Detection and Removal of Lighting & Shaking Artifacts in Home Videos

Wei-Qi Yan
Department of Computer Science, School of Computing, National University of Singapore
Singapore 117541, Tel: (65) 6874-6806
yanwq@comp.nus.edu.sg

Mohan S Kankanhalli
Department of Computer Science, School of Computing, National University of Singapore
Singapore 117541, Tel: (65) 6874-6738
mohan@comp.nus.edu.sg

ABSTRACT
Many amateur videographers, like home video enthusiasts, may capture videos that are not of a professional quality. Many minor but visually annoying distortions like lighting imbalance and shaking artifacts could be introduced by the unskilled operations of the video camcorder. Since home videos constitute footage of great sentimental value, such videos cannot be summarily discarded. Unlike movies and sitcoms, shot re-takes of important events, such as wedding ceremonies are just not possible. Therefore, such distortions need to be corrected. In this paper, we present a novel method to detect segments of videos that have lighting and shaking artifacts. These segments can then be subjected to a restoration process that can remove these artifacts. We present techniques to correct lighting artifacts by appropriately adjusting the luminance. In order to remove the shaking artifact, image mosaicing is first employed to build a mosaic frame for the segment with the aid of edge blending techniques. Subsequently a Bezier-curve based blending of motion trajectory is employed to perform motion-compensated filtering of the shaking artifact. The restored video is then created by appropriately cropping the mosaic frame based on the compensated motion trajectory. We have implemented the developed techniques and the experimental results on home videos demonstrate the effectiveness of our approach. Detection and removal of artifacts are significant in other videos as well as those obtained from autonomous vehicles, robots and remote sensing.

Categories and Subject Descriptors
H.30 [**Information Storage and Retrieval**]: General; H.5.1 [**Information Interface and Presentation**]: Multimedia Information System. 1.2.4.

General Terms
Algorithms, Design, Experimentation, Human Factors, Theory.

Keywords
Artifacts removal, lighting artifacts, video shaking, video mosaic.

Multimedia'02, December 1-6, 2002, Juan-les-Pins, France.

1. INTRODUCTION
Personal home videos are commonly utilized to record significant events involving one's family and friends such as holidays, weddings, and graduation ceremonies. In general, it is used to record people, places, and activities [6, 7, 15]. The amount of home video material can be very large - the total duration may easily accumulate to several dozen hours within a short period. With the steadily dropping prices of digital camcorders, home video usage is bound to proliferate. Home videos usually consist of raw unedited footage and generally have the following characteristics:

- Home videos are usually not shot by professionals. Any person who can press camcorder buttons is a potential home videographer. Thus many home videos tend to be made by amateurs.
- There is no deliberate story, plot or structure in the home video material. Since home videos do not have designated scenes or sets, they tend to be shot anytime at any location by persons of drastically varying skills. Moreover, no special lighting device is usually employed. Thus, very bright shots and very dark shots invariably creep up in many home videos. Moreover some camcorders armed with infrared devices can allow shooting of videos with only diffused green output with a bright center and dark concentric rings.
- Home videos often have shaking artifacts. This is because they are shot with portable hand-held camcorders with the photographer often shooting in motion like during walking or from a moving vehicle. Shaking is sometimes also related to zooming, if the shot object is at a long distance, the effect of shaking is exaggerated.
- Home videos often contain a lot of zoom (in or out) operations since the operation is relatively easy to do in most digital camcorders. However, rapid zooming operations often introduce blurring artifacts due to the lag in the imaging system response. Moreover, looking out from glass windows covered by rain/moisture generates blurred home videos.
- Home videos often appear foggy. When the camcorder is taken from cold place (air-conditioned room) to hot place (outdoors), the lens will be covered by moisture leading to a foggy video.

We are currently developing a digital home video album, in which we propose to provide full content management and distribution support for the average non-expert home video user. In this context, we are aiming to provide a utility for correcting artifacts inadvertently introduced by the user. Our long-term aim of this research is to provide a fully automated system that can automatically detect and correct the annoying visual artifacts that tremendously reduce the viewing pleasure of valuable footage of great sentimental value. The work described in the paper details our efforts towards this direction.

The significant of detection and removal of video artifacts is crucial not only in home videos but also in others, such as those of robot control and military applications. Even though videos shot at night in the field employ infrared techniques, detection and removal of lighting artifacts would be useful. Detection and removal of shaking artifacts play a key role in robot control and are helpful to stabilize the jerky images captured by a camera on moving robot. Detection and removal of blurring artifacts are useful in correcting satellite images. Thus, effective techniques to detect and remove artifacts will be very useful.

This paper is organized as follows. Section 2 will present related work, section 3 will depict lighting artifacts handling, section 4 will narrate our work about shaking artifact handling, section 5 will report the experimental results to show the utility of the techniques and section 6 will draw the conclusions and point out our future work.

2. RELATED WORK

Hitchcock [7] is a home video editing system, which presents a user interface that supports semi-automatic video editing. This system describes the problems that non-professionals have in using existing video editing tools and provides users with an interactive system for composing video that does not require manual selection of the start and end points for video clips. For the Hitchcock system, an algorithm to cluster clips into meaningful piles with an intuitive user interface to combine the desired clips into a final video is introduced in [6]. Reference [15] describes algorithms to abstract home videos with the algorithms adopting a new approach to cluster time-stamped shots hierarchically into meaningful units. However, such algorithms do not distinguish between low quality and high quality videos. Reference [16] details a system for indexing and browsing home videos. The system is capable of extracting both the structure information and the semantic objects. Reference [14] presents an algorithm for text segmentation and text recognition to the specific characteristics of time and date information in home videos. Reference [14] also proposes a new algorithm for clustering time-stamped shots into semantically meaningful units, for shortening shots into interesting clips, for selecting the clips of the video abstract, and for arranging them into the final abstract.

As far as practical techniques for image mosaicing are concerned, reference [10] describes a practical panoramic imaging system called FlyAbout that uses spatially indexed panoramic video for virtual reality applications. Reference [22] presents techniques for constructing full view panoramic mosaics from sequences of images. Global alignment is applied to reduce accumulated registration errors and a local alignment technique is adopted to compensate for small amounts of motion parallax introduced by translations of the camera and other unmodeled distortions. About video editing, reference [2] describes several visualization and interaction techniques that use video metadata, including transcripts, to investigate the problems of editing in this domain.

About removal of shaking artifact in a video, some papers [3,5,11] consider electronic digital image stabilization. Reference [3] describes a video stabilization algorithm using a block-based parametric motion model. In particular, it shows how to apply the algorithm to translate and rotate camera motions. In [5] a robust stabilization algorithm is given based on a simple use of block motion vectors. However, this method relies on the accuracy of block motion vectors. In [11], the authors provided a technique of image stabilization based on a 2D feature based multi-resolution motion estimation algorithm. By estimating the motion of camera, the combination of the estimates from a reference frame is used to warp the current frame in order to achieve stabilization. This method advantageously uses the information present in all images. However, as pointed out by the author in the conclusions, this method has many limitations in terms of the assumptions of camera motion. Reference [11] proposes an inertial model for motion filtering in order to eliminate the vibration of the video sequences and to achieve good perceptual properties. However, they have a limitation in terms of the maximum amount of displacement allowed [18,19]. Our work is very different from [3,5,11] in terms of the technique used as well as not having their limitations.

We now present an overview of our work. Unlike other work in references [3,5,11], our work is focused specifically on removal of artifacts in home videos. We have developed a new two-stage framework for the detection and correction of lighting as well as shaking artifacts. The input to our system is a raw video clip. Because a home video is often very long, it is first segmented into its shots [25]. From these shots, segments having lighting and shaking artifacts are then detected in the first stage. The user is then prompted whether (s)he would like to correct the artifacts. If the user chooses to correct the artifacts, the corresponding artifact-removal operations take place in the second stage. Finally, the processed video segments are saved. Figure 1 depicts the bird's eye-view of our work.

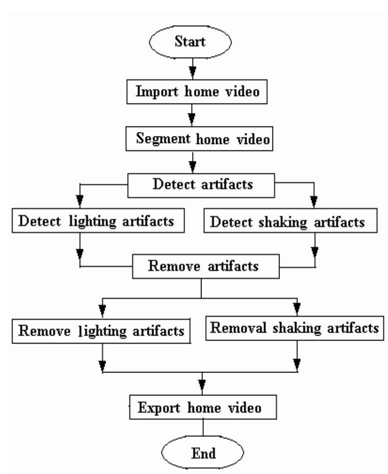


Figure 1. Overview of the system

The novelty of the contributions of this paper is in the detection as well as the removal of lighting and shaking artifacts of home videos. A video shot with lighting artifact is detected and classified into three types – over-bright, over-dark and imbalanced. We present a correction technique for each of the three types. We have also developed a novel technique for detecting and correcting shaking artifacts. We make the key observation that the shaking artifact can be construed as the noise of the motion trajectory. A Bezier-curve based approach is utilized to filter the motion trajectory. An image mosaicing technique has been adopted to construct a large intermediate frame that is employed for restoring the video based on the filtered trajectory.

Although many cameras have devices such as stabilizers, NightShot & Backlight calibrators and red-eyes calibrators to attenuate these artifacts [8,9,20,21,24], they do not always work well. For instance, the camcorders [26,27] for our experiments still shot the videos with these artifacts.

3. LIGHTING ARTIFACT HANDLING

As discussed earlier, home videos often possess segments having visual artifacts. After performing the pre-processing step of segmentation using shot-boundary detection [23], we first need to delineate the exact segments of the video shots that have the artifacts. Only then can further processing for their removal be done. Our strategies to detect and remove lighting artifacts of home video are summarized in Figure 2.

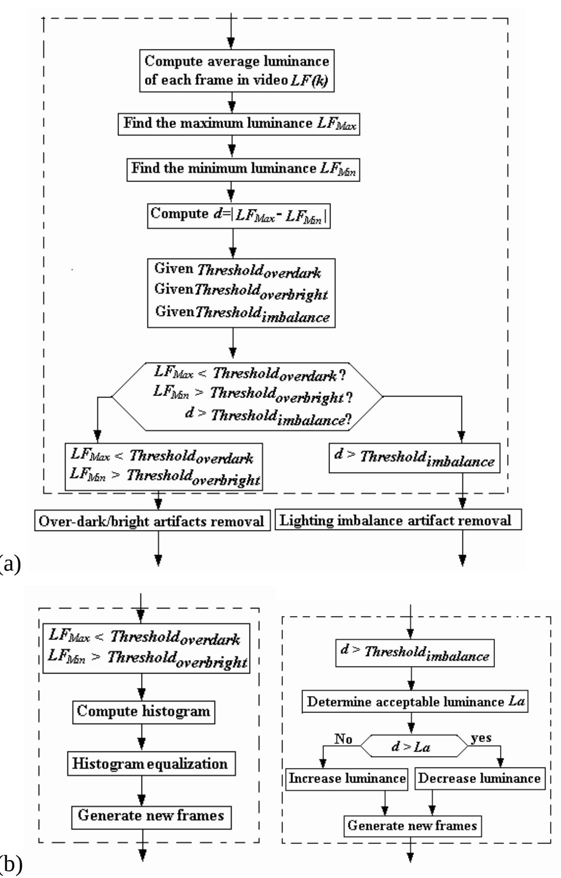


Figure 2. Detection and removal of lighting artifacts (a) Flowchart to detect frames with lighting artifacts (b) Flowchart to remove over-bright and over-dark lighting artifacts (c) Flowchart to remove lighting imbalance artifacts

We consider three kinds of lighting artifacts of home video: unbalanced lighting, over-bright and over-dark. They are described below: a video with lighting imbalance is the segment whose difference between the maximum value and the minimum value of average luminance of each frame is greater than a given threshold; an over-dark video is the video whose maximum value of average luminance in the segment is too low; and an over-bright video is the video whose minimum value of average luminance in the segment is too high. Detection of all the types of lighting artifacts follows the same scheme described in Figure 2(a). To remove the lighting artifacts, different schemes are recommended -- lighting imbalance artifacts will be corrected as shown in Figure 2(c); over-bright video and over-dark video are restored utilizing the scheme shown in Figure 2(b).

3.1 Lighting Artifacts Detection

Lighting artifacts are a common feature in home videos. Since most home videos are shot without a lighting device, the brightness is unbalanced due to the incorrect orientation of camera lens with respect to the ambient lighting. A common error is to point the camcorder towards a light source like a window in a room or towards a lamp. If the camera is then pointed to other directions, there will be a severe imbalance in the brightness. Professional videographers avoid this problem by having mobile lighting equipment.

We first informally define lighting imbalance and then provide an exact computational procedure:

Definition 1. (*Lighting Imbalance*): If the luminance of the frames varies drastically in a segment causing an obvious variation in brightness to the human visual system, this variation is called lighting imbalance.

Thus, in a video with lighting imbalance artifact, some frames are very bright while other frames in the same segment appear very dark which causes obvious visual changes that are perceptually annoying. These frames certainly do not satisfy the users' needs and they have to be detected and restored.

We now describe a method to detect lighting imbalance automatically. To detect this imbalance in brightness, the average luminance of the video is first computed. Assume frame k of a home video is: $(f_r(i,j,k),\ f_g(i,j,k),\ f_b(i,j,k))$, $i=0,1,\ldots,W\text{-}1$, $j=0,1,\ldots,H\text{-}1$, $k=0,1,\ldots,L\ \text{-}1$. The luminance of a pixel is represented as (CIE-XYZ):

$$L(i,j,k) = 0.2627 * f_r(i,j,k) + 0.6558 * f_g(i,j,k) + 0.0815 * f_b(i,j,k) \quad (1)$$

If the luminance is given by YUV of MPEG video, the following YUV to RGB conversion formula can be used:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.140 \\ 1 & -0.394 & -0.581 \\ 1 & -2.032 & 0 \end{bmatrix} \begin{bmatrix} Y \\ U \\ V \end{bmatrix}$$

The average luminance of frame k in a video:

$$LF(k) = \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} L(i,j,k)/(W \cdot H) \quad (2)$$

After computing the average luminance for every frame, we can plot the temporal curve of the average brightness. From this curve, the range of average brightness can be easily determined with the maximum value LF_{Max} and the minimum value LF_{Min} of the average luminance.

$$LF_{Max} \geq LF(k), \; k = 0,...,L-1 \tag{3}$$

$$LF_{Min} \leq LF(k), \; k = 0,...,L-1 \tag{4}$$

$$d = | LF_{Max} - LF_{Min} | \tag{5}$$

where d is the distance between LF_{Max} and LF_{Min}. If the average luminance of video frames in a segment has low variation, e.g. $d<16$, that segment of the home video has balanced brightness. If the variation is large, it shows that the clip has a large variation in brightness and hence needs some adjustment. Basically, we use the average luminance of frames in order to consider the overall luminance distribution in the video. Our experiments show that this simple technique is quite adequate in detecting shots having lighting imbalance.

In order to detect the frames with over-bright and over-dark luminance in a home video, we employ the maximum value LF_{Max} and the minimum value LF_{Min} of the average luminance. In our procedure, if the maximum value LF_{Max} of the average luminance in a video segment is too low, e.g. $LF_{Max} < 16$, the home video is regarded as an over-dark video; if the minimum value LF_{Min} of the average luminance in a segment is too high, e.g. $LF_{Min} > 250$, the home video is regarded as an over-bright video. These thresholds have been arrived at after extensive empirical testing on real home videos.

3.2 Lighting Artifacts Removal

In section 3.1, the method to detect the video segments with lighting artifacts has been described. When such clips are detected and selected, we need to remove these artifacts.

From (3) and (4), the maximum value LF_{Max} and the minimum value LF_{Min} of average luminance in a home video are obtained, and from (2), the average luminance of frame k in a video is $LF(k)$.

Now suppose the acceptable brightness is denoted by β, if the average luminance of video frame $LF(k)$ is less than this value $\beta > LF(k)$, the compensation can be computed as $\Delta(k)=\beta - F(k)$, $\Delta(k)>0$, and we use it to boost the average luminance of the frame. Therefore a parameter t, $t\in[0,1]$ is calculated using $t=\Delta(k)/255$ and this parameter is used to adjust the average luminance in the frames with lighting imbalance. According to equations (6) and (7), luminance ($f_r(i,j)$, $f_g(i,j)$, $f_b(i,j)$) is transformed to ($f'_r(i,j)$, $f'_g(i,j)$, $f'_b(i,j)$):

$$\begin{bmatrix} f_r^{'}(i,j) \\ f_g^{'}(i,j) \\ f_b^{'}(i,j) \\ 1 \end{bmatrix} = \begin{bmatrix} 1.0-t & 0 & 0 & \Delta(k) \\ 0 & 1.0-t & 0 & \Delta(k) \\ 0 & 0 & 1.0-t & \Delta(k) \\ 0 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} f_r(i,j) \\ f_g(i,j) \\ f_b(i,j) \\ 1 \end{bmatrix} \tag{6}$$

If the average luminance of a video frame $LF(k)$ is less than β, then $\beta < LF(k)$, $\Delta(k)<0$, $t\in[-1,0]$ and the average luminance can be adjusted using:

$$\begin{bmatrix} f_r^{'}(i,j) \\ f_g^{'}(i,j) \\ f_b^{'}(i,j) \\ 1 \end{bmatrix} = \begin{bmatrix} 1.0+t & 0 & 0 & -\Delta(k) \\ 0 & 1.0+t & 0 & -\Delta(k) \\ 0 & 0 & 1.0+t & -\Delta(k) \\ 0 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} f_r(i,j) \\ f_g(i,j) \\ f_b(i,j) \\ 1 \end{bmatrix} \tag{7}$$

Assume the acceptable brightness is selected as: $\beta = LF_{Max}$, $\Delta(k)>0$. Since $(f.(i,j)(1.0-t)+255\cdot t)\in[0,255]$, $t\in[0,255]$, it ensures that the luminance of every pixel in the frames is not out of range. At the same time, the maximum luminance and the minimum luminance of every individual frame of the video come closer to each other. After adjusted the average luminance iteratively, the over-bright effect of the video will disappear. A video with a uniform and balanced luminance will result if the average luminance of frames is approximately equal.

One advantage of this simple lighting imbalance removal technique is that it is fast enough to process a home video in real time without any special hardware accelerator. Home video editing users do not need to wait long to see the restored results. Another advantage of this method is that it can also remove lighting imbalance from a single frame effectively when the lighting in the individual frame is unbalanced. For instance, if one region in a frame is very bright, and another region in the frame is very dark, after removed this artifact, not only is the lighting of whole video balanced, but also the contrast of the two regions in any frames of the video reduces. It may be regarded as a beneficial side-effect.

If the difference between the maximum value and the minimum value of the average luminance of the whole video is small, but the maximum value is too low or the minimum value is too high. i.e. $|LF_{Max}-LF_{Min}|<\varepsilon$ and $255-\varepsilon<LF_{Min}<LF_{Max}<255$ or $0<LF_{Min}<LF_{Max}<\varepsilon$, $\varepsilon\to 0$, $\varepsilon>0$. This indicates that the video is over-dark or over-bright. Either of these artifacts can be rectified using histogram equalization. A new index of pixel color values is computed according to the histogram of the frame. Suppose the histogram of a frame is G_i, $i = 0,1,2\ldots,255$, it is easy to compute the new mapping $i\to G_i$, $i = 0,1,2\ldots,255$ using:

$$C_i = \frac{256}{W \times H} \sum_{j=0}^{i} G_j \tag{8}$$

where W and H is the width and height of the frame respectively. With this transformation, the over-bright (or over-dark) lighting artifact will get removed. Our experimental results attest to the efficacy of this technique.

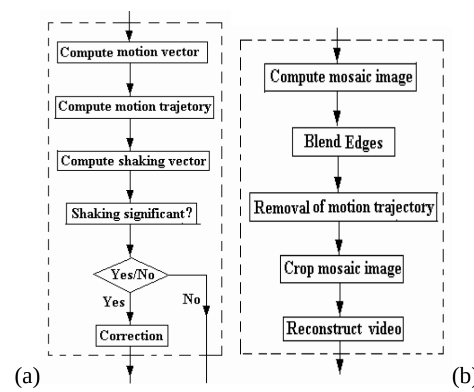


Figure 3. Flowchart for handling shakes in home videos (a) Shaking artifact detection (b) Shaking artifact removal

4. SHAKING ARTIFACTS HANDLING

Since home video takers are not professional photographers, they do not have special lighting devices, and they also do not have any special device that holds the camera while shooting. Moreover, with the current trend of miniaturization of camcorders, most home video camcorders are deliberately designed for hand-held use. One consequence of the hand-held camera is that it may be subjected to a jerky motion with the walking of the photographer. Moreover, home videos are sometimes shot from a moving vehicle like a tour bus. Therefore home videos tend to have many shots that are shaky. To detect shakes in videos is therefore a crucial step in home video editing. In this paper, we follow the steps shown in Figure 3 to detect and remove shaking artifacts in home videos.

4.1 Detection of Shaking Artifacts

Once again, we first provide an informal definition of the shaking artifact and then develop a computational algorithm in order to detect them.

Definition 2 (*Video Shaking*): If most objects in all the frames of a video clip move back and forth repeatedly along same directions during a short period, the video clip is said to be shaking.

In order to describe above definition clearly, a precise mathematical formulation is now given. Suppose a video clip consists of n frames, $V=\{F_0, F_1,\ldots,F_{n-1}\}$, the common region $\Omega=\bigcap_{i=0}^{n-1} F_i$ is divided into m parts, $\Omega=\bigcup_{i=0}^{m-1}\Omega_i$, $\Omega_i \cap \Omega_j=\emptyset$, the position of region Ω_i in the k-th frame of video is $(P_x(\Omega_i)|_{Fx}, P_y(\Omega_i)|_{Fx})$. The temporal sequence of these positions is the motion trajectory of that region Ω_i. If $\exists\Omega_i\subset\Omega$, $P_x(\Omega_i)|_{Fx+1}-P_x(\Omega_i)|_{Fx}>0$ and $P_x(\Omega_i)|_{Fx-1}-P_x(\Omega_i)|_{Fx}<0$; or, $P_y(\Omega_i)|_{Fy+1}-P_y(\Omega_i)|_{Fy}>0$ and $P_y(\Omega_i)|_{Fy-1}-P_y(\Omega_i)|_{Fy}<0$, then the video V is said to be shaking.

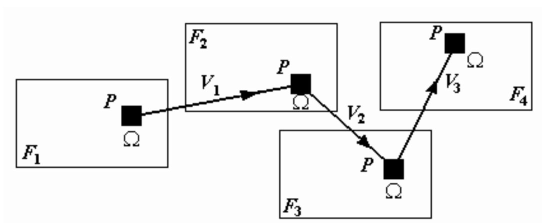


Figure 4. Trajectory of a region in a shaking video

In order to explain this formulation with a concrete example, consider Figure 4, which denotes successive frames of a shaky video. A region (depicted as a black square) is selected in the successive sequence F_r, r = 1,2,3,4 for a shaky video. The shaking vectors $\vec{V_r}$ which depict the motion of that region in successive frames, does not have a steady change --- the length and direction of these vectors vary rapidly. The inner product of two neighboring vectors is:

$$\vec{V}_{r-1}\bullet\vec{V}_r=|\vec{V}_{r-1}|\cdot|\vec{V}_r|\cos\alpha \qquad (9)$$

where α is the angle between vector $\vec{V_r}$ and $\vec{V_{r-1}}$. Thus,

$$\alpha=\cos^{-1}\left(\frac{\vec{V}_{r-1}\cdot\vec{V}_r}{|\vec{V}_{r-1}|\cdot|\vec{V}_r|}\right) \qquad (10)$$

If $\alpha\geq 90^o$ ($0\leq\alpha\leq 180^o$) in some frames F_r during a short period with a high frequency, e.g. if 20 frames in a video with 30 frames per second are shaking, we say the shaking is significant and the video is regarded as a shaking video. The points are named shaking points.

With this definition, we now provide an approach to detect shaking video segments. Assume that the position of a (manually or automatically) selected region in a frame r is $P(r)$ (motion trajectory), r = 0, 1, 2,…,L-1, L is the total number of frames in the home video. The change of the position of the selected region in frame r is:

$$\Delta P(r) = P(r) - P(0),\ r = 0,1,2,\ldots,L-1 \qquad (11)$$

The second-order difference of the change of position is given by:

$$\Delta^2 P(r) = P(r+1) - P(r) \qquad (12)$$

If

$$\alpha=\cos^{-1}\left(\frac{\Delta^2 P(r)\cdot\Delta^2 P(r+1)}{|\Delta^2 P(r)|\cdot|\Delta^2 P(r+1)|}\right)\geq 90^o$$

i.e. $\Delta^2 P(r)\cdot\Delta^2 P(r+1)\leq 0$, the point $P(r)$ is regarded as a shaking point. If such shaking points $P(r)$ appear in many frames during a segment, then this segment of video is said to have the shaking artifact.

If a home video is in the MPEG1 or MPEG2 format, shake detection of home video can benefit from the MPEG motion vectors [5]. However, not all video formats can provide motion vectors, thus image matching is crucial in order to determine the corresponding small region (black square of Figure 4) in all the frames. Generally speaking, manual matching is possible in a short video clip, but it is slow and cumbersome. A precise yet automatic matching would be useful for the corresponding position of the selected region in the next frame to be found quickly. However, video is an image sequence with audio, video frames with affine transformation and perspective transformation make the exact matching very difficult or impossible. Thus, the maximum likelihood block matching technique could be utilized [23].

Suppose a frame of video is $f(i,j)$, i= 0,1,…,W-1, j = 0,1,…,H-1; a region $f'(i, j)$, i= 0,1,…,M-1, j = 0,1,…,N-1 including the target object is searched in the sub-window $0\leq s\leq W-M-1$, $0\leq t\leq H-N-1$, $W > M$, $H > N$ of the frame; the distance (color distance, histogram distance, etc) between two windows in adjacent frames is given by the L_2 norm:

$$E(s, t) = \|f-f'\|_2 \qquad (13)$$

Thus $\quad E=E(s,t)_{\min},\ 0\leq s\leq W-M-1,\ 0\leq t\leq H-N-1 \qquad (14)$

After all the candidate regions are searched, the region with the best matching is obtained. When all the frames in the whole video compute their matched regions $P(m,n,l)=P(s,t)|_{E=E(s,t)_{\min}}$, $0\leq m\leq W-M-1$, $0\leq n\leq H-N-1$, $0\leq l\leq L-1$, L is the length (number of total frames) of the video segment. A coarse motion trajectory for that segment is established: $f(m,n,l)$. Generally speaking, this trajectory is not the best one, a better motion trajectory can be obtained by a process of iterative refinement:

$$f(m^k,n^k,l) = F^k(f(m^k,n^k,l)),\ \ |m^k - m^{k-1}| + |n^k - n^{k-1}|<\delta,\ \delta>0 \qquad (15)$$

where F^k is the refinement operation applied k times. After iterating around the initial trajectory several times, it eventually converges to a better motion trajectory given by:

$$|f(m^k,n^k,l) - f(m^{k-1},n^{k-1},l)| < \varepsilon,\ \ \varepsilon>0 \qquad (16)$$

The above procedure is summarized in the following algorithm:

(1) Select a target region Ω in the first frame of the segment.

(2) Find the best match region for Ω in every frame of the segment.

(3) The centroid coordinates of the matched region provide the initial motion trajectory.

(4) Fix a search window size δ for trajectory refinement.

(5) Repeat step (2) while searching within the corresponding search window centered around the found motion trajectory. This step computes the refined motion trajectory.

(6) Compute the distance between the last two motion trajectories $|f(m^k,n^k,l) - f(m^{k-1},n^{k-1},l)|$.

(7) If this distance is less than a threshold ε, then stop else go to step (5).

4.2 Removal of Shaking Artifacts

When a camera shake is detected, the next step is to remove the shaking artifact. We first make a key observation that since the shaking artifact manifests as a *jerky* motion trajectory, the frames within the shaky segment can be used to construct a mosaiced image that captures the scene information from all frames in that segment. If we want to correct the shaking artifact, the motion trajectory of any regions should be made *less* jerky, preferably should eliminate the jerky motion. Thus if a normal video segment has a normal motion trajectory, the motion trajectory with a shaking artifact can be considered as a normal trajectory corrupted by noise. Thus a filtering technique can be employed to de-noise such a trajectory in order to obtain a smoothed motion trajectory. If we have the mosaiced image corresponding to the segment and if we can smooth the motion trajectory, then we do the inverse process of re-constructing the video frames of the segment from the mosaiced image based on the smoothed motion trajectory. The resultant video segment will be significantly less jerky. Of course, perfect reconstruction of the frames is not possible since the scene is not static. But if the segment is small enough, the reconstruction will be quite accurate. Thus, the shaking artifact removal technique can be summarized as the following three steps:

(1) Construct a mosaic image for the frames of the segment.

(2) Smoothen the shaky motion trajectory by de-noising it.

(3) Reconstruct the segment by sampling the smoothed motion trajectory to carve out the corresponding frames from the mosaiced image.

We will now provide the precise steps with the appropriate mathematical framework for each of the three steps.

4.2.1 Image Mosaicing

We briefly describe the consideration in choosing an appropriate planar image mosaicing technique. We would like to point out that we have not developed any new mosaicing algorithm. But we have carefully chosen the most appropriate algorithm for our purpose since the quality of the mosaicing will directly impact the video reconstruction step. When a big image is constructed out of a sequence of frames from a video segment, the seam between two overlapping frames will appear [12, 17]. General methods for erasing these edges substitute the seam region by the average value between the parts of frames that appear in this region [12]. But this will result in a transparent mosaic, thus the method does not fit our needs, and hence we do not adopt this approach.
To blend different images into a seamless panorama, we need to smoothen all these illumination discontinuities, while preserving image sharpness. A method that fulfills this requirement is described in [1]. In this approach, the frames are decomposed into band-pass pyramid levels, and then combined at each band-pass pyramid level. Final correction of the images from the combined band-pass levels gives the desired panorama.
The alignment to translate frames is easy to be computed, however, frames of home video accompanied with affine transformations and perspective transformations have distortions. For this, the inverse transformations are needed. In [22], the authors describe a method to mosaic images with affine transformation and perspective transformation. Both global and local alignment methods are given. Here we only consider local alignment.

4.2.2 Motion Trajectory Smoothing

If the shaking artifact is regarded as the noise of the motion trajectory of a normal video, to correct the shake may be regarded as de-noising the motion trajectory. The simplest method would be to use polyline simplification [13], which is also called mean filtering. This method substitutes a shaking vector by the average shaking vector of the neighboring shaking vectors. The Bezier curve [4], a kind of B-Spline, can also perform this task because the Bezier curve has some desirable properties such as the convex hull property, existence of derivatives, interpolation of the beginning and end points, ease of computation and ease of control. What is most important is that the interpolating cubic Bezier curve has C^2 continuity. This interpolation thus acts like a noise filter for a jerky trajectory. These properties make the Bezier curve ideal for removal of noise from the motion trajectory.
When polyline simplification is substituted by Bezier-fitting, the turning points of motion trajectory in shaking video are replaced by points on Bezier curve, the trajectory of home video becomes smooth, shaking artifacts are removed. In fact, the de Casteljau algorithm to implement Bezier curves is a recursive procedure for the subdivision of a polyline into a curve [4].
We now provide the details about smoothing motion trajectory of shaking home video. It is possible to find the maximum value $f_{max}(x,y,t)$ and the minimum value $f_{min}(x,y,t)$ of a motion trajectory curve $f(x,y,t)$ within a period $[t_0,t_1]$, where $0 \le t_0 \le t \le t_1 \le T$, T is time axis of a video. If $t_{max} \le t_{min}$, four tuples can be constructed:
$(t_1, f(x,y,t_0))$, $(t_{\max}, f(x,y,t_{\max}))$, $(t_{\min}, f(x,y,t_{\min}))$, $(t_1, f(x,y,t_1))$.
The four tuples can be blended using the function:

$$B(x,y,t)=B_0^3(t_0)f(x,y,t_0)+B_1^3(t_{\max})f_{\max}(x,y,t_{\max})+B_2^3(t_{\min})f_{\min}(x,y,t_{\min})+B_3^3(t_1)f(x,y,t_1) \quad (17)$$

where $B_i^n(t)$, $n=3$, $i=0,1,2,3$ are the Bernstein polynomials [4].
The blending function gives a smoothed motion trajectory that is guaranteed to lie within the convex hull of the four tuples. Figure 5 is an illustration of the shaking artifacts removal algorithm. Figure 5(a) is the graph of mean filtering. Figure 5(b) is an illustration of Bezier curve method. Bezier curve, as a free style curve, is defined by four control points (t_0, f_0), (t_2, f_2), (t_3, f_3), (t_1, f_1); (t_2, f_2) and (t_3, f_3) is selected from (t_{max}, f_{max}) and (t_{min}, f_{min}) according to their orders. i.e. if $t_{max} \le t_{min}$, then $t_3 = t_{min}$, $f_3 = f_{min}$, $t_2 = t_{max}$, $f_2 = f_{max}$, else $t_{max} \ge t_{min}$, then $t_3 = t_{max}$, $f_3 = f_{max}$, $t_2 = t_{min}$, $f_2 = f_{min}$.

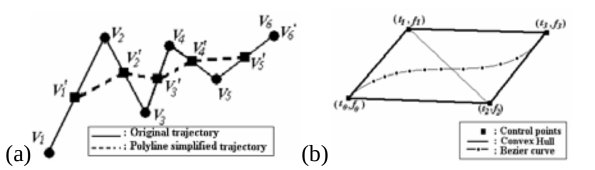
In general, quadric B-spline curves [4] are computed as:

$$P(t)=(1,t,t^2)M(P_i,P_{i+1},P_{i+2})^T \quad (18)$$

P_i, $i=0,1,2\ldots,n$ are control points and $t \in [0,1]$.

$$M=\begin{bmatrix} 1/2 & -1 & 1/2 \\ -1 & 1 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Thus any points $P(t)$ of the smoothed motion trajectory are computed using the above expression.



(a) (b)

Figure 5. Motion trajectory filtering (a) Mean filtering with polyline of motion trajectory, (b) Bezier curve based filtering and its convex hull.

The advantage of using the Bernstein basis is that because of the convex hull property, any interpolated points on the smoothed trajectory will indeed correspond to a valid frame lying within the bounds of the mosaiced image (as illustrated in Figure 6(d)). If we had used any other basis for filtering, perhaps frame extrapolation would have been necessary which is error-prone. Note that the Bezier-curve based trajectory smoothing procedure is iteratively applied to obtain a progressively smoother motion trajectory. This procedure is described as the following algorithm:

(1) Divide the motion trajectory into proper pieces corresponding to at least four points.

(2) Find the maximum and minimum coordinate values of the selected trajectory points.

(3) Use the four extremal points as Bezier control points in order to obtain a de-noised trajectory.

(4) If iterative smoothing is desired, then resample the smoothed trajectory and repeat steps (1)-(3).

(5) Return the coordinates of new points for the final smoothed motion trajectory by resampling the curve.

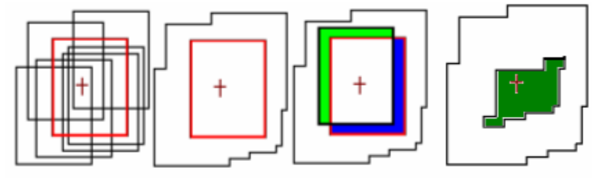
4.2.3 Video Reconstruction

In Figure 6(a), the black-edge images are the frames of the shaking video segment. They are used to construct a mosaic image. A new video without shaking artifacts (or with a small acceptable amount of shaking) is constructed by the region with red–edge in Figure 6(b). Figure 6(c) illustrates a frame of video (with the black-edge), the selected region and the mosaic image, the green part is cropped from the video frame, and the blue part is added to the frame. When all the frames in a new sequence are determined, the reconstructed video is generated.

This procedure can be mathematically described as follows: suppose frames of a home video segment have their domains Ω_i, i=0,1…,L-1, L is the number of total frames. After the image mosaicing step, a region belonging to a synthesized image is constructed:

$$\Omega = \bigcup_{i=0}^{L-1} \Omega_i \tag{19}$$

and our expected domain of every frame is $\Omega'_i \subset \Omega$, from which a new video clip is created. What we should note is that: $\Omega'_i \cap \Omega = \Omega'_i$, therefore, the boundary does exist. The image outside the boundary is not considered while cropping the mosaiced image.



(a) (b) (c) (d)

Figure 6. Construction of a new video frame (a) Overlapping frames (b) The mosaiced image. (c) The blue part is added and the green part is cropped. (d) The green part is the boundary of motion trajectory.

Suppose a frame of video is Ω_i, the expected domain is Ω'_i, the part $\Omega'_i \cap \Omega_i \subset \Omega'_i$ is common area and should not be considered. We pay our attention on the part $\Omega_i - (\Omega'_i \cap \Omega_i)$ and $\Omega'_i - (\Omega'_i \cap \Omega_i)$, the former is the part we should cut it from original frame and the latter is the part we should add it to the new frame. In case of new frame out of range i.e. $(\Omega'_i - \Omega'_I \cap \Omega_i) \cap \Omega \not\subset \Omega$, mean filtering based on polyline simplification and curves-fitting based on Bezier curve interpolating are considered, for their properties ensure the new frame in the boundary i.e.$(\Omega'_i - \Omega'_i \cap \Omega_i) \cap \Omega \subset \Omega$.

If a motion trajectory of a shaking video is constructed properly, we crop a number of new frames for a stable video from the mosaiced image at any point on the continuous motion trajectory. We only select the frames that fit our needs and reconstruct frames corresponding to the integer points by interval [t_0, t_1] in (17) which now become the frames of the restored stable video.

If the video has moving objects for a significantly long duration, there are some shortcomings of our mosaic-based method. Such videos will require construction of a huge mosaic image. Even for a short shot with a moving object, we cannot construct the mosaic accurately. Therefore, we make a minor modification to the mosaicing procedure and only consider mosaicing a few temporally adjacent frames of the current frame (e.g. 15 frames before and 15 frames after). We can thus do piece-wise artifact removal for the shot and the whole video. When we finally put together the corrected frames, the shakes will be reduced significantly. Thus, the change is to consider a 30-frame segment instead of a shot in our algorithm.

5. EXPERIMENTAL RESULTS

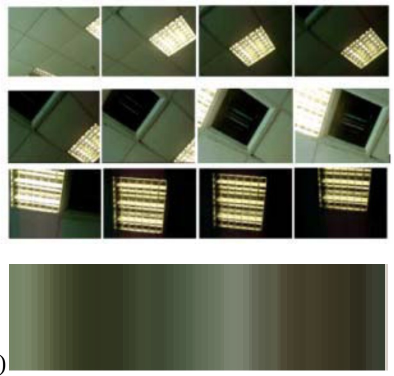
In this section, we will present our experimental results and all the videos (both the original and the restored) are available for viewing at our website [28]. These videos were shot by us to illustrate the various types of artifacts described.

5.1 Lighting Artifact Results

Figure 7(a) shows a few sample frames of a home video with lighting imbalance. In Figure 7(a), the luminance of the first row of frames varies from over-bright to over-dark; in the second row, the luminance of frames diversifies from over-dark to over-bright; and in the third row, the luminance changes from dark to over-dark. Figure 7(b) is a color bar to show the average color variation of each frame in the whole video. Each vertical bar corresponds to the average color of one frame of the video.

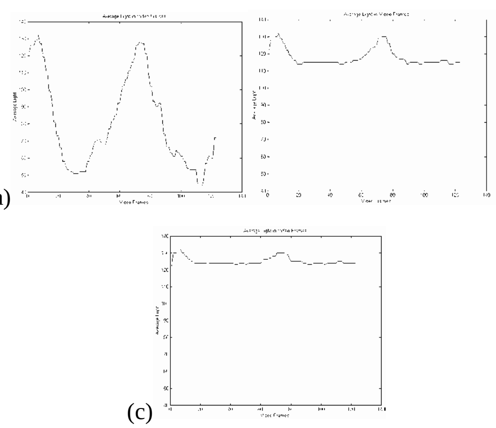
The diagram of average luminance variation of the samples in Figure 7 is provided in Figure 8(a); Figure 8(b) is the diagram of average luminance after once correction; Figure 8(c) is the diagram of average luminance after performing the correction twice.

Figure 9(a) shows a frame in the video, which is over-dark; Figure 9(b) shows the corresponding frame after the lighting artifacts removed; Figure 9(c) shows the color bar of the corrected video removed the lighting artifacts.



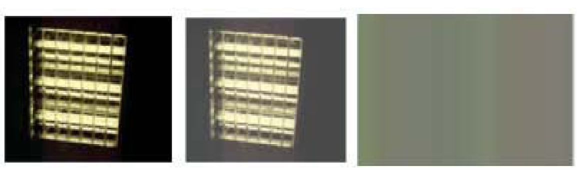
(a) (b)

Figure 7. Original frames of video about ceiling with lighting imbalance artifacts (a) Example frames (b) Color bar of the original video.



(a) (b) (c)

Figure 8. Diagram of average luminance of every frame in the sample video. (a) Average luminance of every frame without lighting artifacts removal. (b) Average luminance of every frame after correction once (c) Average luminance of every frame in the video after correcting twice.



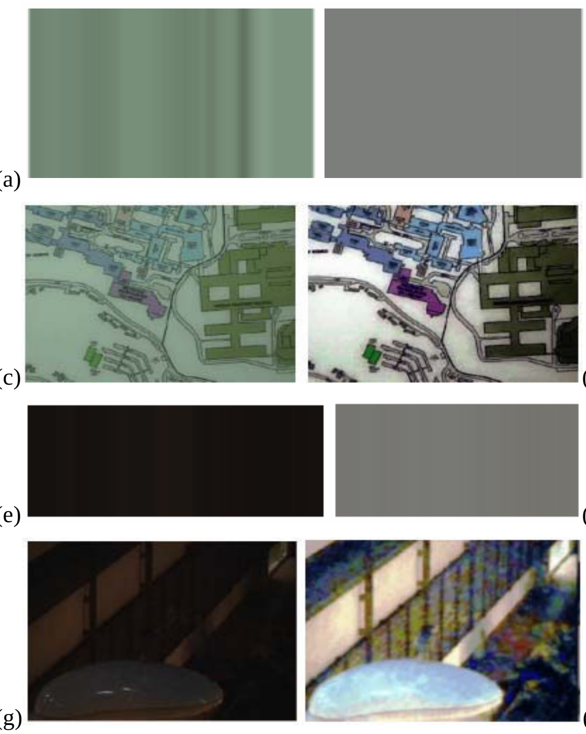
(a) **(b)** **(c)**

Figure 9. Frame with removed lighting artifacts in the video with unbalanced luminance. (a) An over-dark frame (b) The corresponding frame with the lighting artifact removed (c) Color bar after the lighting artifacts removed.

From Figure 9, it is clear that the algorithm for correcting lighting imbalance is efficient and effective. The maximum luminance and the minimum luminance of frames in the video are close to each other after two iterations.

To removal lighting artifacts from frames with too high or too low luminance without imbalance luminance, the method based on histogram equalization is used. Figure 10 gives experimental results. Figure 10(a) is the color bar of sample video with too high brightness. This video is about a brilliantly lighted campus map of National University of Singapore (NUS) and was shot at one of the shuttle bus stops at night. Figure 10(b) is the color bar of the corrected video. Figure 10(c) is one of frames in the raw video and figure 10(d) is the corresponding frame after correction.

Figure 10(e) is a color bar of an over-dark sample video. The video is about a bird on a street lamp post. In Figure 10 (g), one of frames shown is dark, and almost nothing can be seen there. Figure 10(h) demonstrates the corresponding result after correction. The shapes of the lamp and the bird can be seen clearly. Figure 10(f) is the color bar of the sample video after correction. For this video, we have found that post-processing by blurring generates even better results.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 10. Artifacts removal from home video with too high or too low luminance by histogram equalization. (a) Color bar of over-bright video. (b) Color bar of corrected over-bright video. (c) Example frame from original over-bright video (d) The corresponding frame in the corrected video (e) Color bar of an over-dark video (f) Color bar of the corrected over-dark video (g) Example frame from original over-dark video (h) The corresponding frame from the corrected video.

5.2 Shaking Artifacts Results

A video about an office file cabinet is provided as a shaking home video example in Figure 11. Figure 12(a) is the graph of shaking vectors of the video frames in horizontal and vertical directions when compared to the first frame in the video. From the curve, it can be seen that at first, the camera is moving towards right and bottom direction, and then the camera is moving towards the right direction. In Figure 12(b), the shaking vectors of video frames after mean filtering are given while Figure 12(c) gives the shaking vectors in the video frames after Bezier filtering which clearly shows its superiority.

Figure 13(a) is the mosaic image, this image combines all the frames together according to their positions in each frame of the shaking video and the red rectangle shows that the area that will be cropped as new video frames. Figure 13(b) is the same mosaic image with edge blending and it is definitely of a better quality.

Figure 14(a) shows one reconstructed frames from the mosaic image of Figure 13(a). It can be seen that some color in the frame is not quite right. Figure 14(b) shows the same frame cropped from the mosaic image with blending edges of Figure 13(b). The sharp edges of the frame in Figure 14(a) disappear. The interested reader can look at our website for some more results of artifact removal on other videos which cannot be shown here due to space limitations.

Figure 11. Some frames from a shaking video about file cabinet

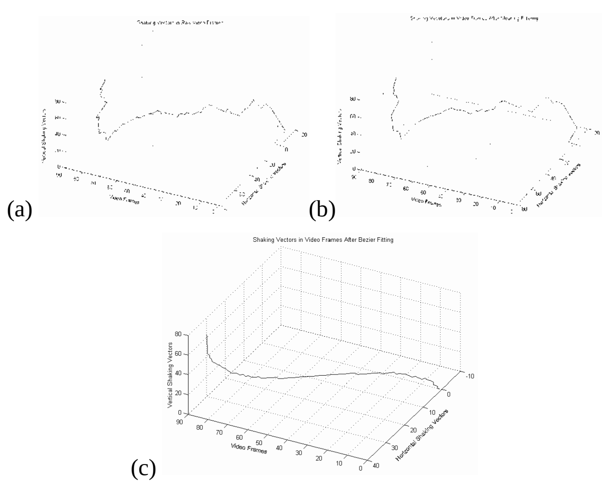


Figure 12. Shaking vectors in the video frames. (a) Shaking vectors in video frames. (b) Corrected shaking vectors in video frames after mean filtering. (c) Corrected shaking vectors in video frames after Bezier fitting.

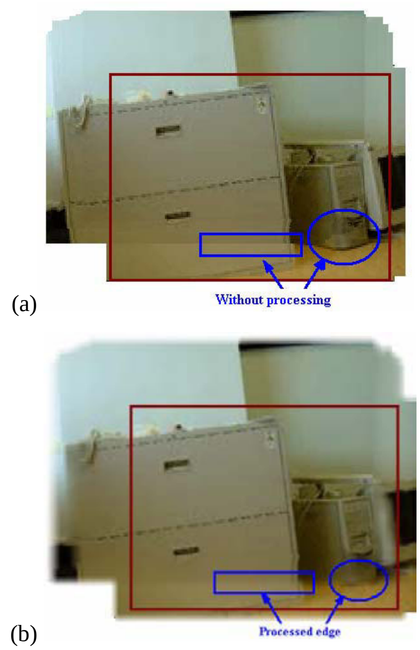


Figure 13. The mosaiced image. The region with a rectangle is the cropped region used to construct a new video clip. (a) Mosaic image with direct frame overlapping. (b) Mosaic image with blended edge frame overlapping.

Figure 14. Cropped frames from the mosaiced image. (a) Frame from the directly overlapping mosaic. (b) Frame from edge-blended mosaic image.

6. CONCLUSIONS

In this paper we have described a method to detect and correct the lighting and shaking artifacts of home videos. These artifacts are quite common given the usual skills and equipment of an average home video user. For the detection and removal of lighting artifacts, the average luminance model of frames is first computed. Based on this, we have developed procedures for detecting video segments which have lighting imbalance or which are over-dark and over-bright. We then give procedures for correcting the lighting artifacts. Our experimental results demonstrate the utility of the proposed methods.

For the detection and removal of shaking artifacts in home videos, the shaking of a home video is considered to be the noise of the motion trajectory and two methods to filter the motion trajectory based on polyline simplification and Bezier curve are given. Compared with general polynomials for the least square approximation, Bezier curves have many useful properties, like the interpolation of cubic Bezier curves satisfy C^2 continuity [4] and the convex hull property. These make the reconstructed frames of a home video to lie within the boundary of the mosaiced image constructed out of the frames in the shaky video segment.

Our experiments on real home videos with shaking artifacts provide compelling evidence for the efficacy of the proposed method.

Our future work is to further improve the quality of our restoration algorithms. As mentioned in the introduction, one other artifact occurring in home videos is the *blurring* artifact. We are trying to develop a method to detect the blurring artifact and eventual restore videos with this artifact. Our long-term goal is to provide an intuitive restoration utility that can correct the visually annoying artifacts of home videos that can be used by any home video user.

7. ACKNOWLEDGMENTS

We are very grateful to our colleagues Ye Shi-Ren and Oon Wee Chong for their help in obtaining the video data for our experiments. We would also like to thank Shih-Fu Chang of Columbia University and the anonymous reviewers for their suggestions and constructive comments.

8. REFERENCES

[1] Burt, P. J. and Adelson, E. H. A multiresolution spline with application to image mosaics, ACM Transaction on Graphics, Vol. 2, No. 4, October 1983, 217-236.

[2] Casares, J., Myers, B. A., and Long, A. C.. Bhatnagar, R. Stevens, S. M. Dabbish, L. Corbett, A. Simplifying video editing with intelligent interaction (submitted).

http://www-2.cs.cmu.edu/~silver/publications.html.

[3] Chen. T.. Video stabilization algorithm using a block-based parameteric motion model. Project report, 2000. http://www-isl.stanford.edu/~tingchen

[4] Farin. G.. Curves and surfaces for computer aided geometric design: a practical guide (3rd Edition). Academic Press, Imprint Boston, 1992.

[5] Filippo V., Alfio C., Massimo M. and Giuseppe M. Robust digital image stabilizing algorithm using block motion vectors, in Proceedings of International Coneference on Consumer Electronics(ICCE'02), (Los Angeles, June, 2002), 234-235.

[6] Girgensohn, A., Bly, S., Shipman, F., Boreczky, J. and Wilcox, L.. Home video editing made easy-balancing automation and user control, in Proceedings of INTERACT '01, IOS Press, 464-471, 2001.

[7] Girgensohn, A., Boreczky, J., Chiu, P., Doherty, J., Foote, J., Golovchinsky, G., Uchihashi, S. and Wilcox, L.. A semi-automatic approach to home video editing, in Proceedings of UIST '00 (San Diego, CA USA, 2000), ACM Press, 81-89.

[8] Gove R. J., Video stabilization system and method. Texas Instruments. United States Patent 5973733, October 26, 1999.

[9] Hansen; M. W. and McDowell L. K. Apparatus and method for removing blank areas from real-time stabilized images by inserting background information. Sarnoff Corporation, Princeton New Jersey. United States Patent 6211913, April 3, 2001.

[10] Jimber, D., Foote, J. and Lertsithichai, S. FlyAbout: spatially indexed panoramic video, in Proceedings of ACM Multimedia 2001 (Ottawa, CA, Oct. 2001).

[11] Jin, J., Zhu, Z. G., Xu, G.Y and Yang, Y.D. Digital video sequence stabilization based on 2.5D motion estimation and inertial motion filtering, Real-Time Imaging Journal, Vol. 7, No. 4, August 2001, Academic Press, 357-365.

[12] Jones, R. C., DeMenthon, D. and Doermann, D. S. Building mosaics from video using MPEG motion vectors, in Proceedings of ACM Multimedia 1999 (Orlando, 1999).

[13] Latecki, L. J., Wildt, D. and Hu. J. Extraction of key frames from videos by optimal color composition matching and polygon simplification, in Proceedings of the workshop on multimedia signal processing (Cannes, October 2001).

[14] Lienhart, R. Abstracting home video automatically, In proceedings of ACM Multimedia 1999(Orlando, 1999).

[15] Lienhart, R. Dynamic video summarization of home video, in Proceedings of SPIE 3972: Storage and Retrieval for Media Databases 2000, 378-389, Jan. 2000.

[16] Ma, W.Y. and Zhang, H. J. An indexing and browsing system for home video, in Proceedings of EUSIPCO'2000, (Tampere, Sept. 2000), 5-8.

[17] McMillan, L. and Bishop, G. Plenoptic modeling: an image-based rendering system, in Proceedings of SIGGRAPH 95 (Los Angels, California, August 6-11,1995), 1995, 39-46.

[18] Morimoto, C. H. and Chellapa R.. Fast electronic digital image stabilization, in: Proceedings of IEEE International Conference on Pattern Recognition, Vienna, Aug. 1996.

[19] Morimoto, C. H. and Chellapa R. Fast electronic digital image stabilization for off-road navigation, to appear in IEE Journal of Real Time Image Processing.

http://www.cfar.umd.edu/~carlos

[20] Sato T. Camera with a shake detection function. Olympus Optical Co.,Ltd., Japan. United States Patent 6067418, May 23, 2000.

[21] Shiomi Y. Image stabilizing device. Canon Kabushiki Kaisha, Japan. United States Patent 6282376, August 28, 2001.

[22] Shum, H.Y. and Szeliski, R. Construction of panoramic image mosaics with global and local alignment, in Proceedings of IJCV (Feb. 2000), 36(2), 101-130.

[23] Tekalp, A.M. Digital video processing, Prentice Hall, 1995.

[24] Yamazaki T. and Ishikawa Y. Image-shake preventing apparatus. Canon Kabushiki Kaisha, Japan, United States Patent 6263162, July 17, 2001.

[25] Zhang, H.J., Kankanhalli, A. and Smoliar, S.W.. Automatic partitioning of full-motion video, ACM Multimedia System, 1993, 1, 10-28.

[26] Operating Instructions of Sony Digital Video Camera Recorder. DCR-TRV15E/TRV17E. Sony Corporation, 2001.

[27] Operating Instructions of Sony Digital Video Camera Recorder. DCR-PC101E. Sony Corporation, 2002.

[28] http://www.comp.nus.edu.sg/~yanwq/rahv.html