

Online Principal Background Selection for Video Synopsis

Shikun Feng¹, Shengcai Liao¹, Zhiyong Yuan², and Stan Z. Li^{1*}

¹*CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences*

²*School of Computer Science, Wuhan University*

{skfeng, scliao, szli}@nlpr.ia.ac.cn zhiyongyuan@whu.edu.cn

Abstract

Video synopsis provides a means for fast browsing of activities in video. Principal background selection (PBS) is an important step in video synopsis. Existing methods make PBS in an offline way and at a high memory cost. In this paper we propose a novel background selection method, “online principal background selection” (OPBS). The OPBS selects n principal backgrounds from N backgrounds in an online fashion with a low memory cost, making it possible to build an efficient online video synopsis system. Another advantage is that, with OPBS, the selected backgrounds are related to not only background changes over time but also video activities in it. Experimental results demonstrate the advantages of the proposed OPBS.

1. Introduction

Surveillance video cameras are widely deployed for security purposes. They work 24 hours a day and generate huge amount of video data. Manual search and retrieval from such videos are time consuming and even impossible. In order to solve this problem, video synopsis has been introduced in recent years [3, 2, 4, 9, 8] as an effective means for video retrieval and search. It condenses a long video in such a way that moving objects originally in different frames can be viewed simultaneously without occluding each other.

An important step in video synopsis is to generate a time-lapse background video. It is usually achieved by selecting n principal backgrounds from N ones with $n \ll N$. We call this “principal background selection” (PBS). The principal backgrounds should meet two properties [3]: (i) It should reflect background changes over time, such as the alternation of day and night. (ii) It should be related with video activities,

that is, backgrounds containing more moving objects are preferred. A simple solution might be that using time-lapse photography [7], selecting a background every N/n frames. However, such homogeneous sampling cannot meet the property (ii) of PBS. In fact, the two properties may conflict with each other. Pritch [3] proposed a PBS approach by combining two temporal histograms with a weight to meet the two properties as a trade-off solution. However, this method needs to store all N original backgrounds, requiring a huge storage space, and processes them offline.

In this paper, we propose an online PBS method (OPBS) for video synopsis. This online method selects backgrounds by processing a background frame at a time without storage of all N background frames. The selected backgrounds reflect not only background changes over time but also video activities. The OPBS enables an online selection of principal backgrounds with a low memory cost, making it possible to build an efficient online video synopsis system.

The rest of the paper is organized as follows: Section 2 presents the OPBS. Section 3 discusses the process of video synopsis. Section 4 presents experimental results.

2. Online Principal Background Selection

In order to select principal backgrounds online, we record two statistics as our preferences for each background. One is a constant 1, which indicates that each background over time has an equal probability to be selected. The other calculates total pixel numbers of the corresponding foreground as activity information, which means that the selection prefers backgrounds containing more moving objects, as required by property (ii). Then we construct two temporal histograms, namely H_t and H_a , with these two statistics in each bin and normalized respectively. In order to meet the two properties of PBS, we follow Pritch’s work to construct a third weighted histogram $H_{new} = \lambda H_a + (1 - \lambda) H_t$. Then the objective of OPBS is to select n principal

*Stan Z. Li is the corresponding author.

backgrounds online so that the histogram H_{new} is divided into n equal parts, which means that we select one background after a certain equal amount of frames and activities. In the following, we present two algorithms of *OPBS*: without and then with prediction. The latter, based on the former, can better divide the area under H_{new} into n equal parts.

2.1 OPBS without Prediction

Refer to Fig. 1. Suppose n principal backgrounds PB_1, \dots, PB_n have been selected at time t and let CPB denote the candidate principal background that will be considered at an appropriate time. Let S_i be the area between PB_i and PB_{i+1} under H_{new} , and denote $S = \{S_i, i = 1, \dots, n\}$ as the set of all divided areas. The x in Fig. 1 denotes the incoming area buffer. There is no need to store every background actually; instead, we just keep PB_1, \dots, PB_n, CPB and the activity information of new coming frames.

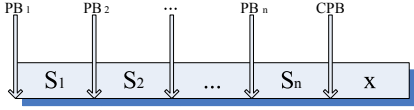


Fig. 1. Illustration of *OPBS*

Denote D_i as the frame set between PB_i and PB_{i+1} , and N_i as the number of frames in D_i . With x we use the notation of D_x and N_x for the same purpose. Also denote $H_a(i)$ and $H_t(i)$ by

$$H_a(i) = \frac{\sum_{j=1}^{N_i} a_{ij}}{a_{all}}, H_t(i) = \frac{N_i}{N_x + \sum_{j=1}^n N_j}, \quad (1)$$

where a_{ij} is the activity information of the frame j in D_i , and the normalization factor a_{all} is the sum of all activity information. Then the area S_i under H_{new} is calculated by $S_i = \lambda H_a(i) + (1 - \lambda) H_t(i)$. The calculation of the area x is similar to S_i . Because both H_a and H_t are normalized histograms, the interpolated histogram H_{new} is also normalized. That is to say, $x + \sum_{i=1}^n S_i = 1$. It should be noted that the values of S_i and x dynamically change over time. In fact, when a new frame arrives, a_{all} will be added by the activity information of this frame, and N_x will be added by 1.

After the calculation of S_i and x , the variance of S will be measured: $var_s = \frac{1}{n} \sum_{i=1}^n (S_i - \mu_s)^2$, where μ_s is the mean of S and it is actually $(1 - x)/n$. For the increasing buffer x , at a certain time we merge it with S in some way as shown in Fig. 2.

There are two choices in front of us. In Fig. 2a, x becomes S'_n , and two adjacent elements of S are merged into S'_j ; while in Fig. 2b, S_n and x are merged

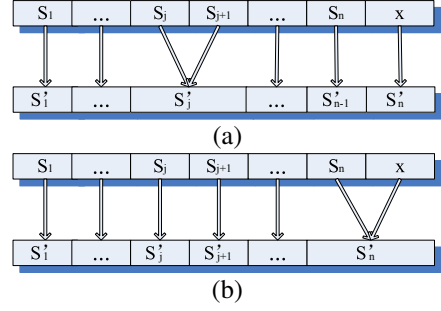


Fig. 2. Two Situations of Merging

into S'_n . x will be set to zero after merging. Obviously the one which has least variance will be chosen: $var_x = \frac{1}{n} \sum_{i=1}^n (S'_i - \mu_{s'})^2$, where $\mu_{s'}$ is the mean of $S'_i (i = 1, \dots, n)$ and it is actually $1/n$.

There are three situations between var_x and var_s . (i) $var_x < var_s$, which means that the merging helps to divide the area under H_{new} into n equal parts. In this case, S will be replaced by S' . (ii) $var_s \leq var_x < \alpha \cdot var_s$, where α is a constant.¹ In this case just let x keep increasing and don't launch the merging (iii) $var_x \geq \alpha \cdot var_s$, which means that var_x becomes undesirably large, so x is forced to merge with S . With this method, var_s always has a low value, so the area under H_{new} can be divided into n equal parts as far as possible.

2.2 OPBS with Prediction

If we analyze the relationship between var_x and var_s , and add the prediction to the method stated above, a better solution will come out. According to the calculation of var_s and var_x , we further have

$$var_s = \frac{1}{n} \sum_{i=1}^n S_i^2 - \frac{1}{n^2} (1 - x)^2, \quad (2)$$

$$var_x = \frac{1}{n} \sum_{i=1}^n S_i'^2 - \frac{1}{n^2}. \quad (3)$$

Let's recall the process of the calculation of var_x . There are two situations to be considered:

- a) S_j is merged with S_{j+1} as shown in Fig. 2a. In this case, $\sum_{i=1}^n S_i'^2$ is given by

$$\sum_{i=1}^n S_i'^2 = x^2 + 2S_j S_{j+1} + \sum_{i=1}^n S_i^2. \quad (4)$$

Combine Equ. (2) (3) (4), we get

$$var_x = \left(\frac{1}{n^2} + \frac{1}{n}\right)x^2 - \frac{2}{n^2}x + var_s + \frac{2}{n}S_j S_{j+1}. \quad (5)$$

¹We find that $1.1 < \alpha < 2.5$ works well.

b) x is merged with S_n as shown in Fig. 2b. This time $\sum_{i=1}^n S_i'^2$ is given by

$$\sum_{i=1}^n S_i'^2 = x^2 + 2S_n x + \sum_{i=1}^n S_i^2. \quad (6)$$

Combine Equ. (2) (3) (6), we get

$$var_x = \left(\frac{1}{n^2} + \frac{1}{n}\right)x^2 + \left(\frac{2}{n}S_n - \frac{2}{n^2}\right)x + var_s. \quad (7)$$

According to Equ. (5) (7), it is shown whatever situation var_x is calculated in, it is approximately a quadratic function of x . We use the word ‘‘approximately’’ because S_j, S_{j+1}, S_n and var_s are actually the functions of x . However, when considering the prediction, they are fixed. Furthermore, a_{all} and $\sum_{j=1}^n N_j$ will become large numbers in Equ. (1) with the online process of growing frames, hence S_i, S_j, S_n and var_s don't change much with x before merging.

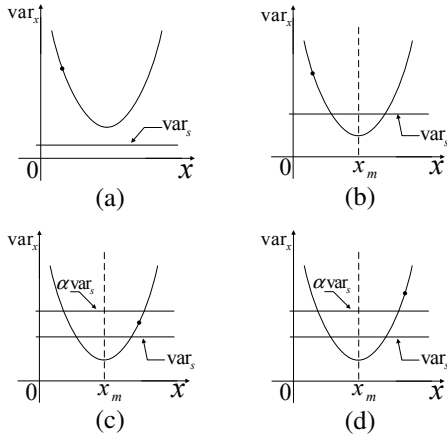


Fig. 3. Four cases of var_x when $var_x \geq var_s$

With Equ. (5) (7), we can predict var_x with x , and adopt appropriate decisions. If $var_x < var_s$, we just merge x with S ; otherwise, there are four cases about var_x as shown in Fig. 3:

- $\min(var_x) \geq var_s$, which means whatever x is, var_x will not be less than var_s . Hence x is merged with S directly.
- $\min(var_x) < var_s, x < x_m$, which means that var_x will be less than var_s as x increases. So this time there is no need to merge x with S .
- $\min(var_x) < var_s, x \geq x_m, var_s \leq var_x < \alpha \cdot var_s$. In this case, the buffer x is always much less than S_i and just let it grow.
- $\min(var_x) < var_s, x \geq x_m, var_x \geq \alpha \cdot var_s$, which means that var_x has no chance to be less than var_s . Hence we force x to be merged with S .

Due to the prediction stated above, the *OPBS* method outputs a more stable result with less var_s .

3. Video Synopsis

The process of video synopsis includes five steps [2]:

- 1) Background Estimation and Foreground Segmentation. Background estimation can be achieved by GMM [5]. In order to get the accurate foreground, the ‘‘graph cut’’ technique [6] is used.
- 2) Background Selection. The job of this step is to generate a time-lapse background video, which is our main contribution in this work.
- 3) Tubes Extraction. It is done by tracking method such as [10].
- 4) ‘‘Play Duration’’ Assignment. In this step, an energy function is minimized to determine when the tubes should appear in the synopsis.
- 5) Stitching the synopsis video. The job here is to stitch the tubes to the selected background. It can be done by Poisson Image Editing [1].

4. Experimental Results

In order to show the efficiency of our algorithm, the experiments are divided into two groups: 1) artificial data; 2) real video data. We apply the same parameters ($\alpha = 1.2, \lambda = 0.35$) for all the experiments.

4.1 Artificial Data

Fig. 4a shows an artificial activity histogram as a mixture of Gaussians. Our goal is to select 50 backgrounds from 4000 frames. The result of selection by *OPBS* with prediction is shown in Fig. 4b. The red vertical lines denote the selection time. It is shown that more backgrounds are selected from active periods while not

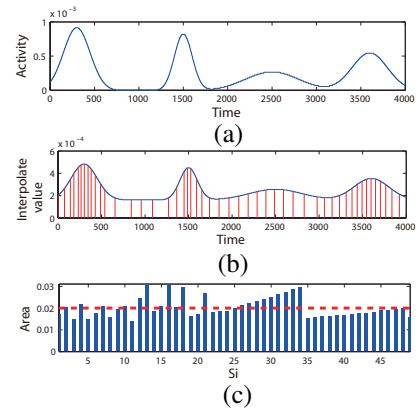


Fig. 4. The Results of *OPBS* with prediction. (a) The input activity histogram as a mixture of Gaussians. (b) Selection result by the prediction method. (c) The area of each S_i .

totally ignoring inactive periods. Fig. 4c represents the area of S_i ($i = 1, \dots, 50$) and the red dashed horizontal line represents the average area of S . We can see that each S_i approximatively shares the same area.

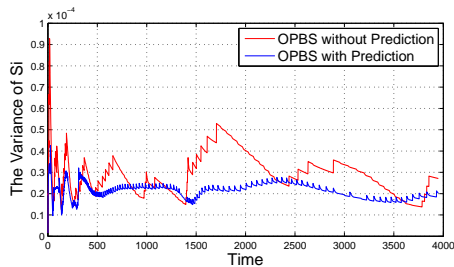


Fig. 5. Comparison between the *OPBS* with and without prediction

Fig. 5 shows the comparison of the *OPBS* with and without prediction. It is shown that the method with prediction has a lower and more stable var_s , which reflects the robustness of the prediction method.

4.2 Real Video Data

Fig. 6a shows the activity information of a real 24 hours street surveillance video (15 fps). The data is normalized for display. It is shown that there are more active objects during the day. This time our goal turns into selecting 50 backgrounds from that of 24 hours! Fig. 6b shows the selection result by *OPBS* with prediction. We can see that more backgrounds are selected from the day time while not totally neglecting the night time. The total processing time for *OPBS* with prediction on the 24 hours video is about 562 seconds at a standard PC hardware (Intel Core2 Duo 3.00GHz CPU, 2GB Memory). Hence the average time for processing a frame is 0.43ms, which is so fast that can be ignored.

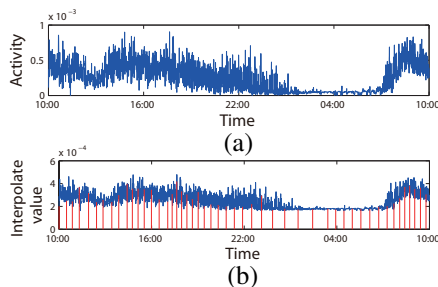


Fig. 6. (a) The activity information of a 24 hours street surveillance video. (b) Selection result by *OPBS* with prediction.

Fig. 7a is from the original street video, while Fig. 7b shows the result of video synopsis with *OPBS*. It is shown that different objects from different time are displayed at the same frame reasonably. Moreover, different backgrounds over time are selected by *OPBS*.



Fig. 7. The result of video synopsis

5. Conclusion

Two versions of online *PBS* method for video synopsis have been presented. Experiments on artificial data and real video data show the validity of our algorithms. Moreover, it can run at low memory cost on on-line videos. As for the future work, we plan to develop an *online video synopsis* system with *OPBS*.

References

- [1] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Trans. Graph.*, 22(3):313–318, 2003.
- [2] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg. Webcam synopsis: Peeking around the world. In *ICCV*, pages 1–8, 2007.
- [3] Y. Pritch, A. Rav-Acha, and S. Peleg. Nonchronological video synopsis and indexing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(11):1971–1984, 2008.
- [4] A. Rav-Acha, Y. Pritch, and S. Peleg. Making a long video short: Dynamic video synopsis. In *CVPR*, pages 435–441, 2006.
- [5] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*, pages 2246–2252, 1999.
- [6] J. Sun, W. Zhang, X. Tang, and H.-Y. Shum. Background cut. In *ECCV*, pages 628–641, 2006.
- [7] R. Vishniac. *ICP Library of Photographers*. Grossman Publishers, New York, 1974.
- [8] U. Vural and Y. S. Akgul. Eye-gaze based real-time surveillance video synopsis. *Pattern Recognition Letters*, 30(12):1151–1159, 2009.
- [9] M. Xu, S. Z. Li, B. Li, X. Yuan, and S. Xiang. A set theoretical method for video synopsis. In *Multimedia Information Retrieval*, pages 366–370, 2008.
- [10] T. Yang, S. Z. Li, Q. Pan, and J. Li. Real-time multiple objects tracking with occlusion handling in dynamic scenes. In *CVPR*, pages 970–975, 2005.