Piecewise Video Condensation for Complex Scenes

Yingying Chen[1,2], La Zhang[1,2], Jinqiao Wang[1,2], Hanqing Lu[1,2]

[1]National Laboratory of Pattern Recognition, Institute of Automation
[2]University of Chinese Academy of Sciences
{yingying.chen, la.zhang, jqwang, luhq}@nlpr.ia.ac.cn

Abstract. Video synopsis or condensation provides an efficient way to video storage and browsing. Lots of improvements have been made for boosting the speed or improving the condensed quality, which have shown promising results. However, most of the existing approaches cannot effectively deal with complex scenes, such as sudden changes or background object movement. In this paper, we propose a robust video condensation approach for complex scenes. A video segmentation method is designed to analyze the background complexity and divide the input video into several segments. The advantage is two-fold: one is to judge the complexity of backgrounds; the other is to generate a piecewise background image for each segment. Then, we adopt a divide-and-conquer strategy for video condensation. We keep the original video segments for complex backgrounds while maximally condense the other segments. Next, we introduce a feedback scheme and a selective diffusion strategy to keep the integrity of foreground objects, followed by a sticky trajectory method to remove noisy fragments and reduce blinking effect. Furthermore, an adaptive truncation strategy is introduced to raise the condensation ratio and improve the visual quality. Experimental results demonstrate the effectiveness of our approach.

1 Introduction

Nowadays millions of surveillance cameras are installed for abnormal incidents, criminal evidences detection and traffic management etc. The world witnesses a large amount of video data recorded for security purposes every day. Browsing and indexing activities in these abundant videos are a time consuming and boring work for viewers. To alleviate the burden for video browsing and searching effectively, many approaches of video condensation were proposed, such as fast forwarding[1–3], video summarization[4], video montage[5], video synopsis[6–8]and ribbon carving[9].

The goal of video condensation is to shorten the original videos with minimum information loss. Video synopsis, first proposed by Peleg and his colleagues[6–8], showed better performance on controlling the loss of information than other previous video abstraction approaches. It mainly involves three steps: (1) Extract moving objects from the original video to constitute the basic processing unit "tube" (a tube is a spatio-temporal sequence of a moving object); (2) Generate background images by shifting temporal median window over original frames; (3) Rearrange extracted tubes and densely stitch them into background images. Online video condensation (OVC) proposed an online framework [10–12], by transforming the tube rearrangement into a stepwise optimization problem. However, these approaches ignored the complexity of surveillance scenes. As illustrated in Fig. 1, in these complex scenes such as background sudden changes or continuous background object movement, moving objects are difficult to be extracted completely. Moreover, it is also hard to generate proper background images for tube stitching. For example, when the elevator door is open or the horizontal sliding door is open in Fig. 1(a), they are all judged as foreground. The condensation results are shown in Fig. 1(b), where we can see that the visual effects are not acceptable when the moving doors are directly stitched into the backgrounds. Therefore, this kind of background complexity analysis is critical to improve the quality of video condensation. To deal with this problem, we adopt a video segmentation method to estimate the background complexity and divide the input video into segments. For these complex segments such as door open in Fig. 1(a), we keep the original video segments into the synopsis. For the other segments, we maximally condense the content and concatenate condensed results with the complex segments.

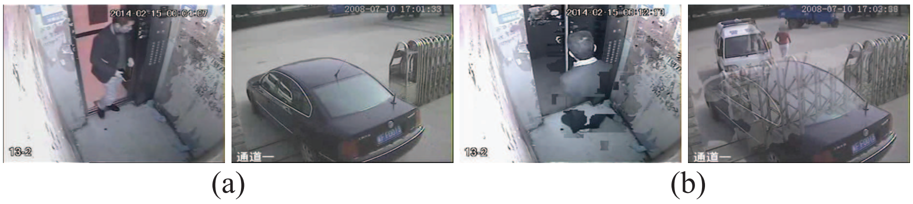

(a) (b)

Fig. 1. Examples of complex backgrounds. (a) Original videos: elevator, sliding door. (b) The condensation results with online video synopsis.

For complex scenes, the integrity of moving objects and the continuity of object trajectories are also important factors for the condensation quality. Therefore, to adapt various changes in complex scenes, we introduce a self-adaptive background modeling approach based on sample consensus with a flexible feedback scheme to automatically adjust the model parameters. Besides, a selective diffusion method is employed to overcome the problems like incomplete foregrounds or false detections brought by intermittent moving objects. For the continuity of object trajectories and reducing blinking effect in a condensed video for better visual effects, sticky tracking was proposed in [11] to merge nearest object cubes before tracking. This method reduces the blinking effect caused by occlusions between objects, however, it also sticks patches caused by background noise and fragments of other objects. Therefore, we argue that it is more reasonable that we generate trajectories by concatenating moving objects in consecutive frames then stick trajectories with overlapping objects. In this way, not only the blinking effect is reduced but also fragments with noise and other objects are removed.

Above all, in this paper we propose a piecewise condensation approach based on the analysis of scene complexity. Based on the background complexity analysis, we divide the input video into several segments. Then we utilize a self-adaptive background modeling approach with a feedback scheme and a selective diffusion strategy to keep the integrity of foreground objects, followed by a sticky trajectory strategy to remove noisy fragments and reduce blinking effect. Finally, we employ an adaptive truncation approach to make the condensed video more compact. The contributions of this paper are summarized as follows:

– We propose a piecewise video condensation approach for complex scenes by dividing the input video into different segments based on the analysis of background complexity and adopting a divide-and-conquer condensation strategy.
– To keep the integrity of moving objects and the continuity of object trajectories, we introduce a self-adaptive background modeling approach with a flexible feedback and selective diffusion scheme.
– We put forward a sticky trajectory strategy to remove noisy fragments and reduce blinking effect.
– We employ an adaptive truncation approach in the process of piecewise optimization to make the condensed video more compact.

2 Related Work

There has been an increasing interest in video presentation and summarization for along time, which is critical for video storage, browsing and indexing.

For video summarization, key frames were usually selected to form a new representative image. Based on maximum frame discrepancy strategies, key frames were usually selected in these approaches. Fast forwarding [1] or video skimming [2] was one efficient video browsing solution, its idea was selecting some representative frames as key frames to replace the whole video, the remaining frames were skipped. For this technique, Choosing the frames with high interest or high activity adaptively adaptively was not an easy task. Some adaptive methods for choosing key frames were proposed [3], but the biggest problem was the big loss of information, especially the fast activities during the dropped frames.

Some researchers generated new images using regions of interest(ROI) beyond the whole frames. For example, video mosaic [13] was a synthetic representation by stitching successive video frames, covering more comprehensive information than a single key-frame. Another typical work was video collage [14]. A video sequence was compacted to get a single image by seamlessly arranging ROIs on a given canvas. Storyboards [15] and narratives [16] represented the course of events by a static image with an explicit temporal cue.

Video montage [5], analyzed both the spatial and temporal information, extracted the informative portions in input videos and condensated them together, its condensate rate was high but caused visual unpleasant and the total loss of context. The ribbon carving method condensed video through removing the ribbons without activities in every frame [9]. It achieved low condensation ratio and also lost a lot of context information. In dynamic video narrative [17], all duplication specific objects were seamlessly stitched into the background video according to its time axis. In terms of a high condense rate, video synopsis [8, 7] had made a big success and attracted the attention of many researchers. Feng et al. [10] proposed an online method, in which tubes were filled in a spatio-temporal volume one by one like playing a Tetris game. However, motion structure was not considered, as well as the time consistency of tubes in their method. Huang et al. [18] regarded the synopsis video generation problem as a maximum a posteriori (MAP) estimation problem, where the appearing frames of object instances chronologically rearranged in real time according to an online updated synopsis table. In [11], the optimization problem of tube rearrangement was transformed into a stepwise optimization problem and used Graphic Processing Unit (GPU) and multicore technique to further improve the speed.

3 Overview

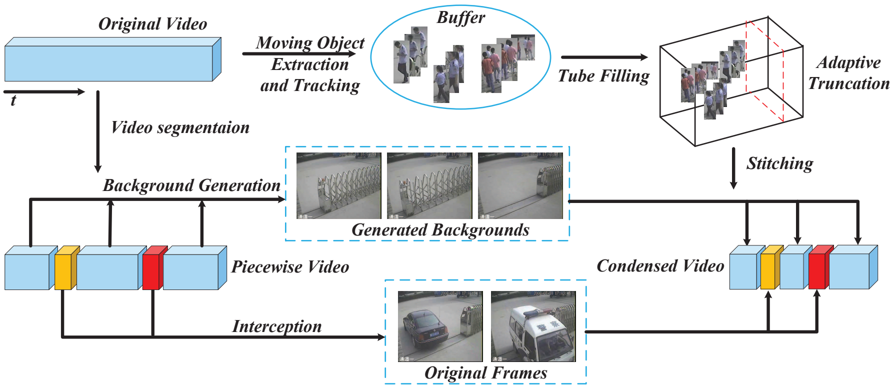


Fig. 2. Overview of piecewise video condensation.

As illustrated in Fig 2, the proposed approach includes video segmentation, background image generation, piecewise condensation, adaptive truncation and object stitching. Firstly, based on the background complexity analysis, we divide the input video into several segments. Then we introduce sample consensus model like ViBe [19] to extract moving objects with a feedback scheme and a selective diffusion strategy to keep the integrity of foreground objects, followed by a sticky trajectory method instead of sticky tracking [10] to remove noisy fragments and reduce blinking effect. Furthermore, an adaptive truncation strategy is introduced to raise the condensation ratio and improve the visual quality. And object tubes are stitched into the generated backgrounds with a modified Poisson editing [7]. In the followings, we will introduce these stages in details.

4 Video Segmentation based on Background Complexity

The complexity of the background in videos can be reflected by temporal changes of background median. Videos are segmented depending on the difference among temporal medians of video clips. The segment process is shown in Fig. 3.

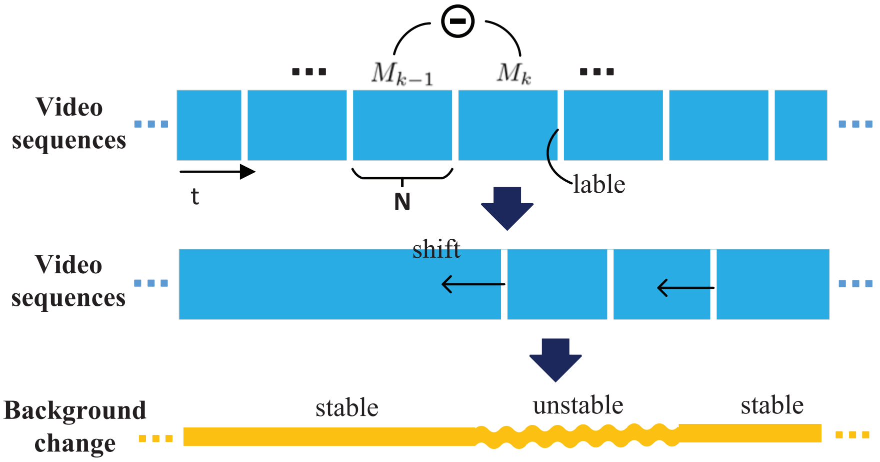


Fig. 3. Video segmentation based on background change.

Firstly, video sequences are divided into several groups in time order, each group contains N frames. Temporal medians are computed in each group, one temporal median frame M_k corresponds to one group k. The last frame id of each group is selected as a label to each group. Let M_{k-1} and M_k be two neighboring temporal median frames, L_{k-1} and L_k are their labels respectively. We compute the frame difference between (in Fig 3, $\ominus$ stands for this process) each pair of temporal median frames in time order. We use D_k to represent the distance between M_{k-1} and M_k. If D_k exceeds the threshold D_{thre}, we consider the background changes a lot, L_k will be recorded, otherwise, M_k will be updated for next computation $\ominus$:

$$\begin{cases} M_k = (1-\alpha)M_{k-1} + \alpha M_k \\ \alpha = \lambda D_k / H_{thre} \end{cases} \tag{1}$$

where α is the updating rate ($\alpha \in [0,1]$), the updating process can help detect some slow changes of background (i.e. illumination variation). According to the distribution of recorded labels in the video, the whole video is segmented into stable periods and unstable periods. Since the temporal median method has hysteretic effects, we set the labels to shift $1.5N$ before. The temporal window size N decides the sensitivity of background change detection. We have tried different values of N on lots of videos and found that $N = 300$ is a proper value choice. Fig. 4 shows background changes reflected by temporal median difference method on different videos and foreground areas measure the distance D_k.

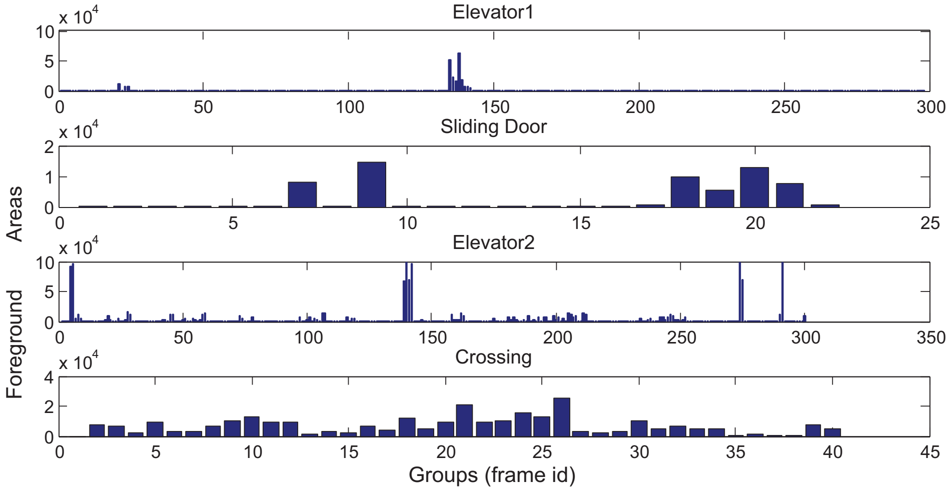


Fig. 4. Background changes reflected by temporal median difference in different complex scenes.

5 Objects Extraction and Sticky Trajectory

5.1 Objects Extraction

To the frames with stable background, we adopt an improved background modeling approach with self-adaptive ability illustrated in Figure 5. Based on the frames of input video sequences, firstly, we build a sample based background model. Like Vibe [20], each pixel model is represented by a sequence of historical samples based on sample consensus. The background model, named B, is formed by a series of pixel models, each of which contains a set of N recent background samples:

$$B(x) = \{B_1(x), B_2(x), ..., B_N(x)\} \tag{2}$$

In our method, we fixed $N = 35$ to strike a balance between accuracy and speed.

$$S_t(x) = \begin{cases} 1 \; if \# \{dist(I_t(x), B_n(x)) < R, \forall n\} < \#min \\ 0 \qquad\qquad\qquad\qquad otherwise \end{cases} \tag{3}$$

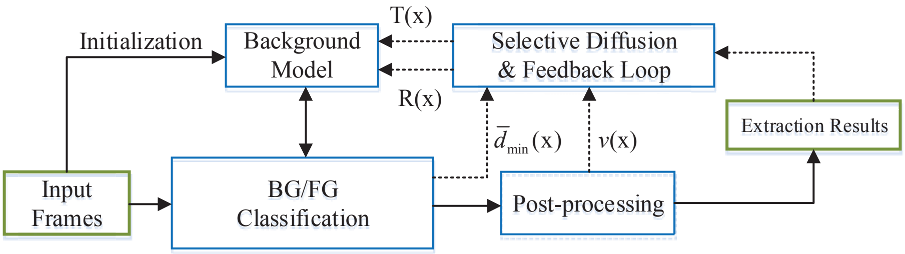


Fig. 5. Moving object detection.

where S_t is the segmentation result, $dist(I_t(x), B_n(x))$ measures the distance between a given background sample and corresponding current observation. R is the distance threshold and $\#min$ is the minimum number of matches required for a background classification. We set $\#min$ to 2 as it was demonstrated in [20].

To adapt various changes in complex scenes, a flexible feedback scheme is presented to automatically adjust the model parameters. Decision threshold R and update rate T are two most important parameters in this modeling process. we consider R and T as two pixel-level state variables. Two frame-size maps are defined to store the current value of R and T. In feedback loops, they are decided by recursive moving average map $\bar{d}_{\mathrm{min}}$ and blinking pixels accumulators v. The recursive moving average map $\bar{d}_{\mathrm{min}}$ is a measure of background dynamics, it is calculated by the distance between samples and current observations.

Then, a selective diffusion method is employed to overcome the problems like incomplete foregrounds or false detections brought by intermittent moving objects. For those intermittent motionless foreground objects, the pixels in their area are different from surrounding background pixels, so we prevent the diffusion from surrounding background to the foreground. For those sudden moving of background objects, the background area that they leave behind is similar with surrounding background, so we accelerate the diffusion from surrounding background to the foreground. So far, we extract moving foreground objects from video sequences frame by frame.

Because the context information in generated background images is corresponding to the original frames which is used to extract tubes, bounding boxes of extracted objects can be used for computing energy cost and object stitching directly, without any additional segmentation method like graph cut [11], which can accelerate the condensation process.

5.2 Sticky Trajectory

After moving objects extraction, we concatenate moving objects in consecutive frames together to obtain object trajectories. Although many tracking methods have been proposed, those methods may be difficult to suit for video condensation. The reason is that the break of object trajectory will cause blinking effect

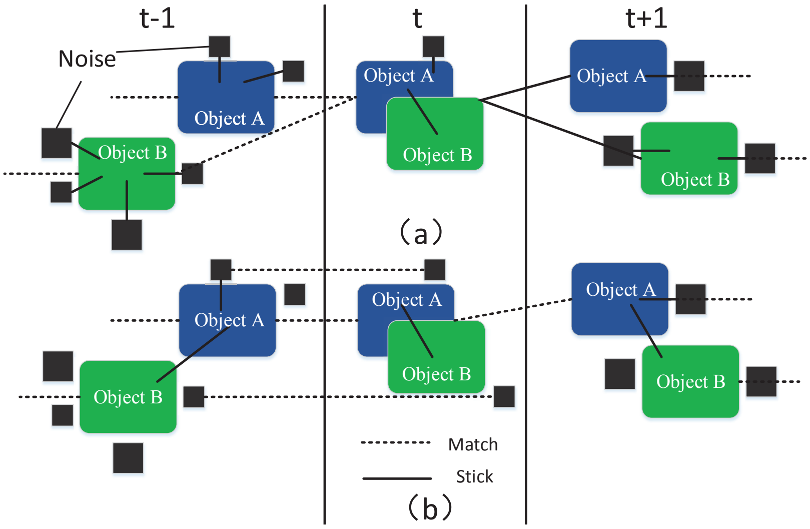


Fig. 6. Comparison between sticky tracking [11] and sticky trajectory. (a) sticky tracking, (b) sticky trajectory.

in condensed video. For example, if there are two objects in a video: object A and B. When a part of object A is occluded by object B at frame t, object A will loses this part in condensed video at some point. If worse, object A is occluded by object B at frame t totally, then object A will disappear abruptly and then appear again in the view. To reduce blinking effect in a condensed video for better visual effects, sticky tracking was proposed in [11] shown in Fig. 6(a). In [11], if occlusions happen to two or more object tubes, they will be merged into a single tube, as if they are sticking together. The key point is to launch merging before matching. This method reduces the blinking effect caused by occlusions between objects, but it also sticks noise caused by dynamic background and fragments of other objects. These noise also increases the number of foreground objects, which will take more time for optimization and reduce the compression ratio.

Therefore, instead of sticky tracking [11], we adopt a sticky trajectory approach to not only reduce the blinking effect but also remove fragments with noise and other objects. As shown in Fig. 6(b), different with [11], we generate trajectories by concatenating moving objects in consecutive frames at first. Then, we remove the noise trajectories that are very short because noise often abruptly appears and disappears. Finally, we stick trajectories that have overlapping objects at the some point. In this way, trajectories are generated before sticky process, which benefits removing noisy fragments caused by dynamic background and other objects easily. In addition, the compression ratio is increased with better visual effects.

6 Piecewise Condensation

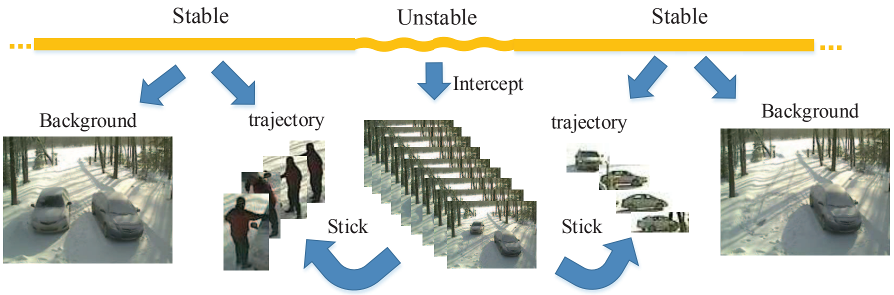


Fig. 7. Background image generation and piecewise condensation.

For the segmented videos with stable backgrounds and unstable backgrounds, we adopt a divide-and-conquer strategy for video condensation. In the following steps, we detail the background image generation and energy minimization:

Background generation. As shown in Fig .7, based on the video segmentation, object extraction and sticky trajectory, background images can be generated respectively using temporal median method in each segment with stable background. Tubes from different segments will be stitched into corresponding background images, so as to keep the context information completely.

Energy Minimization. A divide-and-conquer strategy is adopted for video condensation. We keep the original video segments with unstable backgrounds while maximally condensing the other segments. In order to achieve visually pleasing condensed results, the problem of rearrangement of tubes is formulated as an energy function, visual overlap and lost of information are defined as energy cost [7]. Therefore, the task of video synopsis is to solve a problem of energy cost minimization. We use online strategy to transform the global cost minimization to a stepwise optimization, to make condensation faster and ensure low memory cost. Let B denote the set of tubes, the stepwise optimization is solved by a greedy algorithm [10, 21]:

$$\begin{aligned} {l_i}^* &= \arg\min_{l_i} E(l_i) \\ s.t. E(l_i) &= E_a(l_i) + \sum_{j \in B} E_c(l_i|l_j) \end{aligned} \tag{4}$$

where $l = \{l_i\}_{i=1}^{|B|}$ is the set of play start time of tubes, every tube T_i have one corresponding l_i. The $E_a(l_i)$ named activity cost measures the cost of extracting tube T_i from original video and stitching it into a generated background at time

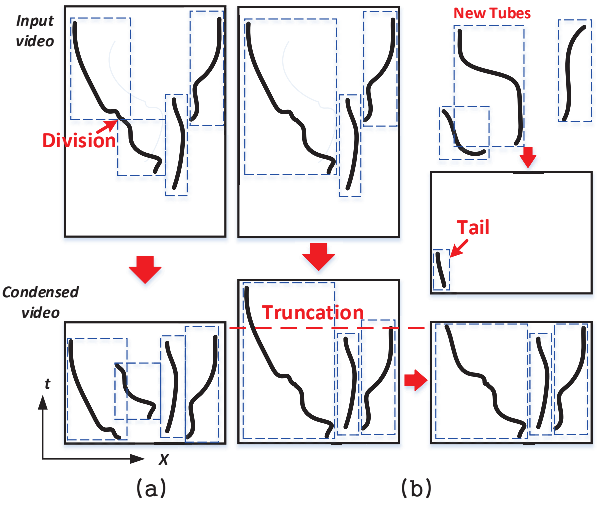


Fig. 8. Adaptive truncation.

l_i. We mainly consider the collision cost between each of the two tubes $E_c(l_i, l_j)$. In online framework, tubes are filled in a condensed space one by one, so simply regard $E_c(l_i, l_j) = E_c(l_i|l_j)$. The greedy algorithm may lead to decreasing of condensation ratio because it condenses the video on finding locally optimal solution in each synopsis clip. The adaptive truncation could decrease this kind of impact.

6.1 Adaptive Truncation

Tubes extracted from original video will be rearranged to give new time labels. Fig. 8 is the top view of tube filling in condensation space. Fig. 8(a) shows the traditional way of tube filling [7]. Tube division is adopted to improve the condensation ratio but causes blinking effect in condensed video (object appears and disappears suddenly in the middle of frames). For a high visual quality, instead of segmenting tube, we truncate condensation space to improve the condensation ratio. As shown in Fig. 8(b), because tubes are always with different temporal length, the compactness in the latter part is very low. In the process of adaptive truncation, the former space with higher compactness will be condensed first. Tubes already filled in the final volume are truncated into two parts, i.e., the body parts and tails. When new tubes come, the former tails are filled into the condensation space firstly, then they will be optimized together with new coming tubes. The temporal truncated location in condensation space can be estimated by mean length approximately:

$$T_{tra} = \frac{1}{n} \sum\nolimits_{i=1}^{n} (l_i + L_i) \tag{5}$$

where l_i is the start time label of tube T_i, and L_i is its length. The adaptive truncation breaks one-off process of tube filling into several steps and makes the condensed video more compact in each step. That would help to alleviate the problem brought by greedy optimization through considering a part of future tubes. Moreover, the discontinuity of tubes between condensation space is avoided [21]. Fig. 9 shows the comparison results between condensation with adaptive truncation (AT) and without adaptive truncation. Condensation result with AT shows more compact.

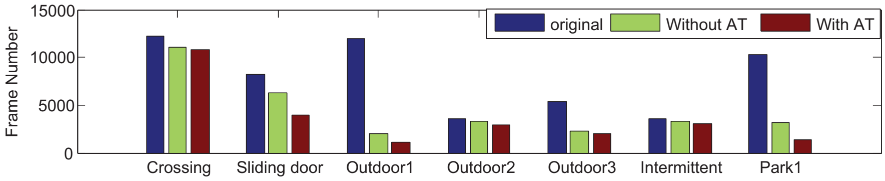


Fig. 9. Comparison results with adaptive truncation and without truncation.

7 Experiments

To evaluate the performance of the proposed approach, we carried experiments on two video datasets. One is a public dataset from [11], composed by 9 videos captured from indoor and outdoor scenes. Another one was collected by ourselves, including 9 complex scenes ("intermittent" is captured by moving camera). Table 1 presents the results of our approach. The condensation ratio (CR) denotes the frame number ratio between condensed videos and original videos, and AoMU, PoMU are abbreviations of average of memory usage and peak of memory usage, respectively.

Condensation ratio. As shown in Table 1, the lowest condensation ratio is 1.16 while the highest one is 32.4. The lower condensation ratio generally results from complex background changes, which are truncated as unstable video segments.

Speed. As shown in Table 1, the speed decreases with the increase of the the pixel resolution. For the video sequence with solution (320 × 240), the processing speed is about 100 fps. For high resolution (740 × 576) video sequences, the process speed still has about 41 fps.

Memory usage. For those high resolution (740 × 576) video sequences, the memory usage peak of our system is lower than 2.0 GB.

Subjective evaluation. The robustness of video condensation can be reflected by the subjective evaluation of condensation quality for different scenes. Two criteria are proposed for evaluating the visual quality of condensed video, including visual pleasing and comprehensible:

1. Visual pleasing: Do you think this synopsis is comfortable for viewers ? You can score based on the following aspects: Overlap, blinking effect and object

Table 1. Our condensation results on 18 videos.

Video	Resolution	#Frame (Num)	Speed (fps)	CR	AoMU (MB)	PoMU (MB)
Overpass [11]	320 × 120	23950	103.1	20.91	150	167
Exit [11]	320 × 240	81538	100.6	18.36	293	313
Garden [11]	320 × 240	33826	100.9	13.68	280	284
Outdoor [11]	320 × 240	138583	98.4	2.99	316	348
Park1 [11]	352 × 288	10221	109.9	7.70	372	374
Passage [11]	352 × 288	51041	98.9	10.01	377	395
Street [11]	704 × 576	100114	42.5	8.48	1465	1530
Staircase [11]	704 × 576	46109	41.5	3.76	1466	1545
T-junction [11]	704 × 576	637470	43.6	32.40	1493	1899
Elevator1	352 × 288	89992	93.2	15.00	371	374
Elevator2	352 × 288	90001	94.0	3.47	401	439
Crossing	740 × 576	12266	42.2	1.13	1508	1959
Slidingdoor	352 × 288	8244	107.1	2.10	378	390
outdoor1	480 × 360	11999	82.8	11.46	660	716
outdoor2	724 × 416	3495	74.2	1.21	1082	1085
outdoor3	352 × 288	5323	102.3	2.70	385	410
Irondoor	352 × 288	90001	97.3	2.93	385	410
intermittent	568 × 376	3500	40.22	1.16	762	771

completeness.
2. Comprehensible: Can you infer the original object behavior information from this synopsis ?
3. Overall Satisfied: Do you think the synopsis overall is comfortable for viewers ?
We set the score scale as 1-5 for these two criteria, where the score below 3 means the visual quality of this condensed result is bad and the score of condensed results with accepted quality is above 4. We invited 36 participants to score the synopsis results. All the participants have strong background knowledge in video surveillance, they were requested to watch the original videos first, then compare the condensed results with original videos and give their scores on two aspects: visual pleasing and comprehensible. We compared average scores by two criteria between our method and online video condensation (OVC) method [11]. The statistical results of subjective feedbacks on two datasets are illustrated in Fig 10. The score of our approach is close to OVC in dataset1 [11] because most videos have relatively simple backgrounds. But in the dataset of complex scenes, our method achieves better performance on visual quality.[1] The scenes "intermittent" captured by intermittent moving camera also can be condensed with a high visual quality by our approach.

Table 2 gives a summary of comparison between our approach and other state-of-art video condensation methods. It shows that compared with video

[1] http://pan.baidu.com/s/1i329jzn

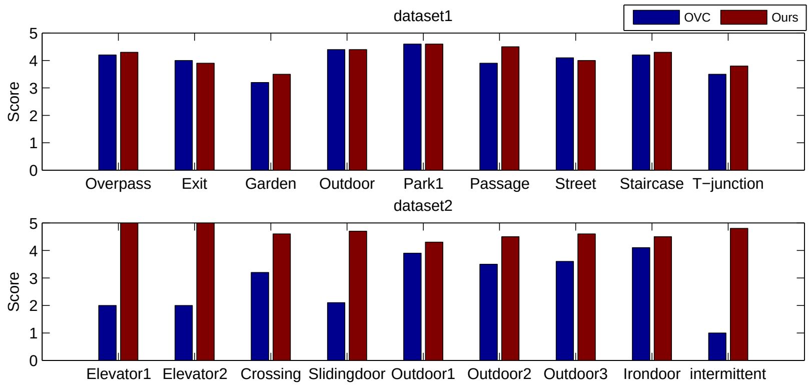


Fig. 10. Comparison results of Subjective evaluation between our method and OVC [11].

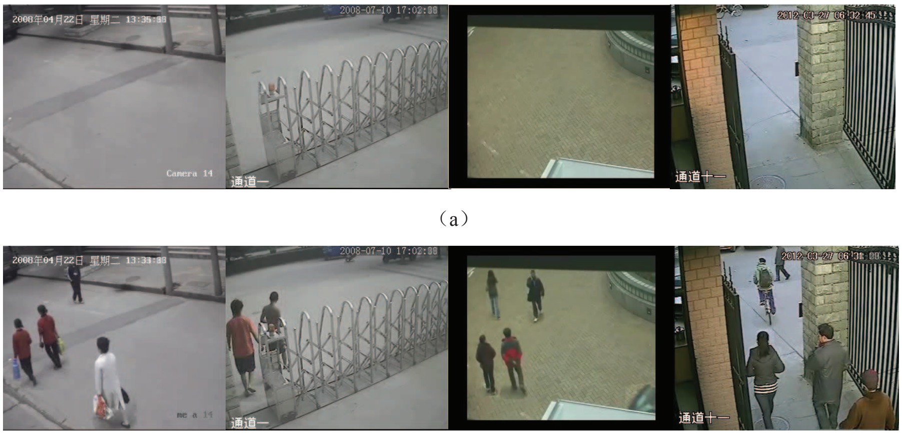
Table 2. Comparison results with the state-of-art methods.VS:Video Synopsis, R-C:Ribbon Carving, CR:Condensation Ratio, OVC:Online Video Condensation.

Method	Speed	Memory	CR	Blinking Effect	Robustness
VS [7]	10fps	High	User	High	Low
RC [22]	Slow	Huge	Low	-	-
OVC [10]	100fps+(GPU)	Low	High	Low	Low
Ours	40fps+(CPU)	Low	High	Low	High

synopsis [7] and Ribbon carving [10], our method has faster speed with lower memory, better visual quality. We cannot directly compare our processing speed with [11], since they run in GPU. But compared to online video condensation [11, 22], our approach is more robust to different scenes. Example frames of the condensation videos are shown in Fig. 11.

8 Conclusion

A robust condensation approach based on piecewise condensation framework has been proposed in this paper. The piecewise condensation framework can condense results with high visual quality in different scenes, even the sequence captured by intermittent moving camera. We divide the input video into several clips. Then we present a self-adaptive background modeling approach with a feedback scheme and a selective diffusion strategy to keep the integrity of foreground objects, followed by a sticky trajectory strategy to remove noisy fragments and reduce blinking effect. The process of condensation is with high speed and low memory cost. Besides, an adaptive truncation is designed to refine the low condensation ratio brought by online tube filling. Experimental results show the superiority of the proposed approach.



(a)

(b)

Fig. 11. Example frames of condensation video in four different scenes, which are outdoor3, Slindingdoor, outdoor1, Irondoor, from left to right respectively. (a) one frame in input video, (b) one frame in our condensation video.

9 acknowledgment

This work was supported by 863 Program 2014AA015104, and National Natural Science Foundation of China 61273034, and 61332016.

References

1. Petrovic, N., Jojic, N., Huang, T.S.: Adaptive video fast forward. Multimedia Tools and Applications **26** (2005) 327–344
2. Smith, M.A., Kanade, T.: Video skimming and characterization through the combination of image and language understanding. In: Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop on, IEEE (1998) 61–70
3. Höferlin, B., Höferlin, M., Weiskopf, D., Heidemann, G.: Information-based adaptive fast-forward for visual surveillance. Multimedia Tools and Applications **55** (2011) 127–150
4. Kim, C., Hwang, J.N.: An integrated scheme for object-based video abstraction. In: Proceedings of the eighth ACM international conference on Multimedia, ACM (2000) 303–311
5. Kang, H.W., Chen, X.Q., Matsushita, Y., Tang, X.: Space-time video montage. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Volume 2., IEEE (2006) 1331–1338
6. Pritch, Y., Rav-Acha, A., Gutman, A., Peleg, S.: Webcam synopsis: Peeking around the world. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE (2007) 1–8

7. Pritch, Y., Rav-Acha, A., Peleg, S.: Nonchronological video synopsis and indexing. Pattern Analysis and Machine Intelligence, IEEE Transactions on **30** (2008) 1971–1984
8. Rav-Acha, A., Pritch, Y., Peleg, S.: Making a long video short: Dynamic video synopsis. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Volume 1., IEEE (2006) 435–441
9. Li, Z., Ishwar, P., Konrad, J.: Video condensation by ribbon carving. Image Processing, IEEE Transactions on **18** (2009) 2572–2583
10. Feng, S., Lei, Z., Yi, D., Li, S.Z.: Online content-aware video condensation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2082–2087
11. Zhu, J., Feng, S., Yi, D., Liao, S., Lei, Z., Li, S.: High performance video condensation system. IEEE Transactions on Circuits & Systems for Video Technology **25** (2015) 1113–1124
12. Sun, L., Xing, J., Ai, H., Lao, S.: A tracking based fast online complete video synopsis approach. In: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE (2012) 1956–1959
13. Irani, M., Anandan, P.: Video indexing based on mosaic representations. Proceedings of the IEEE **86** (1998) 905–921
14. Liu, X., Mei, T., Hua, X.S., Yang, B., Zhou, H.Q.: Video collage. In: International Conference on Multimedia 2007, Augsburg, Germany, September. (2007) 461–462
15. Fu, W., Wang, J., Zhao, C., Lu, H.: Object-centered narratives for video surveillance. **8556** (2012) 29–32
16. Goldman, D.B., Curless, B., Salesin, D., Seitz, S.M.: Schematic storyboarding for video visualization and editing. Acm Transactions on Graphics **25** (2006) 862–871
17. Correa, C.D., Ma, K.L.: Dynamic video narratives. Acm Transactions on Graphics **29** (2010) 2010
18. Huang, C.R., Chung, P.C.J., Yang, D.K., Chen, H.C., Huang, G.J.: Maximum a posteriori probability estimation for online surveillance video synopsis. IEEE Transactions on Circuits & Systems for Video Technology **24** (2014) 1417–1429
19. Barnich, O., Van Droogenbroeck, M.: Vibe: A universal background subtraction algorithm for video sequences. Image Processing, IEEE Transactions on **20** (2011) 1709–1724
20. Olivier, B., Marc, V.D.: Vibe: a powerful random technique to estimate the background in video sequences. In: ICASSP. (2009) 945–948
21. Fu, W., Wang, J., Gui, L., Lu, H., Ma, S.: Online video synopsis of structured motion. Neurocomputing **135** (2014) 155–162
22. Van Droogenbroeck, M., Paquot, O.: Background subtraction: Experiments and improvements for vibe. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, IEEE (2012) 32–37