Appl. Math. Inf. Sci. 7 No. 1S pp. 43S-49S (2013)

Applied Mathematics & Information Sciences
An International Journal

A Shot Detection Approach to Synchronize Stereoscopic Video

Shi-Wei Lo[1, 2] and Fang-Pang Lin[1]
[1] National Center for High-performance Computing, Hsinchu, Taiwan
[2] Department of Biomedical Engineering and Environmental Sciences, National Tsing Hua University, Hsinchu, Taiwan
Corresponding author: Email: shiwei.lo@gmail.com

Received Nov. 25, 2011; Revised Jan. 8, 2012; Accepted Jan. 31, 2011
Published online: 2012

Abstract: One of the challenging problems in stereoscopic (3D) video playback is the synchronization of paired-video of binocular vision. This paper deals with the issue of synchronization of binocular video that acquire from dual-camera system. Arranging a paired-video into an identical timeline has been suggested as one mechanism for synchronizing video information. This approach requires a suitable method to automatically locate synchronized points in a paired-video. The developed method of finding such synchronized points is to determine the similarity of temporal feature curve between pair of "*Left*" and "*Right*" video in timeline. In this paper, we address a shot change detection approach to determine synchronized points and present three tests to examine hypothesis. Experimental results on real date are presented. The experimental results show that the statistical approach permits accurate detection of synchronized points, in particular, the performance of the approach was found extremely satisfactory with determine all synchronized points in longer video sequence.

Keywords: Shot Detection, Stereoscopic Video, Synchronization

1 Introduction

The 3D (or stereoscopic) video is becoming more and more common today. As far as 3D video is concerned, the primary requirement is the synchronization of paired-video of binocular vision when playback. The synchronizing can be achieved using dedicated hardware to ensure the paired-video are recorded in a synchronized way. However, this is only possible for professional applications such as commercial film, industrial recorded and major sports broadcasting, where more expensive equipment can be used. Another way is to manually perform the synchronization, but it is time and manpower consuming. Moreover, it does not always produce sufficiently accurate results.

Tuytelaars and Van Gool [1] explain how to synchronize two video sequences of the same event when cameras are non-static. They compute a rigidity measure of frames then find the best fitting epipolar geometry, and back projecting the epipolar lines into the 3D world and looking for intersecting lines. Sinha and Pollefeys [2] propose an automatic approach to synchronize a network of uncalibrated and unsynchronized video cameras. They also compute the epipolar geometry but form dynamic silhouettes, to deal with unsynchronized sequences and find the temporal offset between them. Moreover, to synchronize the motion capture and video, Sigal, et al. [3] manually labeled visible markers on the body. These labeled can be used in the optimization procedure of synchronization. Akbarzadeh et al. [4] using hardware, IEEE-1394 sync units, to directly synchronize the multiple cameras. Therefore, each video sequences recorded respectively are synchronized. Several researchers already looked into the problem of video synchronization in network- and multi- camera system. However, there is lack of simple approach to automatically synchronize the paired-video.

In spite of the several approaches proposed in the literature currently available, research on automatic multi-cameras synchronization is ongoing, and even commercial and freeware applications for manipulating synchronization are being deployed. "*StereoMovie Maker*" functions both as a versatile stereo movie editor and stereo movie player (*http://stereo.jpn.org/eng/stvmkr/*). It allow user to manually displace one of binocular video timeline relative to the other in order to effect synchronization. There is some commercial package, "*Adobe® Premiere®*" and "*Cineform Neo3D*" (*http://www.adobe.com/Premiere*), these software together offer a workflow for editing stereoscopic video content and 3D effects with manual timeline arrangement. Moreover, *Sony Vegas* (*http://www.sonycreativesoftware.com/*) use synchronized timecode to synchronizing same shooting event and also functions with manually align the events in the timeline. In our previous work, the script-based batch, "*StereoImageMaker"* (*http://sourceforge.net/projects/tdwtool/*) is a shell script batch that using *ImageMagick*'s tools to stitch pair image sequence from frame-based but it also need manual procedure to find out the synchronization in timeline before stitching the pair images. There is deficient of automatic synchronization approach for unsynchronized paired-video.

Stereoscopic video is created using two synchronized video cameras generally produced paired-video (left and right video). These cameras are used to emulate the human eyes, and have similar properties such as separation, vergence and field of view. However, the missing of synchronization is leading human visual discomfort. With a stereoscopic video, playback is generally very important as it requires specialized hardware. For these reason, one of the challenging problems in stereoscopic video playback is the synchronization of paired-video.

The paired-video consists of Left and Right video (Side-by-Side mode) that is created with same scene but slightly different camera properties, the content (scene) of video have a great similarity. A shot or take in video refers to a contiguous recording of one or more video frames depicting a continuous action in time and space [5]. Shot transitions can be divided into two categories: abrupt transitions and gradual transitions. Gradual transitions include camera movement and video editing special effects, and abrupt shot transitions are known as hard cuts which are very easy to detect since the two frames [6]. In our approach, we chose the abrupt shot detection method because of the video is raw and not include any editing effects. The abrupt shot detection is to detect the change between two frames. The histogram-based Chi-square test is the most common and reliable method to detect shot [5, 7]. Histograms are less sensitive to object motion and disparity noise between two views of cameras.

The shot detection result represents a feature curve of scene transformation of video sequences. These two feature curves from Left and Right video can be matching in timeline by using Hausdorff distance method. The Hausdorff distance can be used in a wide range of applications, such as real scene recognition, tracking, engineering drawing understanding, and aerial image analysis [8].

If the shot boundaries from paired-video are identical in timeline, then using shot boundary to be a synchronized point will more easy to synchronize the timeline of paired-video. Unfortunately, in normal recording with unsynchronized cameras, it does not guarantee the shot boundaries are identical in timeline. In this approach, we do not detect the cut or boundary of video sequence but using the shot detection characteristic value to form a feature curve in timeline.

In this paper, we introduce an approach for automatic synchronize side-by-side stereoscopic video captured by two independent unsynchronized cameras. The results of this experiment clearly showed that we achieved the best accuracy and most consistent results when video sequences were rather longer and varied.

2 Shot Detection and Timeline Matching

The purpose of this paper was to determine the shot detection method to establish the synchronized points for paired-video sequences. And to produce a feature curve of video which could be usable for evaluating the synchronized points of paired-video.

2.1 Shot Detection

The chi-square test is the most accepted test for comparing two binned distributions and be used in shot change detection [5]. To decision a shot every frames of the sequence is characterized by a values. This value includes disparity values, which indicate the degree of similarity between consecutive frames based on histograms. In this paper, the disparity values were used to represent the feature curve of video in time domain.

The chi-square test (X^2) used for the area between the distributions is given in (1) as proposed by [9].

$$X^2 = \sum (R_i - S_i)^2/(R_i + S_i) \quad (1)$$

Where i is number of gray levels (bins), R_i the number of events in bin i for first frame, S_i the number of events in the same bin i for the second frame. For chi-square value, a low score represents a better match than a high score. If this value is large enough, the two processed frames are declared to belong to different shots.

The paired-video processed by chi-square test and chi-square value exporting as the time serial feature curves. The feature curve represents the time variance of scene of video that used to determine the synchronized points of paired-video.

2.2 Timeline Matching

The Hausdorff distance measures the distances between two sets of points. Thus, it can be used to measure the similarity between two patterns of points when they are superimposed on one other [8]. Here the Hausdorff distance was used to searching synchronized points in feature curves of chi-square value.

Given two point sets A and B, the Hausdorff distance between A and B is given in (2).

$$H(A,B) = max(h(A,B),h(B,A)), \quad (2)$$

where $h(A,B) = max_{a \in A} min_{b \in B} \| a - b \|$, from A to B. And $h(B,A)$ is from B to A.

The Hausdorff distance detects the similarity of the feature curves, the point sets of chi-square value, as synchronized points. The seeking of synchronized points is accomplished by using chi-square value as feature curves represented in timeline.

3 Dataset

The experimental dataset consisted of the sequences listed in Table I. The test sample was extracted from the source sequences which comprise single shot only; such test sets were able to eliminate the shot boundary which is ideal synchronize point. The test samples were complex with extensive motion and no graphical effects.

The source sequences have well synchronized and stored with side-by-side format. Each test sample extracted from the source sequences has variance frames; *Baku_block*, *Baku_river* and *740D_tour* have 290, 650 and 920 frames respectively. All test video samples have been separated to left and right part, say paired-video, and set frame shift to 0, 5, 10, 20, 30, 60, and 120 to simulate the synchronized and non-synchronized condition.

The *Baku_block* have least frames in three test samples, scene consisted of camera tile/zoom to a block of building. The *Baku_river* consisted of camera tile/pan to a river. The *740D_tour* have longest video length (920 frames) consisted of camera tracking a car in tour of driving.

4 Results

We have validated our approach by experiments with a variety of video contents. The results of this experiment were reported here: *Baku_block* in figure 1, *Baku_river* in figure 2 and *740D_tour* in figure 3. The searching of synchronized points which used Hausdorff distance on feature curves of paired-video set was evaluated. Each lowest peak in results represented the most possible synchronized points.

5 Discussion

The proposed approach which used shot change detection and Hausdorff distance on three video set were evaluated.

For the *Baku_block* (figure 1) sequence, the reported synchronized points were found where frame shift less than 30 frames. The frame shift, for example, when set to 30 frames shift from the start then the paired-video were also cut 30 frames from the end. This is because left and right video must make sure have same video frames for combination and playback. The *Baku_block* have 290 frames, in case shifting 30 frames, the feature curve only remained 230 frames. When frame shifting more than 60, 90, 120 then feature curve remained 170, 110, 50 frames respectively. If total frames of feature curve less than 230, this results in a lack of basis for searching the features.

In *Baku_river* (figure 2) sequence found the same situation as *Baku_block*, because *Baku_river* have more frames (650) the synchronized points found when shifting frame less than 60 frames.

The *740D_tour* have longest video length (920 frames) and also have more complex scene, all the all synchronized points found in all shifting frames condition (shifting from 0 to 120 frames). Thus, the shorter video sequence in this approach will miss some synchronized points, but in longer and varied sequence it achieved best accuracy. This was caused by longer and varied video sequences have a large number of chi-square values and resulted in significantly variation of Hausdorff distance of feature curves.

6 Conclusions and future work

We have described in this paper an original and efficient approach to synchronize stereoscopic (3D) video. It involves the chi-square test and the characterization of feature curve. We have evaluated the three test sequences in this paper. These results indicate the conditions under which longer and varied video sequences can provide best accuracy of synchronized points of paired-video. Furthermore, the estimation of the feature-based curve model involved in the proposed technique can also be exploited for further video analysis steps such as video retrieval, video matching and video database.

TABLE I. VIDEO SOURCE AND TEST SAMPLE PROPERTIES.

Source Sequence	Width * Height	FPS	Codec	Length	Test Sample	During (frames)	Frames
Baku_3D_LR[a]	1920*1080	25	XVID MPEG-4	28"33	*Baku_block*	7735-8025	290
Baku_3D_LR[a]	1920*1080	25	XVID MPEG-4	28"33	*Baku_river*	28900-29550	650
740D_3D_LR[b]	2560*720	24	H.264	9"7	*740D_tour*	7756-8676	920

a. The Baku_3D_LR : 3d2010ru (*http://3d2010.ru/*) b. The 740D_3D_LR : 3D Vision Blog (*http://3dvision-blog.com*)

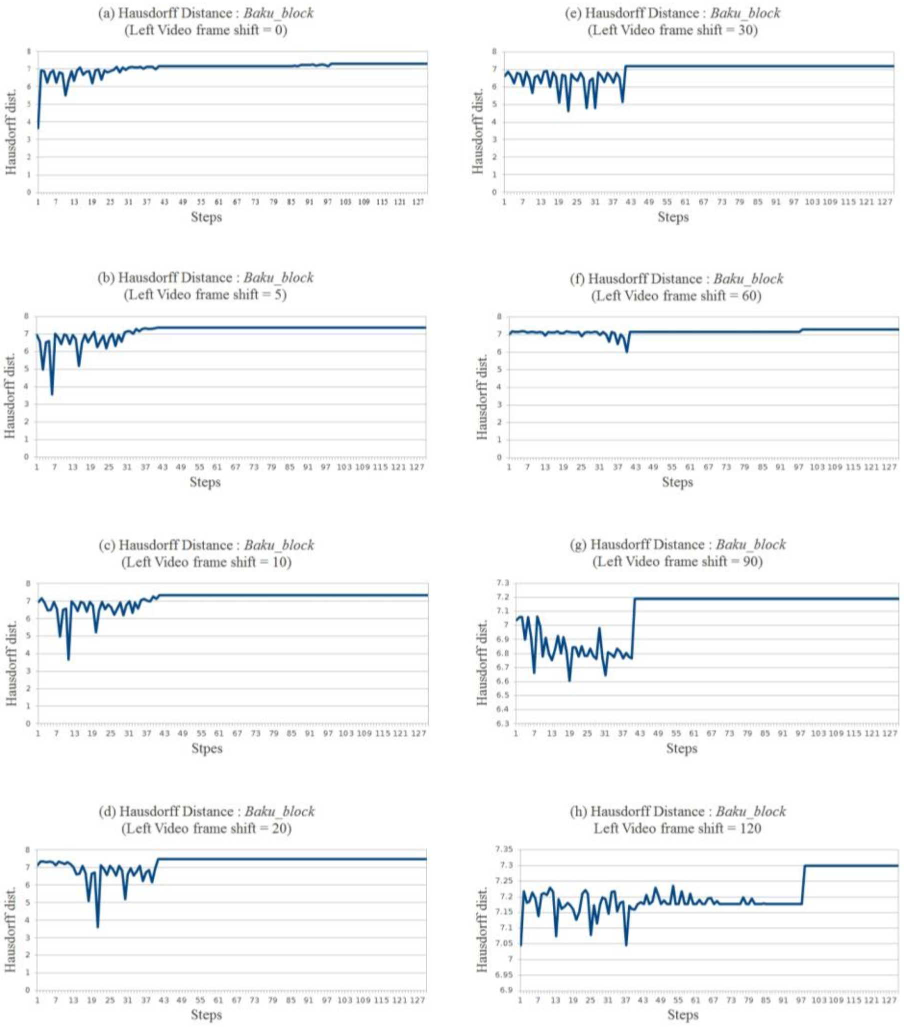

Figure 1: The seeking synchronized point results of *Baku_block* sequence (a) non frame shift as well as synchronized condition, left video set shift to (b) 5, (c) 10, (d) 20, (e) 30, (f) 60, (g) 90 and (h) 120 frames. The graphs shows the most possible synchronized point at lowest peak and were good matching for shift less than 30 frames.

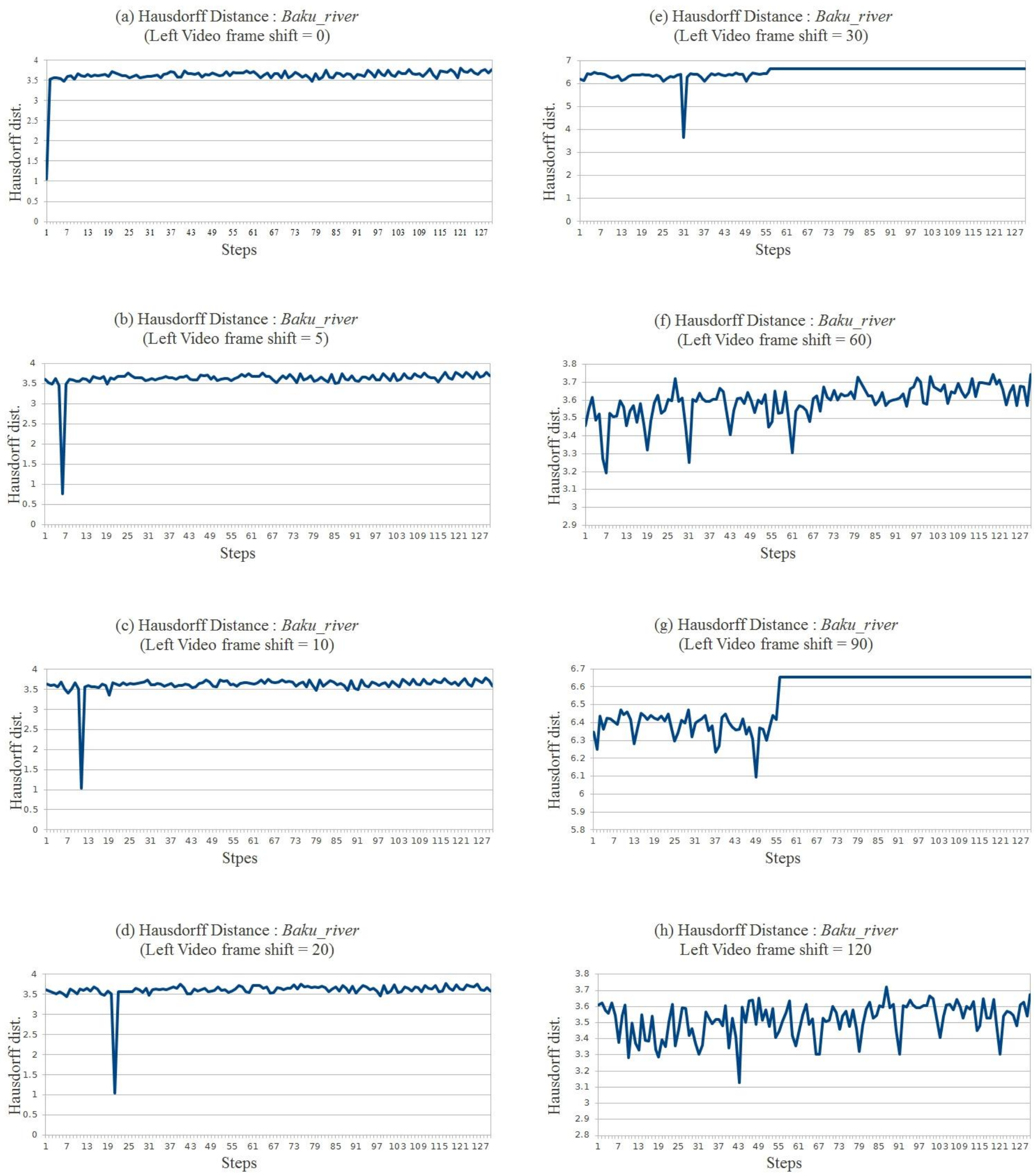

Figure 2: The seeking synchronized point results of *Baku_river* sequence (a) non frame shift as well as synchronized condition, left video set shift to (b) 5, (c) 10, (d) 20, (e) 30, (f) 60, (g) 90 and (h) 120 frames. The graphs shows the most possible synchronized point at lowest peak and were good matching for shift less than 60 frames.

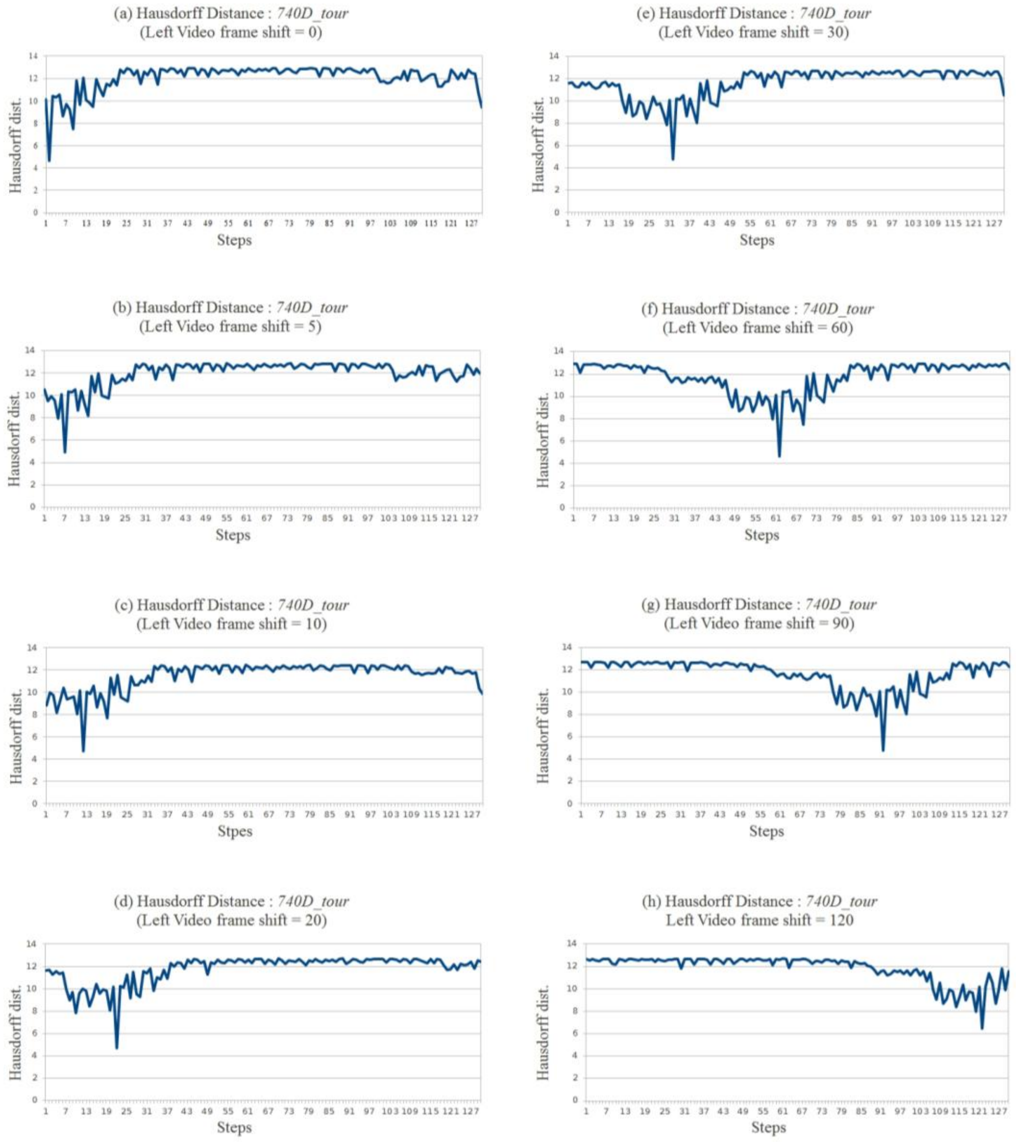

Figure 3: The seeking synchronized point results of *740D_tour* sequence (a) non frame shift as well as synchronized condition, left video set shift to (b) 5, (c) 10, (d) 20, (e) 30, (f) 60, (g) 90 and (h) 120 frames. The graphs shows the most possible synchronized point at lowest peak and were good matching in each condition of frame shift.

Acknowledgements

The authors would like to acknowledge the National Center for High-performance Computing in providing huge computing resources to facilitate this research. This research has been funded in part by National Science Council of Taiwan, Grant Number: NSC 100-2923-I-492-001.

References

[1] T. Tuytelaars, L. Van Gool, Synchronizing video sequences. Proc Cvpr Ieee. (2004), 762-768.

[2] S.N. Sinha, M. Pollefeys, Synchronization and calibration of camera networks from silhouettes. Int C Patt Recog. (2004), 116-119.

[3] L. Sigal, A.O. Balan, M.J. Black, HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion. Int J Comput Vision. Vol.87, (2010), 4-27.

[4] A. Akbarzadeh, J.M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, H. Towles, D. Nister, M. Pollefeys, Towards urban 3D reconstruction from video. Third International Symposium on 3D Data Processing, Visualization, and Transmission, Proceedings. (2007), 1-8.

[5] I.K. Sethi, N. Patel, A Statistical Approach to Scene Change Detection. P Soc Photo-Opt Ins. Vol.2410, (1995), 329-338.

[6] E. El-Qawasmeh, Scene Change Detection Schemes for Video Indexing in Uncompressed Domain. Informatica.

Vol.14, (2003), 19-36.

[7] G. Lupatini, C. Saraceno, R. Leonardi, Scl-Dea, Scene break detection: a comparison. Eighth International Workshop on Research Issues in Data Engineering - Continuous-Media Databases and Applications, Proceedings. (1998), 34-41.

[8] Y. Xilin, I.C. Octavia, Line-Based Recognition Using A Multidimensional Hausdorff Distance. IEEE transactions on pattern analysis and machine intelligence. Vol.21, (1999), 901-916.

[9] W.H. Press, Numerical recipes in C : the art of scientific computing, 2nd ed., Cambridge University Press, Cambridge ; New York, 1992.

Shi-Wei Lo received the Master degree in Biomedical Engineering from Chung Yuan Christian University, Taiwan, in 2002. Now study the PhD of Biomedical Engineering and Environmental Sciences in National Tsing Hua University, Taiwan. He is currently an Assistant Researcher in National Center for High Performance Computing, Taiwan. His research interests include image/vid eo processing and parallel computing.

Fang-Pang Lin received the PhD degree in Civil Engineering from University of Swansea Wales, in 1995. He is currently an Associate Research Scientist in National Center for High Performance Computing, Taiwan. His research interests include Grid Computing, Multigrid Methods, Optimization Design and Computational Fluid Dynamics.