this figure, the trajectories of moving objects from different frames are synthesized. It can produce a condensed video with the shortest length and preserve as more details as possible. This system will save a huge amount of monitor time for video forensics in all video scenarios.

II. SYSTEM DESCRIPTION

There are two phases of our proposed video condensation system. The first phase is moving objects extraction, and the second phase is video condensing where some condensation methods can be selected according to user's demand. In the moving objects detection phase, we use GMMs [7] to segment the foreground in a video. Next, label moving objects by tracking techniques, and extract the objects appearing in consecutive frames individually. At last, compute the trajectory information of the moving objects, such as occurrence time, moving angle, and moving speed, then save them into an object database [8]. Figure 2 shows the procedure of the moving objects extraction phase.

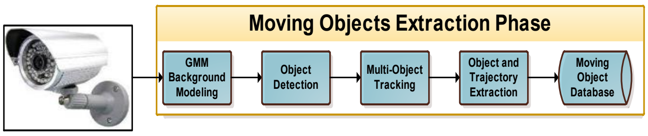


Figure 2 The procedure of the moving objects extraction phase.

Before the video condensing phase, the object database has been set up. The candidate selection is on user's demand. There is a constraint that objects do not occlude each other, so the objects appearing sequence in a given video will influence the result of condensing the video, including the condensed video length, as well as the spatial distribution of moving objects. Through spatial and temporal rearrangements, the moving objects selected from a candidate list in the database are put to the condensed video. If a moving object occludes other objects in the condensed video, choose another one which does not. After all objects in the database are put in the video, the condensation is completed. The procedure of the video condensing phase is shown in Figure 3.

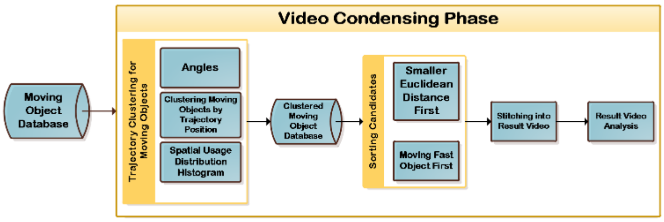


Figure 3 The procedure of the video condensing phase.

III. STRATEGIES FOR VIDEO CONDENSATION

Besides reducing the video length, the appearance of the moving objects in a frame should not be too messy; otherwise, it won't be comfortable to watch. If trajectories with similar features are put to adjacent time sequences, the condensed video length will be shorter and readability will be better. In this section, we propose some trajectory sorting rules to condense surveillance videos with shorter length, and make them more comfortable for viewing what happened, without missing any information.

A. First In First Out

The first in first out sorting rule is the most naïve way for sorting trajectories. Although picking a moving object from the candidate list is according to the order of the moving object appearing in the original video, in the video condensing phase it still need to check whether the candidate has been deposited in the resulting video. Therefore, the sequence of the moving objects in the condensed video is not always the same as the sequence of the original video.

Principally, the sequence of object selection is the same as that of the objects stored in the candidate list. However, when switching into the video condensing phase, a selected object will be checked whether it occludes another object in the current frame of a condensed video. Consequently, the object appearance sequence is different from the object saving order in the candidate list.

B. Sorting by Position Proximity

The distance between two moving objects should be enough to prevent from object occlusion. Thus, in a condensed video, two moving objects should have a safe distance. In the same time, the condensation rate maintains the least. To achieve this, we select an object which has the shortest distance between it and each object in the condensed video, where the distance is computed by the Euclidean metric.

C. Sorting by Moving Speed

The rate of capturing a frame is fixed. Accordingly, when an object stays in the video for longer time, the captured frames of the object are more. Therefore, the moving speed is in inversely proportional to frame counts. If a fast moving object is arranged after a slow moving object, for preventing occlusion, the fast moving object must be delayed to appear; otherwise, it will bump into slow objects. However, the delay incurs traffic jam. The longer a moving object stays in the field of view, the more time for objects that occupy the same space. It will influence the result of condensing a video.

D. ClusteringMoving Objectsby Usage Distribution Histogram

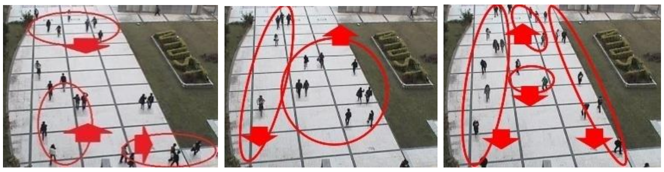
When we want to synthesize a moving object into the condensed video, we would choose the objects which pass less traffic rate regions. The traffic rate of each pixel is obtained by computing how many objects have passed that pixel. Pixels with higher traffic rates are the spatial bottleneck regions of the video [9]. If we always choose moving objects that pass higher traffic rate regions, it is likely that the objects occlude each other in the condensing phase. Therefore, selecting the moving objects which pass the less traffic rate regions first is helpful for reducing the length for the condensed video. Otherwise, it is recommended to choose objects that pass higher and less traffic rates alternatively.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Many experiments are done to evaluate the effectiveness of our proposed system. In the experiments, we examine the condensation results from three videos in different scenarios, each of which comprises 3,600 frames originally, using different clustering and sorting algorithms [10]. We aim at producing a condensed video with shorter video length, and more harmonic trajectory between moving objects, under the constraint that the moving objects never occlude each other.

A. Condensation Results on the Campus Scenario

In the campus scenario, the moving speed seems slower because the camera is placed in a high place. There is no specific entrance or exit in this case, so trajectory clustering is difficult. The best approach for this case is to compute the pass through rate of pixels and obtain bottleneck regions. Objects traversing the region with a high pass through rate are selected first, and the traffic of bottleneck regions can be reduced. Moreover, we use the shortest path approach to help selection. With the methods mentioned above, both the clustering and condensation results are made better. In Figure 4, the degree of movement harmonic among moving objects is higher when sorting by the Euclidean distance, while lower without sorting that uses the FIFO mechanism only.



(a) (b) (c)

Figure 4 The condensation result of the campus scenario based on: (a) the FIFO mechanism; (b) sorting by the position proximity; (c) sorting by the moving speed.

In Table I, when the moving object sequence is clustered by the "spatial distribution histogram," all the produced video lengths are less than 20% of the original video. It shows that some regions with high traffic have an impact on the result of reducing the video length, but appropriately clustering moving objects can alleviate the impact. It is surprising that using the FIFO mechanism together with normal condensation management (i.e., without any clustering manipulation) also achieves good result (15.51% video length of the original), which may because in natural movements, moving objects (e. g., people in the video) would certainly avoid being collided with others and choose to walk on the path with less traffic in the scene. In this case, we can see that choosing the clustering strategy has more impact on the result than choosing the sorting strategy.

TABLE I PERFORMANCE EVALUATION OF THE CAMPUS SCENARIO WITH DIFFERENT CONDENSATION STRATEGIES

Condensation Strategy	Sorting Scheme	Video Length	Execution Time
Normal Condensation Management	FIFO	1,425 (15.51%)	668ms
	Moving Speed	1,560 (16.90%)	678ms
	Position Proximity	1,566 (17.04%)	827ms
Clustering Moving Objects by Motion Orientation	FIFO	806 (22.30%)	95ms
	Moving Speed	827 (22.90%)	74ms
	Position Proximity	826 (22.90%)	149ms
Clustering Moving Objects by Spatial Distribution Histogram	FIFO	1,610 (17.52%)	11,305ms
	Moving Speed	1,546 (16.82%)	11,330ms
	Position Proximity	1,451 (15.79%)	11,171ms
Clustering Moving Objects by Trajectory Position	FIFO	1,713 (18.64%)	2,127ms
	Moving Speed	1,544 (16.80%)	1,963ms
	Position Proximity	1,971 (21.45%)	2,027ms

It is noticed that although clustering moving objects by motion orientation does not acquire the shortest video length, the execution time on condensing a video spends significantly shorter. This may result from objects moving in a similar manner, but the traffic rate differs from various regions. For attaining higher object selection efficiency, choosing the objects with similar motion orientation can greatly reduce the execution time of putting a selected object into the condensed video where any existing object does not occlude each other.

B. Condensation Results on the Highway Scenario

In the highway scenario, there are only two moving directions, and the region in which vehicles run is limited. Since only two clusters exist, moving objects clustering is not meaningful in this case. The object selection that we adopt the shortest distance sorting method works well in the video condensing phase. And the speed sorting method further improves the condensation rate as Figure 5 shows. However, the moving speed for vehicles in the highway is very high; the condensation result is too dense for examination.

(a) (b) (c)

Figure 5 The condensation result of the highway scenario based on: (a) the FIFO mechanism; (b) sorting by the position proximity; (c) sorting by. the moving speed.

Because the vehicles in the video identically run in two opposite directions and the moving region are limited, In Table II, it can be found that the different combination of sorting and clustering strategies does not make too much difference on video lengths after condensation. In this scenario, the video is condensed at the rate of 22%~26% in different strategies.

TABLE II PERFORMANCE EVALUATION OF THE HIGHWAY SCENARIO WITH DIFFERENT CONDENSATION STRATEGIES

Condensation Strategy	Sorting Scheme	Video Length	Execution Time
Normal Condensation Management	FIFO	808 (22.0%)	118ms
	Moving Speed	848 (23.5%)	94ms
	Position Proximity	838 (23.3%)	208ms
Clustering Moving Objects by Motion Orientation	FIFO	806 (22.3%)	95ms
	Moving Speed	827 (22.9%)	74ms
	Position Proximity	826 (22.9%)	149ms
Clustering Moving Objects by Spatial Distribution Histogram	FIFO	795 (22.0%)	4,865ms
	Moving Speed	829 (23.0%)	5,101ms
	Position Proximity	827 (22.9%)	5,048ms
Clustering Moving Objects by Trajectory Position	FIFO	924 (25.6%)	1,541ms
	Moving Speed	909 (25.3%)	1,510ms
	Position Proximity	904 (25.1%)	1,570ms

C. Condensation Results on the Square Scenario

In the square scenario, clustering is useful for grouping moving objects. The moving directions are separated into different ones and are shown sequentially in the condensed video, which lightens the burden of video examiners. In Figure

6, we can see that "sorting by the position proximity" between moving objects acquires the best movement harmonic in the condensed video, which is as good as clustering moving orientation of objects.

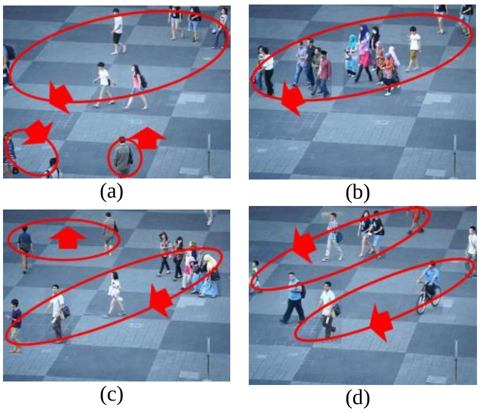


Figure 6 The condensation result of the square scenario based on: (a) the FIFO mechanism; (b) clustering moving objects by motion orientation and sorting by the position proximity; (c) clustering moving objects by usage distribution histogram; (d) sorting by the position proximity.

In Table III, we find that "clustering moving objects by the usage distribution histogram" again produces the best result for shrinking video length, which the video lengths are all shortened to be less than 45%, and the different combination of sorting and sorting strategies yields little difference. Because the moving objects in this scene all move in unpredictable directions, clustering them by motion orientation does not reduce too much computational time than that without any clustering manipulation, but is about 10 times less than using the spatial distribution histogram.

TABLE III PERFORMANCE EVALUATION OF THE SQUARE SCENARIO WITH DIFFERENT CONDENSATION STRATEGIES

Condensation Strategy	Sorting Scheme	Video Length	Execution Time
Normal Condensation Management	FIFO	5,182 (44.97%)	1031ms
	Moving Speed	5,205 (45.17%)	1,083ms
	Position Proximity	5,042 (43.76%)	1,157ms
Clustering Moving Objects by Motion Orientation	FIFO	5,573 (48.37%)	1,118ms
	Moving Speed	5,372 (46.63%)	1,135ms
	Position Proximity	5,586 (48.48%)	1,365ms
Clustering Moving Objects by Spatial Distribution Histogram	FIFO	5,073 (44.03%)	12,103ms
	Moving Speed	5,087 (44.15%)	12,225ms
	Position Proximity	4,997 (43.37%)	12,544ms
Clustering Moving Objects by Trajectory Position	FIFO	5,307 (46.00%)	2,430ms
	Moving Speed	5,585 (48.48%)	2,379ms
	Position Proximity y	5,258 (45.64%)	2,241ms

V. CONCLUSIONS

In this paper, we have presented a surveillance video surveillance condensation system based on the spatial and temporal rearrangement of moving object trajectories, which mainly consists of moving objects extraction and video condensing phases. The users are able to choose their interested time sections of the original video. The proposed system allows user to choose an object within a time section, and perform clustering according to the moving direction , spatial distribution, position proximity within the time section. Through our video condensation system, users can quickly browse many hours of video content. By selecting different clustering strategies, the result of shirnked video length and movement harmonic can be improved. Such a system can greatly ease the work of the viewer finding a suspect in surveillance videos.

From the experiments, we can see that the FIFO mechanism has the shortest execution time on average, but the produced condensed video is just acceptable. Due to lack of clustering motion orientation, the objects would move in a staggered pattern, causing the difficulty to do forensics. When we perform clustering moving objects by position proximity, the condensation rate becomes worse. However, in this approach, it is more likely that the moving objects with similar walking paths appear together, which is easier for forensic officers to watch

ACKNOWLEDGEMENT

The authors thank the Ministry of Science and Technology of Taiwan (R. O. C.) for supporting this work in part under Grant MOST 104-2221-E-011-032-MY3.

REFERENCES

[1] N. Petrovic, N. Jojic, and T. Huang. "Adaptive video fast forward." *Multimedia Tools Application*, vol. 26, no. 3, pp. 327-344, 2005.

[2] J. You, G. Liu, L. Sun, and H. Li, "A multiple visual models based perceptive analysis." *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 3, pp.273-285, 2007.

[3] A. Rav-Acha, Y. Pritch, and S. Peleg. "Making a long video short: Dynamic video synopsis. " in *Proceedings of the Sixth IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 435-441. New York, 2006.

[4] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg. "Webcam synopsis: Peeking around the world." in *Proceedings of the 11th International Conference on Computer Vision*, pp. 1-8. Rio de Janeiro, Brazil, 2007.

[5] Y. Pritch, A. Rav-Acha, and S. Peleg. "Nonchronological video synopsis and indexing." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1971-1984, 2008.

[6] Y. Pritch, S. Ratovitch, A. Hendel, and S. Peleg. "Clustered synopsis of surveillance video." in *Proceedings of the Sixth Advanced Video and Signal Based Surveillance*, pp. 195-200, Genoa, Italy, 2009.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337-1342, 2003.

[8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd Ed., Addison-Wesley, Reading, Massachusetts, 1992.

[9] H. John and W. Manchek, "A K-means clustering algorithm," *Journal of the Royal Statistical Society. Series C* (*Applied Statistics*), vol. 28, no. 1, pp. 100-108, 1979.

[10] C. C. Liu, "A novel video condensation approach based on the spatial and temporal rearrangement of moving object trajectories," *Master Thesis*, Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, 2013.