LOCALIZED TEMPORAL PROFILE OF SURVEILLANCE VIDEO

Saeid Bagheri, Jiang Yu Zheng

Dept. of Computer Science, Indiana University-Purdue University Indianapolis
bagheris@cs.iupui.edu jzheng@cs.iupui.edu

ABSTRACT

Surveillance videos are recorded pervasively and their retrieval currently still relies on human operators. As an intermediate representation, this work develops a new *temporal profile* of video to convey accurate temporal information in the video while keeping certain spatial characteristics of targets of interest for recognition. The profile is obtained at critical positions where major target flow appears. We set a sampling line crossing the motion direction to profile passing targets in the temporal domain. In order to add spatial information to the temporal profile to certain extent, we integrate multiple profiles from a set of lines with blending method to reflect the target motion direction and position in the temporal profile. Different from mosaicing/montage methods for video synopsis in spatial domain, our temporal profile has no limit on the time length, and the created profile significantly reduces the data size for brief indexing and fast search of video.

Index Terms— Video profile, video synopsis, video indexing, temporal profile, browsing, retrieval, surveillance video.

1. INTRODUCTION

The amount of recorded surveillance video is growing as cameras are distributed widely. Problems with storage, indexing and retrieval of such video data arise when their growth rate becomes high. While automatic retrieval has not reached a satisfactory accuracy, viewing and analyzing such footage becomes labor-intensive and time consuming. It is crucial to summarize videos in a compact form that is intuitive and preserves most of the information in the video. For surveillance video from a static camera, several works have removed segments without motion events and thus shortened the video length. As a spatial indexing approach for condensing video, video synopsis [1] based on spatial mosaicing (or called montage, onion skinning) methods compose different actions of targets in a single key frame. However, such a method has limitations on representing temporal changes for long videos. It becomes cluttered and confusing as the video length or the number of targets increases [2]. It also fails to present the time instance of actions intuitively, thus it will not serve well for indexing purposes.

We take a different approach for the video indexing problem. Because the space that a surveillance camera covers is fixed, we focus on representing the time progress of dynamic events at critical positions in the video, which is also a temporal index of the video. The previous time line style index of video was not built to the frame granularity [3]. In this paper, we introduce a method that will create a temporally accurate profile of a video in a 2D image while preserving some shape and spatial information for recognition. The important information in a video recorded with a static camera is foreground motion. Therefore, we develop a technique to intuitively present the transitive motion and show their temporal context in video.

For the temporal video profiling of surveillance video, previous effort [4] has realized a line sensor to capture dynamic targets through a monitoring line. For dynamic camera motion, a temporal profile has been generated from a moving sampling line to mainly summarize background [5, 6, 7]. However, the spatial information is lost when a temporal slice is cut from the spatial-temporal volume of video. To solve this problem, this work samples multiple lines at a critical location with a major target flow. We blend multiple slices at related positions into a single temporal profile. Transparency is assigned to indicate the direction of target motion.

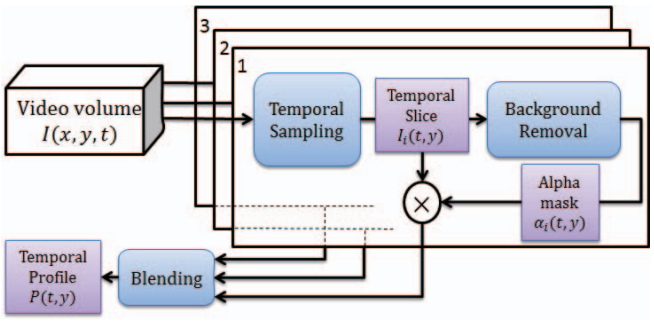


Fig. 1 Framework of generating temporal profile from video.

A diagram of the approach is shown in Fig. 1. The resulting profile is easy for measuring the time instance and duration of actions. Multiple targets can be compared in the time domain as well. By clicking a point in the temporal

This is the author's manuscript of the article published in final edited form as:
Bagheri, S., & Zheng, J. Y. (2014). Localized temporal profile of surveillance video. In 2014 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1–6). http://doi.org/10.1109/ICME.2014.6890143

profile, we can directly locate the video frame that the point was extracted for further examination.

In the following, Section 2 starts from a single sampling line to obtain a temporal slice of moving targets. Section 3 revisits its properties in temporal and spatial domains. Section 4 addresses the multi-line sampling and blending of slices to form a temporal profile. Section 5 is for experiments and discussion.

2. FOREGROUND EVENTS SAMPLING

A surveillance video captured with a static camera is usually aimed at monitoring a wide area and observing "critical" locations where targets pass through. The background is static throughout entire video and its flow is zero in the video volume. However, a moving object with translation motion leaves a trajectory tube non-parallel to the time axis as illustrated in Fig. 2(a). Other motions such as object rotation, human articulate actions, and non-rigid motion of smoke, water, and tree leaves also leaves disturbing traces in the video volume. Our temporal profile proposed in this paper will particularly capture the transitive action at critical positions, because this counts for target in-and-out events in the field of view. Other motion such as waving and body action may also leave signal footage in the profile for further examination in the indexed video frame.

Assume a path exists in the camera view that causes a major foreground flow as in Fig. 2(b). The path direction for targets to pass is possible to be considered as velocity vector $\boldsymbol{m}=(u,v)$. We set a line $\boldsymbol{l}$ in the video frame $I(x,y,t)$, and sample the pixel data on it over consecutive frames ($t=1,2,\ldots$), a *temporal slice,* $I(t,l)$, is obtained from the volume, where l is the coordinate on the sampling line and it is the linear combination of x and y.

If the line orientation is set non-parallel to the motion direction of foreground flow in the image, the foreground will leave some shapes in the temporal slice, otherwise known as flow traces [7]. As a real example, Figure 3(a) shows several frames and an obtained temporal slice in the video volume. This gives the first criterion to set the sampling line.

Criterion 1: *A fixed sampling line to obtain a temporal slice in the video should be set non-parallel to the flow direction of passing targets at critical location.*

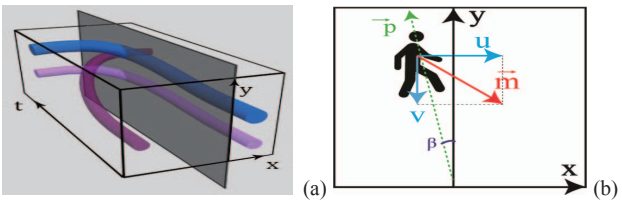


Fig. 2 Sampling foreground motion with a line. (a) An illustration of video volume with foreground flow trajectories (colored tubes). The plane along the time axis is a temporal slice obtained from sampling a pixel line over time. (b) A foreground object passing the sampling line with the moving direction $\boldsymbol{m}=(u,v)$ in the video.

Now, which direction is better to fix the sampling line in the frame after excluding the flow direction? We consider principal direction of targets, $\boldsymbol{p}$, in the image (Fig. 2) for improving the shapes generated in the profile. For example, the principal direction of a person is the up-right pose and it may be slightly slanted in the image (with angle β from an axis) when a camera overlooks a site. Our second criterion to select the direction of sample line is as follows.

Criterion 2: *A sampling line is set to cross the foreground flow in one of the principal directions that is more orthogonal to the major flow direction.*

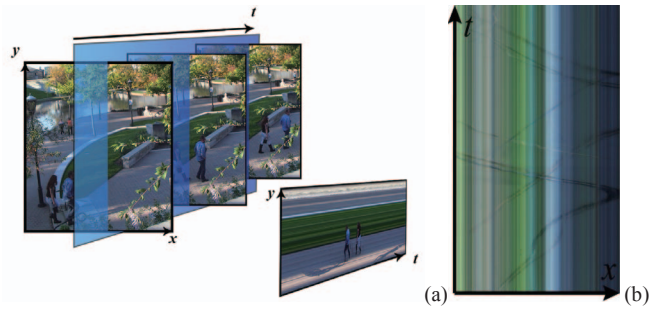


Fig. 3 Motion and shape information in the video. (a) Cutting a temporal slice along the time axis at a location with apparent transitive motion in the video. (b) A *condensed image* of video obtained from averaging the pixel values in y direction in the volume to observe the flow. The traces of people are visible in the condensed image.

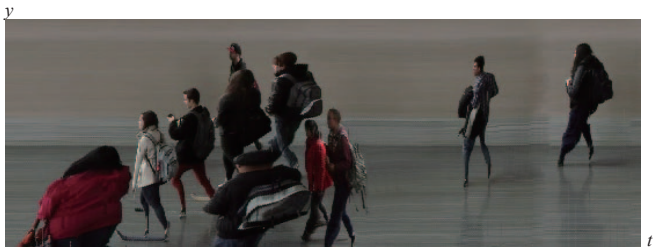


Fig. 4 Original resolution of temporal slice $I(t,y)$ where people are detailed enough for identification.

Although a sampling line is possible to be set orthogonal to the major flow in the view, we select a horizontal or vertical line here to maintain the quality and sampling speed of temporal slices. Without losing the generality, a coordinate system O-xy is set along the selected sampling line $\boldsymbol{l}$ and the temporal slice is $I(t,y)$ (l becomes y under this condition). The flow vector passing the line for an individual target is $\boldsymbol{m}$ as shown in Fig. 2(b), where u is the component orthogonal to the line (in x direction) for revealing shapes, and v is the component parallel to the line that leaves skewed shape in the slice.

Where to locate a sampling line? The position is determined before its orientation at a path of target. It can be selected manually once in the video, or by accumulating the flow amount in the field of view over a long period of video. A *condensed image* of video [6, 7, 8, 9] indicates the global motion trajectories in the video. In Figure 3(b), we can

observe traces of background parallel to the time axis and the foreground flow moving in different directions. Such foreground flow will appear on paths in the field of view.

3. SHAPE ANALYSIS IN TEMPROAL SLICES

We sample the pixels on the selected line at each frame to obtain an array of pixels, and the arrays from consecutive frames are connected along time axis. This results in a *temporal slice* $I(t,y)$ in the spatial-temporal volume as shown in Fig. 3(a). The slice thus shows accurate temporal resolution to a frame, and is able to preserve certain characteristics of target shape and environment as shown in Fig. 4. From the slice, we can index to a frame t at the precision of 1/60 second, if the interlace format is used in the sampling. This is much more accurate than the indexed resolution of a clip by mosaicing [1] or tapestry [3]. Sometimes slicing at an obscured location can even reveal acute and deliberate details in the video, often unnoticed by the human. In Figure 5, the visual attention may not notice an object (marked in green) sneaking away, while it appears clearly in the temporal slice. Figure 6 shows another example where a horizontal line scans vertical motion to produce fine shapes of passing targets.

Fig. 5 A temporal slice captures a passing car in its complete shape (right), which is unnoticeable as a complete shape in the video frame (left). The red line is sampled for the temporal slice.

Subject to certain shape deformations, the temporal slice reveals all the passing targets in the video for retrieval. By identifying a point of interest in the temporal slice, the original frame of the point is thus located for further check. In addition, the deformation also reflects some valuable information about the foreground targets. The shape analysis for the temporal slice is given as follows.

(a) The length of a target in the temporal slice is related to its image velocity, i.e., its length is inversely proportional to u at the sampling line. If the target crosses the sampling line at a high velocity, its shape appears narrow in the slice. On the contrary, if its velocity is low, its shape appears stretched in the slice.

(b) If the principal pose $\boldsymbol{p}$ of a target has an angle β with respect to the sampling line in the frame, the projected shape is skewed along the t direction in the slice. This may happen when the camera overlooks a site from a high position under perspective projection. The image flow component v scales the target height in the temporal slice as well. The targets can be inversely skewed to improve the shape in the temporal profile.

(c) All targets face the same direction (left) in a single temporal slice; the slice lacks the spatial location and moving direction for targets. Targets are scanned at their front when they move forward. This can be improved by multi-line sampling in the following.

4. MULTI-LINE TEMPORAL PROFILE

4.1. Foreground Extraction in Temporal Slices

In order to overcome the problem of a temporal slice in lacking spatial information, we set multiple parallel slices at a critical location in the video to sample dynamic events. If these lines are spatially apart from each other with the distances in between roughly wider than target widths, a target will not pass multi-lines simultaneously. Thus the foreground flow crossing these lines will have delays, and the shapes appearing in the corresponding temporal slices will not overlap exactly. This temporal order in the slices helps to determine the motion direction of target. Therefore, we blend these temporal slices together according to their spatial locations to create a *temporal profile* of video that shows the dynamic flow of foreground clearly.

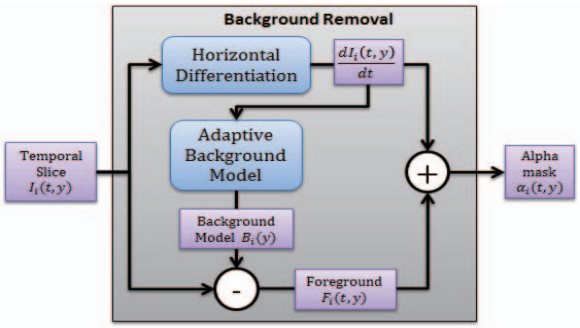


Fig. 7 Foreground detection from temporal differentiation and background subtraction in each temporal slice.

Fig. 6 A horizontal line samples apparent vertical motion to obtain a temporal slice. The targets are skewed due to slanted motion vector with respect to the sampling line (blue). The high image quality allows capturing details.

In each temporal slice, background is visible as parallel stripes along time axis as depicted in Figs. 3, 4, and 5, and it will occlude other slices in the blending. Analysis of a temporal slice alone can yield sufficient information for background removal [4], which deals with a smaller data set than analyzing the video volume [10, 11]. Therefore, we perform a series of steps to remove the background in each slice as depicted in the diagram in Fig. 7.

For temporal slice $I_i(t,y)$ sampled at position $i = 0,1,2$ we compute the temporal derivative of the slice as:

$$\mathrm{I}'_i(\mathrm{t}, y) = \frac{\partial I_i(\mathrm{t}, y)}{\partial t} \tag{1}$$

which is implemented by Sobel operator and the output is mapped to range [0, 1]. Since each background pixel leaves a uniform trace parallel to the time axis, it will be removed after this temporal differentiation with a threshold τ_1. The remaining pixels are dynamic foreground, which may also include some slow movements such as the waving of leaves, flag, and water. However, this derivative computation may miss some foreground pixels if the foreground has a homogeneous color distribution such as a surface on a bus.

We use the differential image to produce and update a *background map*, $B_i(y)$, which will be used for robust foreground extraction. For consecutive time instances t_j ($j<N$) without foreground activities, i.e., $|I_i'(t_i,y)|<\tau_1$, the background map is estimated as the average of those columns.

$$B_i(y) = \frac{1}{N}\sum_{j=0}^{N} I_i(t_j, y) \tag{2}$$

We then subtract the entire temporal slice from this background array to get foreground regions; the resulting value of subtraction is thresholded by another threshold value τ_2.

$$F_i(t,y) = \begin{cases} 0, & |B_i(y) - I_i(t,y)| < \tau_2 \\ 1, & otherwise \end{cases} \tag{3}$$

This output is further combined with the result from Equation (1) to generate a *mask* of dynamic foreground as the maximum of background subtraction and temporal differentiation for slice *i*.

$$mask_i(t,y) = \max\left\{F_i(t,y), \frac{dI_i(t,y)}{dt}\right\} \tag{4}$$

Fig. 8 Masks generated for foreground objects in a temporal slice.

It is worth mentioning that the differences are all performed in three color channels, and the result is converted to an 8-bit grayscale image. Figure 8 is an example of detected foreground regions from a temporally sampled slice. The resulting masks are used for blending multiple slices.

4.2. Integrating Slices to Temporal Profile

At a critical location, a temporal profile is integrated from three different temporal slices sampled on parallel planes in the video volume. We blend the slices with different transparencies according to their spatial locations in the video frame. This generates a distance based haze effect that can illustrate the spatial positions of targets from the margin of the field of view.

Denote $I_0(t,y)$, $I_1(t,y)$ and $I_2(t,y)$ as the slices on locations from right to left in the video frame respectively. Each slice has a blending coefficient α_i that determines its contribution to the final temporal profile. In general, assume the video is sampled at n different locations, where $n\geq 3$, e.g., as depicted in Fig. 9, we blend slices into a temporal profile as

$$P_i(t,y) = \left[1 - \alpha_i mask_i(t,y)\right] P_{i-1}(t,y) + \alpha_i mask_i(t,y) I_i(t,y)$$
$$P_0(t,y) = I_0(t,y), \qquad i = 1,\ldots,n \tag{5}$$

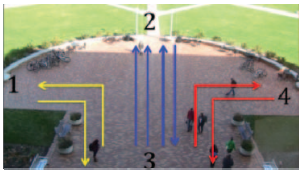
where P_0 is the slice at location 0. It contains a background to provide the profile with a context. If the value of $mask_i$ at a position is zero (i.e., background), the color value in P_{i-1} is used. As long as the slices are blended in such spatial order, the motion direction will be clear in the resulting profile. An example is shown in Fig. 9, where α_1 and α_2 for slices $I_1(t,y)$ and $I_2(t,y)$ are set as 0.75 and 0.5, respectively. Therefore, one can observe that, if a target moves to the right in the frame, the shapes will tend to be more and more opaque in the profile. Inversely, a leftward motion generates shapes more transparent in the profile. This design considers the viewing direction of the temporal profile from right side of the video volume, because the time axis is always preferred to be aligned towards the right. The transparency add in the temporal profile solves the moving direction problem in a single temporal slice. For the horizontal sampling lines in the frames, we assign the lowest slice the most opaque values and transparency increases for higher slices, because the lowest sampling line is always the closest one to the camera in most surveillance cameras overlooking the sites of interest.

5. EXPERIMENTS AND DISCUSSION

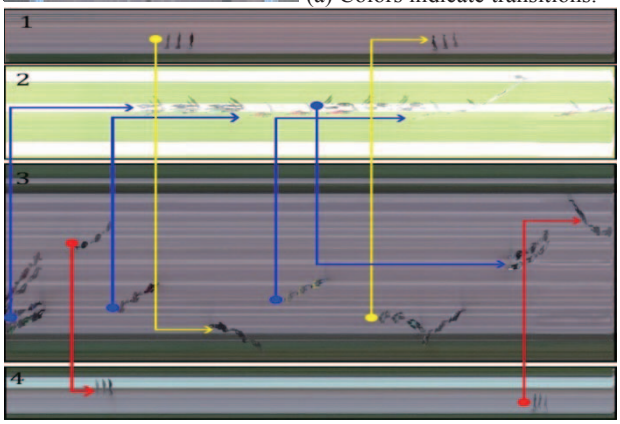
We have tested various experimental videos including indoor and outdoor scenes. For each long video, sampling locations are set manually at critical locations according to the layout of the site in the field of view and the motion directions there. The quality of the figures in profile is compatible to that in video frame in most cases. Figure 10 displays several video profiles at different locations. We use multi-lines to capture temporal slices with synchronized time stamps. Background and foreground separation is implemented stably in the slices. Blending of slices gives a

natural visual effect. By incorporating the shape and temporal information acquired in the profiles, we obtain global motion directions in the video. Finding the frames with passing targets in the video becomes extremely efficient by browsing the temporal profile. Furthermore, cutting off segments without targets from profiles is straightforward.

It is possible that a target has both vertical and horizontal motions at a sampling location ($v \neq 0$). If velocity component v parallel to the sampling line is large, the target shape will be stretched along the spatial axis in the temporal slice. If passing velocity component $u \neq 0$, we have the opportunity to skew the target in y direction to improve the foreground shape. Before blending slices into the temporal profile, we inversely skew the bounding boxes enclosing the targets in the temporal slices according to the direction of $\boldsymbol{m}$ obtained from the path. Figure 11 shows such inversely skewed results, where local shapes of targets are improved. The moving direction in the temporal profile (moving up or down in the frames) is preserved by their positions in the temporal profile. Hence, only the scaling in t direction remains.



(a) Colors indicate transitions.



(b)

Fig. 13 Temporal profiles from four different locations for finding transitions of people between them. (a) Critical locations and their spatial relations. (b) Profiles showing transitions from one location to another.

In surveillance videos, a large portion of motions in the view are random without consistent image velocity. They include fluttering flags, waving trees, human body motion in rotation and translation towards the camera. Such actions do not fit into our framework to present complete shapes in the time domain. However, our algorithm can still produce a temporal slice or profile that contains important motion information in the video. Although a slice shows incomplete patterns of targets, the happening time instance of the event is exactly punctured. Figure 12 displays such a profile from a scene with non-transitional movement, such as minor waving and shaking. We can at least understand the actions of a target over time in the temporal profile, despite of the small body motion to be considered as translational.

Compared to tracking a target path in video frames, our framework yields the motion transition between critical locations as shown in Fig. 13. Distributed surveillance cameras, even without view overlapping can take advantage of temporal profiles. Based on the current results, our future work will be the re-identification of people in the temporal profiles from non-overlapped distributed cameras. The trajectory of a person can sorted out in a large area with a number of cameras.

6. CONCLUSION

This paper proposed an original method to generate a temporal profile of surveillance video at critical locations without time length limitation. The temporal profile will not only provide an intuitive summary of moving targets in surveillance video, but also have the accurate time passing those locations. Moreover, it provides certain spatial shape with transparency and deformed shape of targets for further examination in the video frames. The temporal profile can greatly facilitate the fast search of targets and guide examination to the frames in large database retrieval of surveillance video.

7. REFERENCES

[1] Y. Pritch, A. Rav-Acha, S. Peleg, "Nonchronological video synopsis and indexing," IEEE Trans. PAMI, 30(11), 1971-1984, 2008.

[2] J. Varadarajan, R. Emonet, J.-M. Odobez, "A Sequential Topic Model for Mining Recurrent Activities from Long Term Video Logs," Int. Journal Comp. Vision, 103,100-126, 2013.

[3] C. Barnes, D. B. Goldman, E. Shechtman and A. Finkelstein, "Video tapestries with continuous temporal zoom," ACM Trans. Graphics (TOG), 29 (4), 89, 2010.

Fig 9. Temporal profile of a video recording an indoor environment with three sampling lines (red). The transparency of slices increases from right to left. The arrows on top of targets show their passing in leftward or rightward direction.

[4] J. Y. Zheng, S. Sinha, "Line cameras for monitoring and surveillance sensor networks", ACM Multimedia 07, pp. 433-442.
[5] J. Y. Zheng, H. Cai, K. Prabhakar, "Profiling video to visual track for preview," IEEE Int. Conf. Multimedia and Exposition 2011, pp. 1-6, 2011.
[6] H. Cai, J. Y. Zheng, "Video anatomy: cutting video volume for profile," 19th ACM Multimedia, 2011, 1-4.
[7] H. Cai, J. Y. Zheng, "Automatic heterogeneous video summarization in temporal profile", Intl. Conf. on Pattern Recognition 2013, pp. 2796-2800.
[8] G. R. Flora, J. Y. Zheng, "Adjusting route panoramas with condensed image slices," ACM Multimedia, 815-818, 2007.
[9] J. Y. Zheng, Y. Bhupalam, H. T. Tanaka, "Understanding vehicle motion via spatial integration of intensities," Int. Conf. Pattern Recognition, 1-5, 2008.
[10] C. Stauffer, W. E. L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. PAMI vol.22, no.8, pp.747-757, 2000.
[11] J. Yao and J.-M. Odobez, "Multi-layer background subtraction based on color and texture," IEEE Conf. Comp. Vision Pattern Recognition, 1-8, 2007.
[12] D. Zhong and S.-F. Chang, "Spatio-temporal video search using the object based video representation," IEEE Int. Conf. Image Processing, vol.1, 21-24, 1997.

(a) Profile at paths on two sides of a river. Aspect rates are different for people walking at almost the same speed on different depths.

(b) Sampling at an outdoor area with multiple paths. Leave waving is also observed in green part.

(c) Profile from a surveillance video at a building entrance with multiple targets passing in different directions at the same time.

(d) Sampling at an intersection where pedestrians and vehicles pass.

Fig. 10 Experimental results of temporal profiles from surveillance videos. Time axes are horizontal. Key frames are also displayed.

Fig. 11 Inverse-skew operation for shape improvement in the temporal profile. (left) An indoor area where a passing flow (yellow) is very slanted in the view. (middle) Original temporal profile. (right) Inversely skewed profile in dynamic foreground regions.

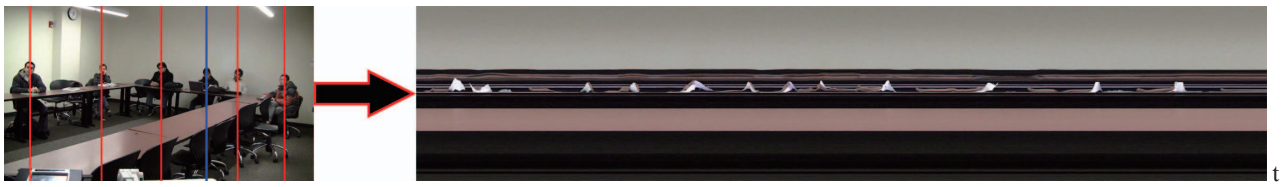


Fig. 12 Temporal slice for non-translational body movement. One can observe the events (flipping book page) in the profile.