

Content Aware Video Condensation Using Low Cost Ribbon Carving Technique

Suhandas ¹, Dr. Linganagouda Kulkarni ²

Assistant Professor, Department of ECE, VCET, Puttur, DK Karnataka, India¹

Professor, Department of CSE, BVBCET, Hubli, Karnataka, India²

ABSTRACT: The amount of captured video is growing with the increased numbers of video cameras, especially the increase of millions of surveillance cameras that operate 24 hours/day. Since video browsing and retrieval is time consuming, most captured video is never watched or examined. Efficient browsing of long video sequences is a key tool in visual surveillance. Video condensation aims at providing condensed representations of video datasets that can be easily captured from digital cameras. Video condensation, that is proposed here, is novel in the way information is removed from the space-time video volume is conceptually simple and relatively easy to implement. Ribbons are recursively carved out by minimizing an activity-aware cost function. The ribbon model is flexible and permits an easy adjustment of the compromise between temporal condensation ratio and anachronism of events. And also sliding-window ribbon carving is proposed to handle streaming video. Video condensation can be applied to create a synopsis of endless video streams, as generated by webcams and by surveillance cameras.

KEYWORDS: Ribbon model, Video condensation

I. INTRODUCTION

Efficient browsing of long video sequences is a key tool in visual surveillance. Millions of cameras have been deployed in transportation hubs (e.g., airports, railway and bus stations), on city streets, highways and in office buildings producing over four billion hours of video footage each week. They generate vast amounts of visual data, which makes continuous monitoring by human operators impractical due to high cost and reliability issues (e.g., operator fatigue). Therefore, there is a need to develop automatic algorithms to aid human operators, by identifying relevant data segments and eventually to condense them. In video condensation the information is removed from the space-time video volume. In this work we focus on one class of video condensation algorithms that compute a summary of video activities in the field view of a surveillance camera. The resultant video is called as video digest or video synopsis. A video digest is an abbreviated video that preserves the most relevant activities while removing relatively static temporal and/or spatial segments. A video digest permits fast browsing through recorded surveillance video, thus increasing efficiency in video forensics, but it can also be used for quick review of motion pictures or home videos. The schematic video clip representation is shown in Figure 1.

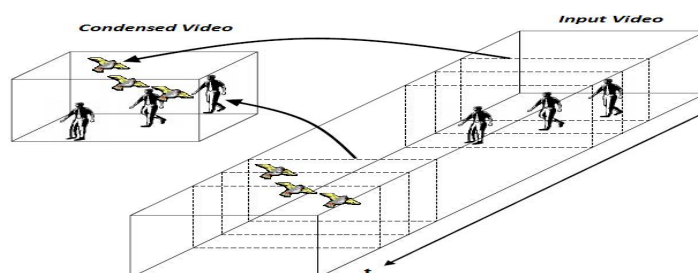


Figure 1: The input video shows a walking person and after a period of inactivity displays a flying bird. A compact video condensation can be produced by playing the bird and the person simultaneously [11].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 9, May 2016

II. LITERATURE SURVEY

Various approaches to compute video condensation have been proposed are:

- Fast forwarding, where individual frames or groups of frames are skipped in fixed or adaptive intervals [8].
- Video summarization, where key frames are extracted and usually presented simultaneously as a storyline [4].
- Video montage, where relevant spatio-temporal (3-D) segments are extracted and combined into a spatio-temporal puzzle [6].

While fast forwarding is conceptually simple, it has limited capabilities since only complete frames can be removed. As a consequence, the video condensation ratio defined as the ratio of video length prior to processing to that after processing, is relatively low. The computation of a video digest can be as simple as skipping frames at regular intervals (e.g., every second frame for 2:1 condensation ratio). However, higher condensation ratios of interest here necessitate skipping numerous consecutive frames, thus potentially leading to the exclusion of relevant activities. Although content-adaptive (irregular) skipping of frames has been developed [8], it removes only complete frames that are inactive or exhibit low activity, thus limiting the realizable condensation ratios.

Very high condensation ratios can be achieved by means of video summarization where key frames are extracted and presented sequentially, like a slide show or simultaneously as a storyline. In the former case, temporal continuity of events is compromised due to usually, large temporal gaps between consecutive key frames while in the latter case the time aspect is removed altogether. A related idea is that of video skimming where short representative sub-sequences from the original video are extracted and combined into a dynamic summary. This approach is a special case of content-adaptive skipping with the same deficiency of removing only complete frames which limits the attainable degree of condensation.

The approaches discussed above preserve complete video frames. An alternative is extracting spatial segments from different frames to combine them into a new entity. The video condensation approach proposed here is closest to video synopsis as we allow spatial segments from different frames to be combined in the output video while permitting only temporal transformations (spatial location of dynamic events is preserved). At the same time, our approach does not require object extraction instead; we use simple background subtraction to guide the video condensation process. We introduce the concept of a ribbon in space-time video volume as a novel extension of the concept of a seam in 2-D images used for content-aware image resizing [10]. There are two methods by which image resizing can be achieved. That is by content aware image resizing and intensity based image resizing. In content based image resizing the cost of a seam is determined based on the object present that is the object present in the frame will always have high cost than the background. In intensity based image resizing depending on the intensity of each frame the cost is decided. This method has the major drawback in some cases the intensity of the object will be low which in turn determines the object as the low cost seam and it will be deleted. Therefore in the proposed video condensation method content based image resizing is followed. We also introduce a ribbon flex-parameter to trade off the video condensation ratio and anachronism of events and develop a hierarchical sliding-window approach to process endless video in multiple passes.

III. RIBBON CARVING FOR VIDEO CONDENSATION

Video Seams

In video condensation, the length of video is reduced while preserving the activities that appear in video. Thus we extend the idea of a seam to a ribbon; the approach of video condensation results by removing the ribbon with minimum cost. For example if there is a segment of N consecutive video frames where each frame is W pixels wide and H pixels tall. After condensation, the process needs to end up with a new segment of N' consecutive frames with $N' \leq N$. The condensation procedure should hopefully preserve all the important events, e.g., moving vehicles, walking pedestrians and their relative timings while not degrading the quality of the perceived video in a noticeable way. This is achieved by mimicking the idea of seam deletion for image down-sizing.

Deleting the pixels in a video seam should:

- a) Preserve the spatial size of the resulting video segment.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 9, May 2016

- b) Reduce the number of frames in the video segment.
- c) Prevent significant event distortion in either space or time.

The key idea of seam carving for video condensation is to associate reliable activity-aware costs with seams and then recursively remove seams one at a time, recalculating the costs if needed, until a certain user-defined stopping criterion is met. There are two types of ribbon, a vertical ribbon and a horizontal ribbon, a vertical ribbon is a set of pixels (x, y, t). As a result, the vertical ribbon becomes a curve line when we view it from side. Similarly, a horizontal ribbon is a set of pixels (x, y, t). As a result, the horizontal ribbon becomes a curve line when we view it from top.

Finding Minimum-Cost Ribbons

To delete a ribbon from a video segment, the first step is to find the ribbon with minimum cost. The cost of a ribbon is defined as the summation of cost at the individual pixels that constitute the ribbon. Let C be a cost function which associates with each pixel having coordinates (x, y, t) and cost is denoted by C(x, y, t). If R denotes a horizontal or vertical ribbon then the cost of the ribbon is given by equation 1.

$$C(R) = \sum C(x, y, t) \tag{1}$$

Where R denotes a ribbon. Therefore from a given video both least-cost vertical and horizontal ribbons can be found. Then compare the minimum cost and remove the ribbon which has smaller cost. The process is recursively carried out to carve the ribbon with minimum cost until some stopping criterion is met.

Cost Functions

In video condensation, the temporal dimension is decreased while preserving events and their relative timings. The most significant events are those due to the motion of objects. Object motion expresses itself in the form of intensity and color changes at the resolution of an individual pixel. Thus, a good pixel-level cost function should reliably indicate motion-related activity.

A crude indicator of such activity is the local intensity gradient. If I_x , I_y and I_t denote local estimates of horizontal, vertical and temporal derivatives of intensity at (x, y, t) respectively, one possible cost function is the length of intensity gradient in space-time (3-D) that is given in equation 2. That is as shown

$$C(x, y, t) = \sqrt{(I'_x(x, y, t))^2 + (I'_y(x, y, t))^2 + (I'_t(x, y, t))^2} \tag{2}$$

This can be achieved by Sobel edge detection algorithm that computes edge detection and computes vertical and horizontal gradient. Due to the contribution of spatial derivatives the cost will be high in moving areas and also in textured still areas.

Sliding-window Approach

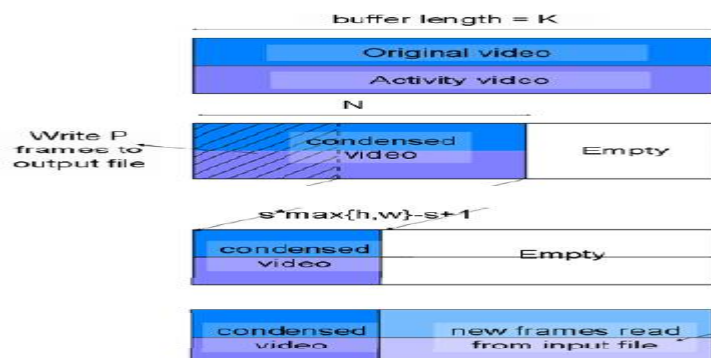


Figure 2: Procedure of sliding-window algorithm for processing streaming video

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 9, May 2016

To apply ribbon carving algorithm to streaming video efficiently and overcome the limitation of memory, we adopt a sliding-window approach. The basic idea is that we use a sliding-window buffer to process a segment of video at a time before it can save some frames if appropriate and read in new frames to process. This way we do not need to process the whole video at a time.

Suppose that we have a sliding-window buffer filled with N frames. After ribbon carving, we have N' frames. Then we compare N' with M_ϕ , where M_ϕ is the maximum possible frame in buffer and ribbon with flex parameter ϕ cannot span more than M_ϕ . If $N' > M_\phi$, then we save the first $N' - M_\phi$ frames, push the remaining M_ϕ frames to the front of the buffer and read in new frames from the video to fill up the buffer. The reason why we can save the first $N' - M_\phi$ frames is that no ribbon from the new frames beyond N' can reach this set of frames. Ribbons passing through the first $N' - M_\phi$ frames will all have cost greater than ϵ . Thus saving these frames will not affect the result of finding the least-cost ribbon. If $N' \leq M_\phi$, then no frames can be saved.

IV. RESULTS

Condensation ratio is defined as the ratio of the length of the original video to the length of condensed video. It depends on several factors, such as how frequently events happen in the original video, the thresholds of background subtraction and ribbon carving, and the flex parameter of ribbons. Parameters used in Table 7.2 are defined as follows:

M = Number of frames to compute the background.

T_1 = Time duration of original video in seconds.

T_2 = Time duration of condensed video in seconds.

$$\text{Condensation ratio} = \frac{T_1}{T_2} \tag{3}$$

Table 1: Comparison of condensation ratios attained

No. of frames to compute the background (M)	Time duration of original video in seconds (T_1)	Time duration of condensed video in seconds (T_2)	Condensation ratio
100	0.35	0.25	1.4
150	0.40	0.25	1.6
200	0.45	0.25	1.8

V. CONCLUSION

The result shows that ribbon carving is effective and efficient, with flex parameters $s = 0$ and $s = 1$ ribbon carving, original video can be condensed by the factor of about 1.5~2. The condensation preserves all the important contents with little degradation of the quality of the perceived video. By using the sliding-window algorithm, the implementation of ribbon carving can handle streaming video. The computational complexity is further reduced by the means of sliding-median filter for background subtraction.

The operator presented for content-aware resizing of images using seam carving technique. Seams are computed as the optimal paths on a single image and are either removed or inserted from an image. This operator can be used for a variety of image manipulations including: aspect ratio change, image retargeting, content amplification

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 9, May 2016

and object removal. The operator can be easily integrated with various saliency measures, as well as user input, to guide the resizing process.

Video condensation is an enabling technology for identification and localization of dynamic scene content, a prelude to scene understanding in video forensic applications where the relative timings of events are crucial. By using a novel flexible-ribbon model, higher condensation ratios can be realized, at the cost of greater event anachronism by increasing the ribbon flex parameter. The best-performing variant of video condensation in this proposal is based on activity (motion) cost rather than intensity-gradient cost.

REFERENCES

- [1] A. Rav-Acha, Y. Pritch, D. Lischinski, and S. Peleg, "Dynamosaicing: Mosaicing of dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1789–1801, Oct. 2007.
- [2] B. Chen and P. Sen, "Video carving," presented at the EUROGRAPHICS, 2008.
- [3] H.-W. Kang, Y. Matsuhita, X. Tang, and X.-Q. Chen, "Space-time video montage," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Jun. 2006, pp. 1331–1338.
- [4] J. Oh, Q. Wen, J. Lee, and S. Hwang, "Video abstraction," in *Video Data Management and Information Retrieval*, S. Deb, Ed. Hershey, PA: Idea Group, Inc./IRM Press, 2004, pp. 321–346, ch. 3.
- [5] J. Konrad, "Videopsy: Dissecting visual data in space-time," *IEEE Commun. Mag.*, vol. 45, no. 1, pp. 34–42, Jan. 2007.
- [6] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "Efficient representations of video sequences and their applications," *Signal Process., Image Commun.*, vol. 8, no. 4, pp. 327–351, May 1996.
- [7] M. Ristivojevic and J. Konrad, "Space-time image sequence analysis: Object tunnels and occlusion volumes," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 364–376, Feb. 2006.
- [8] N. Petrovic, N. Jovic, and T. Huang, "Adaptive video fast forward," *Multimedia Tools Appl.*, vol. 26, no. 3, pp. 327–344, Aug. 2005.
- [9] Rav-Acha, Y. Pritch, and S. Peleg, "Making a long video short: Dynamic video synopsis," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Jun. 2006, pp. 435–441.
- [10] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.
- [11] Y. Pritch, A. Rav-Acha, and S. Peleg, "Non-chronological video synopsis and indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, Nov. 2008.
- [12] Zhuang Li, Prakash Ishwar and Janusz Kornad, "Video Condensation by Ribbon Carving," *IEEE transactions on image processing*, vol. 18, no. 11 2009.