PROFILING VIDEO TO VISUAL TRACK FOR PREVIEW

Jiang Yu Zheng, Hongyuan Cai, Karthik Prabhakar
Dept. of Computer Science, Indiana University Purdue University Indianapolis (IUPUI)
jzheng@cs.iupui.edu hocai@cs.iupui.edu kprabhakar@gatech.edu

ABSTRACT

Video indexing is important to video browsing, editing, retrieval, and summarization. This work takes a new aspect view to profile a video volume to a visual track and create a digest of video for preview. Our projected profile of video contains spatial and temporal information inclusively in a 2D image scroll that is continuous, scalable, and indexing to frames. We analyze the camera kinematics including zoom, translation and rotation, and categorize camera works for profiling various types of video. The key idea is to use a sampling line to sweep the video volume across the major optical flow so as to obtain an intrinsic scene space that is less influenced by the camera motion. We also use motion blur technique to render dynamic targets in the profile. The resulting video track can provide a video preview for guiding the access to the frames. It will facilitate video surveillance, visual archiving of environment, video retrieval, and video editing.

***Index Terms*—** video index, spatial-temporal slice, motion, camera kinematics, camera work, profile of video

1. INTRODUCTION

Browsing a video is not easy as browsing a collection of photos, because of its sequential format containing a large amount of data, and thus requires fast-forward operation. The video indexing so far uses key frames from video clips or shots [1]; various algorithms have been proposed to extract representative frames from relatively constant scenes. However, the key frames are discrete and only contain spatial information at certain moments. It can only index to a video clip rather than to each frame.

This work creates a profile from recorded video data. It is a 2D image belt that contains one axis as the timeline, and the other indicating one space dimension in the video frame. It provides a novel aspect view of a video from side of the video volume, and the time belt can be presented in a video track used in video editing and production. Figure 1 shows a video track in a video software, which is scalable, scrollable, and indexing to frames, but without visual contents displayed currently.

We analyze the intrinsic camera works in capturing videos and categorize videos in several types. We separate optical flow in a video clip into the camera caused motion with respect to a static background, and the dynamic object motion against the background. For each type of video clip, we design a cutting in the video volume to yield a planar or curved slice for revealing the video content in the temporal visual track. This is implemented by sweeping a sampling line across the video volume. For dynamic objects in the video, we further generate motion blur against the clear background in the profile to show their motion trends.

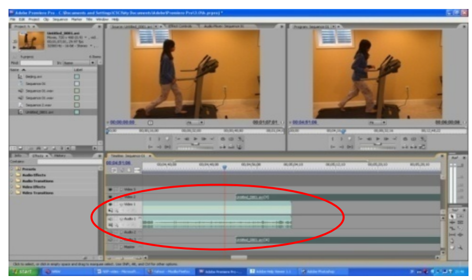

Fig. 1 Current video indexing and editing using video tracks. A frame indicator (vertical red bar) shifts along timeline for picking up a video frame to display.

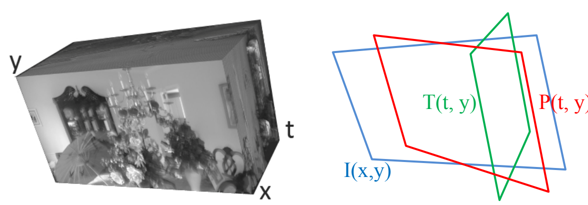

Fig. 2 Spatial-temporal volume of a video clip, a spatial key frame I(x,y), temporal slice T(t,y) and a spatial-temporal slice in red used for our profile of video.

The significance to generate such a new video track lies in its continuous form detailed to frames, which is impossible by key frames; it reveals more temporal information. At the same time, the spatial information in the video is also visualized in the visual track as complete as possible, although some deformation and changes in ordering are brought in. It is easy to be embedded into video software to enhance video editing, retrieval, analysis, and visualization in general.

The related works include spatial indexing approaches by key frames (see Fig. 2) and its extension to a larger spatial domain by mosaicing camera panning views [2] and dynamic panning views [3][4][5][6][15], and overlapping time-different events in the same spatial domain [7]. The drawbacks of such spatial indexing are (1) lack of temporal order, (2) becoming clutter if the video clip lasts for long or targets are crowded in the video, and (3) camera motion limited on pan/tilt only. The matching of background and segmentation of foreground figures are also not robust for complex scenes, never mentioning the inconsistent background due to motion parallax coming from a translating camera as in [10]. In contrast, the video index created in temporal domain (see Fig. 2) has good results on rotating and translating camera motion realized by fixed slit [10][17] and dynamic slit scanning [12]. The shortcomings are a low temporal resolution in showing the video contents for a short clip, and a different projection than the normal perspective projection. Other video analysis and visualization methods include spatial-temporal slicing, EPI [11] and condensed image [8][13]. Our profile of video is a spatial-temporal slice that can overcome the problems in resolution, camera motion types, and robustness in above indexing methods. It is designed to work on all camera kinematics and does not require matching and segmentation. Moreover, it keeps the temporal order of video scalable in time to a resolution better than key frame.

This paper is organized as follows. Section 2 starts from categorizing camera works that generate different flows in video according to the camera kinematics. Section 3 gives a general framework of this visual profiling for both static background and dynamic foreground. Section 4 introduces the visualization of dynamic objects in the visual profile by motion blur technique. Section 5 is the experiments followed by conclusion.

2. CAMERA KINEMETICS AND FLOW

As a camera moves, video frames have partial overlaps on field of view (FOV). We can thus profile a video clip to a 2D image by removing redundant data. The profile should include all the stably visible scenes in video without redundancy, keep a proper resolution of scenes, and render dynamic effect of moving foreground. We start from the analysis of camera motion styles first.

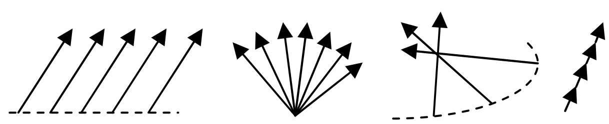

Fig. 3 Typical camera works with zoom, translation and rotation. The arrows indicate the movement of camera axis, and their lengths correspond to the camera focal lengths. (a) Translation, (b) pan/tilt, (c) around-object rotation, (d) zoom in/out.

A camera can be static or undergo motions such as zoom, rotation (pan/tilt/roll), translation (forward/backward/ sideways), or their combinations. Practically, a video can be categorized by its camera work such as pan/tilt rotation, rail/vehicle based movement, focusing or moving around object (composed of translation and rotation), forward moving or zooming, and so on as illustrated in Fig. 3. In addition, dynamic scenes in the FOV may further reveal a variety of motions.

In general, we can describe the camera kinematics in Fig. 4, which yields typical camera works (dotted boxes). The camera works generate distinct optical flows that can be classified as diversified flow or directional flow in the FOV. For the video clips with directional flow generated from the camera ego-motion or the directional movement of a target crowd, we can specify a *major flow* in the FOV. Then, we consider the flow component orthogonal to the major flow as *minor flow*, and other flow components such as instantaneous rotation (from camera roll) and shaking as *unstable flow*. To obtain a profile $P(t,y)$ from video, a sampling line $\boldsymbol{l}$ is set to scan the major flow from one end of the video frame to the other. The scanning direction is opposite to the flow direction and thus is the camera moving direction in many cases as depicted in Fig. 5. All the pixel columns in $P(t,y)$ are from the shifting line $\boldsymbol{l}$. If the video clip is long or the length is unpredictable as from a surveillance camera, we also fix the line in the frame to scan the scenes.

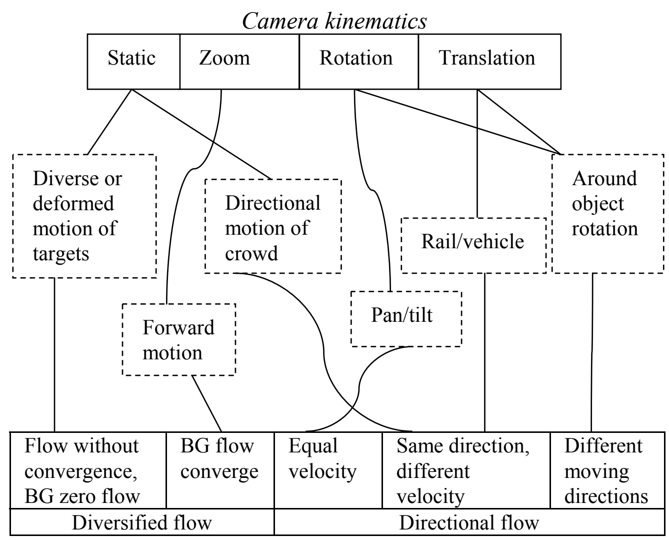

Fig.4 Categorizing camera motion, camera works and flow styles

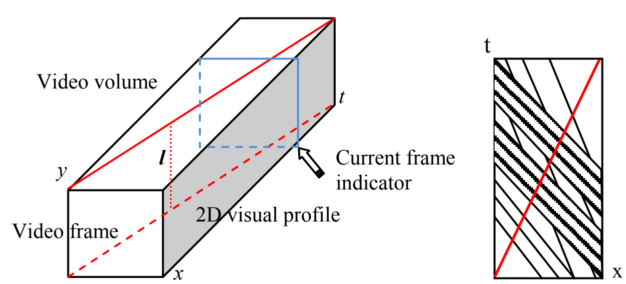

Fig. 5 Video volume and a possible cutting of diagonal slice across a horizontal major flow in the video clip. (a) A video volume, (b) An EPI showing directional flow traces of scenes in the video clip.

The profile is representative because all the scene points stably appearing in the FOV are included. It is also compact because every scene point is scanned only once by the sampling line. We cut a 2D profile without redundant pixels out of a video volume, because the scene transition causes overlaps in the consecutive frames.

The minor flow orthogonal to the major flow is from hand shaking, jaggy walking, and vehicle waving during video capturing [8]. Its effect is visible in our visual profile as tilt changes during panning, translation, and zoom. Minor flow can be kept in the profile to reflect the dynamics of the camera, or can be removed by video deshaking algorithm before and after profiling [3]. The *unstable motion* is from an instantaneous camera rolling, as well as the articulate motion and deformation of targets. They mainly reduce the quality of the visual profile and will not be used for determining the slice cutting. Although the unstable flow can be deshaked first, we avoid cosmetic rectification since a visual profile should also record the true motion.

3. GENERATION OF PROFILE OF VIDEO

We assume a smooth camera motion in order to focus on the video profile generation for all types of camera kinematics.

3.1. Profiles from static camera

For a static camera shooting mild motion such as a talk show, we cut a vertical slice (consistent to gravity) across the video volume diagonally. A diagonal control line either straight or curved through the EPI is longer than the frame width so that the profile of video, $P(t,y)$, has a good resolution even if it is scaled up along the timeline. Figure 6 gives an example compatible to a key frame. The cutting line is set as a curve $\boldsymbol{x}_s(t)$ to distribute targets properly in the profile.

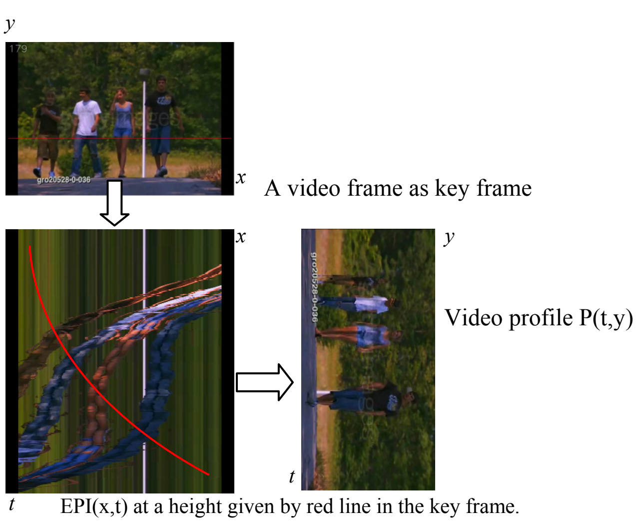

Fig. 6: Four teenagers walking in a static background. (a) A key frame. (b) An EPI where leg movements are visible as traces. The slice cutting location is marked in red curve, (c) A profile that can be scaled along the time axis in a video GUI.

If a camera shoots directional flow, e.g., a surveillance camera monitors people and vehicles through pathways, a sampling line parallel to the dominant structure lines in the scenes is set to cut the flow for the profile as in [14]. The profile shows the shapes and time of arrival of passing targets. If a camera is shooting diversified flow, we can set multiple sampling lines at pathways where major flows occur. With this profile, a surveillance video lasting for many hours can be briefly browsed to locate a special time for examination or to count the total number of passages.

3.2. Zoom in/out

Extending from a static camera, camera zoom yield a flow expending from a Focus of Expansion (FOE). Considering the gravity direction projected in the frame, we can specify the major flow direction as horizontal and set a vertical line to sweep the video volume from either left or right. The captured people and objects thus will have less distortion in the profile. The cutting curve $\boldsymbol{x}_s(t)$ is bended in such a way to preserve the resolution of the enlarged portion in the zoomed FOV. From the shape enlarged along time axis in the profile of Fig. 7, we can figure out the scenes being zoomed up in the video clip. If the scene is zoomed out, we can obtain a time-flipped curve for the profile of video; the shape is preserved but the order is reversed.

3.3. Pan/tilt clip

Panning appears most frequently in video for increasing the FOV and tracking a target of interest. To identify the motion direction and align a cutting slice, we use a *condensed image* [13] to compute the major flow from homogeneous traces as in Fig. 8, which is robust to vertical camera shaking and small rolling. Although the generated panorama is bumpy, it reflects the minor flow caused by tilting and can be rectified through deshaking. Similarly, tilting clips can be processed in a symmetric way for a profile $P(t,x)$.

3.4. Sideway translating camera

The camera translation is always visible in movie shots captured by vehicle/rail sets [9][10]. Such shots can be also captured from air planes, ships, cars, etc. The camera translation in a sideway direction creates a parallel flow field in the FOV with non-homogeneous motion parallax due to varied scene depths. A vertical line can scan the horizontal major flow diagonally in the video volume for a profile. If the camera is translating in a direction facing forward, it creates a flow field expending from FOE; it has both effects of sideway translation and zoom in/out as analyzed in [9]. A vertical line cutting the video frame generates a forward aspect view for all the visible scenes, which is slightly different from what a perspective projection can obtain; it is named *parallel-perspective projection* [9][10].

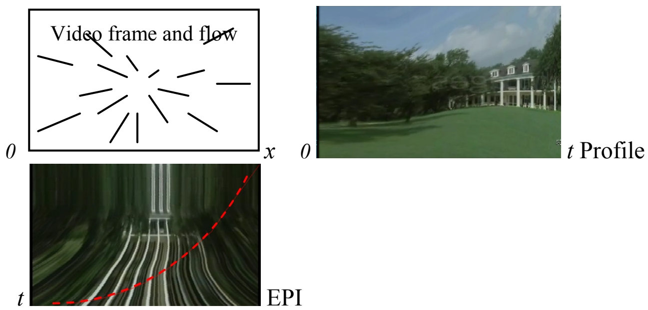

Fig. 7: Zoom in video clip and a generated profile cut at a red curve in the EPI according to the expending flow.

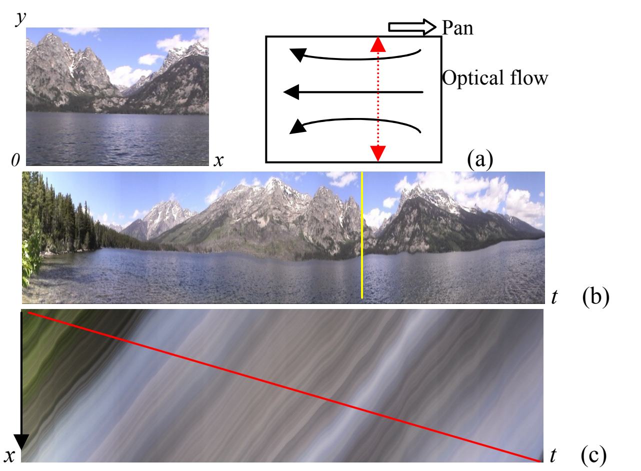

Fig. 8 Generating a visual profile from a panning camera. (a) A video frame panning towards right with major flow in horizontal direction. (b) A profile in which the yellow line indicates the position of frame in (a). (c) Condensed image where a diagonal slice is cut for the profile in (b).

3.5. Around object motion

In many movies and video clips, a camera focuses on an object while its position is moved around the target as in Fig. 9a. Such an around-object motion (typically on a circular path) shows various aspects of a target. The camera has a simultaneous translation and rotation around the path center (green cross point). The path center divides the space into two parts: the space closer than it from the camera has the optical flow opposite to the camera translation, while the space farther than it has the optical flow in the inverse direction (same as the camera translation as shown in the EPI in Fig. 7b). The video captures the foreground (target) and background flow in FOV. The background is usually distant than the path center since the center is usually planned near the target.

We locate a *fixation point* of lines of sight at a position between background and foreground scenes as in Fig. 9a. This corresponds to cutting the profile along a red line that intersects the background motion diagonally in the video volume (Fig. 9b). The slice also intersects the foreground to capture the foreground object of interest, though the width of it will be extended (it is emphasized in a good sense), and the order of foreground may be reversed. Figure 10 shows an example in which the focused object (flower) is elongated in the profile while the background scenes are squeezed along the time axis. Our designed slice in the video volume cuts the flow (traces) of background. This selection is more reasonable than cutting the opposite diagonal line aligning with the background trace in the EPI, because this opposite cutting will extend a partial background and squeeze the foreground in the time domain.

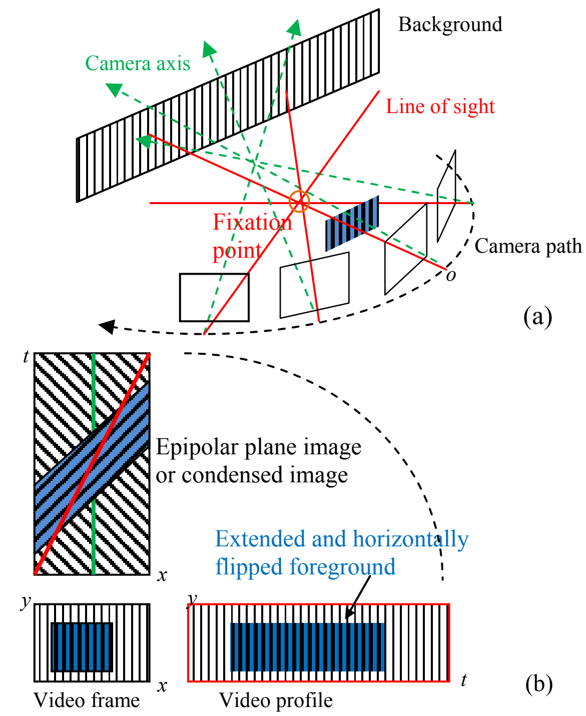

Fig. 9 Around-object motion of camera with translation dynamically changed by rotation, and its profile generation.

Fig. 10 Around-object motion in CCW direction. (top-left) end frame of the clip. The yellow arc shows the flow direction around the path center. (lower-left) An *EPI*(*x*,*t*) at the height indicated in the frame. (top-right) profile of video affected by minor flow.

We deal with general around-object motion along an arbitrary smooth path and camera rotation. Equivalently, a camera moving on a straight rail with a rotation towards a focused object can be considered as the same type of motion (maximally for half circle). Further, a rotating target in front of a static camera can be treated as this type. All the other objects draw sinusoidal trajectories in the EPI [16].

4. VISUALIZE DYNAMIC FOREGROUND SCENES

In this work, we are aiming at present both shape and motion information in the profile of video. To avoid severe shape distortion, we employ motion blurring technique on dynamic foregrounds. If we extend the exposure time of an image, dynamic objects are motion-blurred because of the temporal accumulation of intensities at each point. Static objects have consistent intensities over time and their average are still sharp. People can perceive the motion information when the motion blurred is presented.

As a sampling line scans the video volume as Fig. 11 shows, even a mild motion of the person is profiled with a distortion. We create a degree of motion blur by averaging intensity temporally around the cutting slice. As shown in Figure 11a, we prepare a pixel wall along the slice with thickness of γ for averaging. The output of the slice is then obtained from

$$P(t,y) = \frac{1}{\gamma}\sum_{b=1}^{\gamma} I(x_s(t), y, t+b) \quad (1)$$

where $I(x,y,t)$ is the intensity in the video volume. For Fig.11b, the profile is improved by motion blur in Fig. 11c where γ=35 frames.

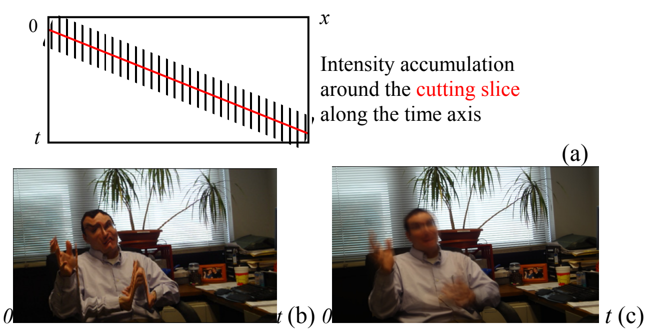

Fig. 11 Motion blur in the profile of video for visualizing dynamic foreground. (a) Accumulating intensities along time axis during slice cutting. (b) A simple slice cutting without motion blurring. (c) The profile with motion blurred person and sharp background.

More generally, we deal with the scenes from a moving camera. For a camera panning, assume its horizontal image velocity is $R(t)$, which is constant in the entire frame at time t. We estimate the orientation of the flow traces in the condensed image, $C(x,t)$, from their gradient. For each time instance t, the positions $x_i(t)$, $i=1,2,..N$ with the gradient values larger than a threshold are picked for estimating velocity $R(t)$, i.e.,

$$R(t) = \frac{1}{N}\sum_{i=1}^{N} \theta(x_i, t) \quad (2)$$

This yields a stable result of trace orientation. We use it for temporal accumulation in the flow direction to obtain

$$P(t,y) = \frac{1}{\gamma}\sum_{b=0}^{\gamma} I(x_s(t) + b\tan(R(t)), y, t+b) \quad (3)$$

This accumulation has two effects. It motion-blurs the dynamic objects different from background (major flow) as in Fig. 12. It also enhances the background in the profile that may be motion blurred in each individual video frame.

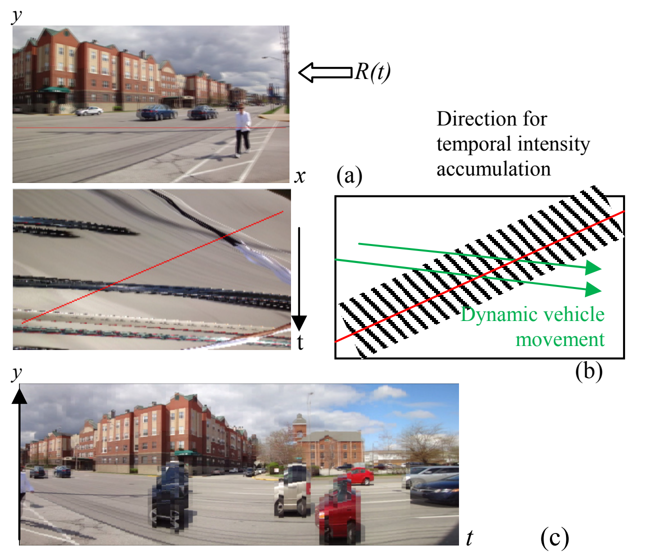

Fig. 12 Motion blurring for dynamic foreground and rotating background. (a) A video frame during camera panning left. (b) An EPI and averaging intensities along the background flow direction. (c) Scaled profile of video in time with sharp background and blurred cars from the diagonal cut in (b).

Further, if a camera involves translation, the motion parallax is then not consistent; it depends on the object depth from the camera. We can only compute a dominant parallax at each time instance for intensity accumulation along that direction. This means only the objects in the dominant depth are clear and objects off the depth will have a certain degree of motion blur. The dominant motion direction is computed by a median filter on flow directions, i.e., $R(t)=median(\theta(x_i,t))$ from where the gradient is larger than a threshold. We then use a small γ for intensity accumulation.

5. EXPERIMENTS AND DISCUSSION

A software has been developed for direction selection for condensing video, cutting EPI, and visualizing the flow field. The condensed image is displayed for verification and iterative adjustment of the slice locations. The camera/flow style is discriminated by human. The program allows interactive allocation of cutting slices for the profiles of video. The implementation of motion blurring is also carried out based on the major flow direction calculation. We use video clips from several websites and our own archives as well for experiments. For consecutive clips in Fig. 13, the spatial-temporal profile reveals the shape information as complete as possible in the temporal domain after selecting a series of slices in the video volume properly. It provides a preview of video even if the shape has some distortion and disordering as comparing to the spatial order in the video frame.

For general camera works containing more actions such as panning during translation, zoom during panning, etc, we will identify the major flow direction and style in order to align cutting slices. Because a general motion can be decomposed to translation and rotation, we will combine the rules of slice cutting in each individual motion for an arbitrary camera motion. Future works will further include automatic detection of flow styles for profiling.

6. CONCLUSION

This work discussed a new version of video indexing by profiling video to a visual track. It generates a continuous visual profile along the timeline in which both spatial and temporal information is visualized. After the clip detection, we analyzed the camera work from the corresponding flow. Then we cut the slices across the major flow for background profile, on which dynamic objects and events are rendered with motion blur. The profile can be embedded into video software as a visual track to assist video browsing, editing, indexing, and retrieval. Although our profiling cannot preserve complete shape and time information in 2D tracks from 3D video volumes, we have presented rich information in a perceptible style in the profile based on the camera kinematics and video characteristics. The video track is further compact and scene orientated (less frame dependent) for video analysis and recognition.

7. REFERENCES

[1] B. Janvier, E. Bruno, T. Pun, S.M. Maillet, "Information theoretic temporal segmentation of video and applications: multiscale keyframes selection and shot boundaries detection", *Multimedia tools and application*, 3(3), 273-288, 2006.

[2] Y. Taniguchi, A. Akutsu, Y. Tonomura, "Panorama Excerpts: extracting and packing panoramas for video browsing", *ACM Multimedia* 97, 427-436.

[3] M. Gleicher, F. Liu, "Re-cinematography: improving the camera dynamics of casual video", *ACM Multimedia 07*, 27-36, 2007.

[4] Liu, Y. Hu, M. Gleicher, "Discovering panoramas in web videos", *ACM Multimedia 08*, 329-338.

[5] A. Rav-Acha, Y. Pritch, D. Lischinski, and S. Peleg, "Dynamosaicing: Mosaicing of Dynamic Scenes", *IEEE Trans. PAMI*, 1789-1801, 2007.

[6] A. Bartoli, N. Dalal, R. and Horaud, "Motion panorama", *Journal of Computer Animation and Virtual Worlds*, 15(5), 501-517, 2004.

[7] Y. Pritch, A. Rav-Acha, S. Peleg, "Nonchronological Video Synopsis and Indexing", *IEEE Trans. PAMI*, 30(11), 1971-1984, 2008.

[8] J. Y. Zheng, G. Flora, "Adjusting route panoramas with condensed image slices", *ACM Conf. Multimedia 07*, 815-818, 2007.

[9] J.Y. Zheng, Y. Zhou, P. Mili, "Scanning scene tunnel for city traversing", *IEEE Trans. Visual. Comput. Graph*. 12 (2) (2006) 155–167.

[10] J. Y. Zheng, "Digital route panorama", *IEEE Multimedia,* 10(3), 57-68*,* 2003.

[11] H. Baker, R. Bolles, "Generalizing epipolar-plane image analysis on the spatial-temporal surface", *IEEE CVPR88*, 2-9 (1988).

[12] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall, "Mosaicing new views: the crossed-slits projection", *IEEE Trans. PAMI*, 2003, 741-754.

[13] J. Y. Zheng, Y. Bhupalam, H. Tanaka, "Understanding Vehicle Motion via Spatial Integration", 1-5, *19th Inter. Conf. Pattern Recognition*, Florida, 2008.

[14] J. Y. Zheng, S. Sinha, "Line cameras for monitoring and surveillance sensor network", *ACM Multimedia 07*, 433-442.

[15] D. N. Wood, A. Finkelstein, J. F. Hughes, C. E. Thayer, D. H. Salesin, "Multiperspective panoramas for Cel animation", *SIGGRAPH 97*, 243-250, 1997

[16] J. Y. Zheng, "Acquiring 3D model from a sequence of contour", *IEEE Trans. PAMI*, 16(2), 163-178, 1994.

[17] J. Y. Zheng, S. Tsuji, "Generating dynamic projection images for scene representation and understanding", *Computer Vision and Image Understanding*, 72(3), 237-256, 1998.

Fig. 13 A girl twirling to music focused by a camera panning left and right repeatedly. (top) Continuous profile of video. (left) a video frame, (lower) an EPI shows the camera panning trajectory and a twisted trace of the girl. The red line indicates the location for slices cut.