2010 International Conference on Computational and Information Sciences

Camera Motion Detection for Conversation Scenes in Movies

Liu Jiayin*, Wang Chuang*, and Jae Ho Kim
Department of Electronics Engineering
Pusan National University
Pusan, Korea
liujiayin@pusan.ac.kr

Abstract—**In this paper, camera movements including zoom, pan and tilt could be efficiently detected for conversation scenes in movies using an improved feature point matching method. RANSAC algorithm is applied to reduce the error rate of feature matching. This implementation possesses notable significance for conversation scenes' editing and analysis. Besides, it is an essential fundamental step for constructing a visual attention model.**

Keywords-feature matching; RANSAC; camera motion; conversation scenes; movie analysis

I. INTRODUCTION

Camera motions are widely utilized to direct viewer attentions in movies. Zooming-in and tilting is always used to emphasize something. Panning and zooming-out is applied to neglect something. The faster the panning is, the less important the contents are.

In visual attention model, camera motion plays a role of magnifier, which is multiplied with the sum of other visual attention models such as spatial attention model, facial attention model[1]. Therefore, an accurate and efficient camera motion detection system is essential for an attention model. Our research is based on conversation scenes. Obviously, conversation scene is one of the most essential elements in a movie. A good conversation scene could be easily become one of the high spots. Besides, conversation scenes are usually full of conflicts and always chosen as key frames to analyze a movie. Furthermore, in conversation scenes, individual objects movement is not in prominent position. In contrary, camera movement causes the dominant motion in conversation scenes. To analyze the camera motions in conversation scenes is helpful to set up a visual attention model and editing conversation scenes.

Camera motion detection has been studied in several literatures [2][3]. Most of them are achieved by analyzing motion vectors of macro blocks. However, Motion vectors are very computational consuming as every macro block needs to be calculated. Besides macro block motion vectors are too sensitive. In this paper, we propose a camera detection method based on features. RANSAC algorithm is applied to get rid of outliers.

This work was supported by the 2th-phase Brain Korea (BK) 21 program funded by the Ministry of Education, Korea.

*The first two authors have contributed equally in the study and to be considered as co-first authors

II. PROPOSED METHOD

A. Sampling

In this paper, a database of conversation scenes in movies is used for experiments. To analyze the camera motions in those scenes, it is not necessary to calculate frame by frame. Actually, in our experiment, sample images are extracted every three frames. It shows that those sample images possess enough information for us to determine the camera motion types and also the amount of calculation is reduced by sampling.

B. Feature points detection

Analyzing those sampled images pixel by pixel is time consuming and infeasible. Consequently, we extracted enough feature points from each sampled image to represent the global characteristics. Feature extraction could be achieved by several methods such as SIFT algorithm [4] and SURF algorithm [5]. In this paper, Harris corner detection algorithm [6] is applied to extract feature points in this paper because it is a popular interest point detector due to its strong invariance to: rotation, scale, illumination variation and image noise. Besides, Harris corner detection can provide us enough information to determine the type of camera motion. Harris corner feature extraction results for two continuous sampled images are shown in Figure1.

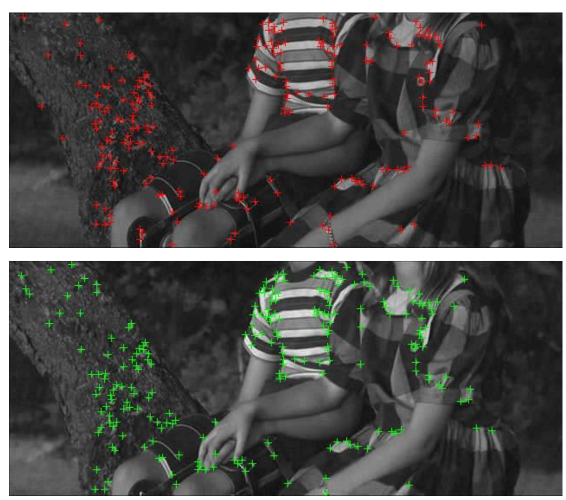


Figure 1: Harris corner detection results for two continuous sampled images

978-0-7695-4270-6/10 $26.00 © 2010 IEEE
DOI 10.1109/ICCIS.2010.180

C. Feature points matching improved with RANSAC algorithm

After the feature points are extracted from two consecutive sample frames, all the feature descriptors in the later frame could be matched with those in the former one. Our algorithm generates putative matches between previously detected feature points in two images by looking for points that are maximally correlated with each other within windows surrounding each point. Only points that correlate most strongly with each other in both directions are returned. The matching result is shown in Figure2.

Figure 2 Feature matching result

As we have noticed, there are some error matching pairs in this process. To reduce the error rate, RANSAC algorithm is introduced. In the computer vision literature, RANSAC is one of the most effective algorithms for model fitting to data containing a significant percentage of gross errors. It is an iterative method to estimate parameters of a mathematical model from a set of observed data. RANSAC Algorithm was proposed by Fishler and Bolles [7] and it is an abbreviation for "RANdom SAmple Consensus". It is applied to find "outliers" which are data that do not fit the model in the data set. It is an iterative method to estimate parameters of a mathematical model from a set of observed data and has been used effectively in a lot of works [8] [9].

In this paper, applying with RANSAC algorithm, we can get the improved matching result from every two continuous sampled frames by cutting of some error matching. The vectors started from the feature points in previous sample frame to their corresponding positions in the latter sample frame are defined as Feature Motion Vectors (FMVs) denoted with $\bar{v}$ for short. Moreover, the start and end point of FMV are denoted with v_0 and v_1 , respectively. Figure3 shows the FMVs generated from the two continuous sample frames in Figure1.

Figure 3 Improved matching results applied with RANSAC algorithm

D. Determination of camera motion types

After the FMVs are extracted, a process to determine the camera motion type is proposed.

First of all, we separate the FMVs into four groups according to their starting points' coordinates (Figure 4):

Group 1: $G1 = \{\bar{v} \mid v_{0x} <= H/2, v_{0y} <= W/2\}$

Group 2: $G2 = \{\bar{v} \mid v_{0x} <= H/2, v_{0y} > W/2\}$

Group 3: $G3 = \{\bar{v} \mid v_{0x} >= H/2, v_{0y} <= W/2\}$

Group4: $G4 = \{\bar{v} \mid v_{0x} >= H/2, v_{0y} > W/2\}$

Where H, W, v_{0x}, v_{0y} stand for frame height, frame width, FMV's starting point's coordinate in x direction , FMV's starting point's coordinate in y direction, respectively.

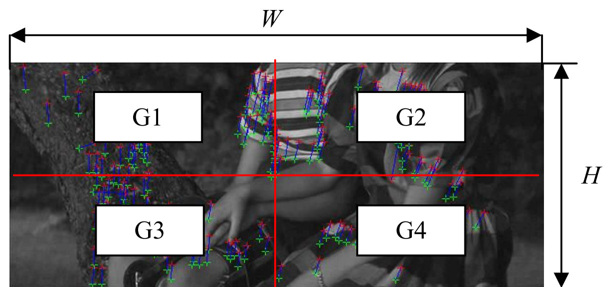


Figure 4: Sketch map of FMVs grouping

Obviously, the orientation angles of each group for different camera motions can be approximately expressed by figure 5

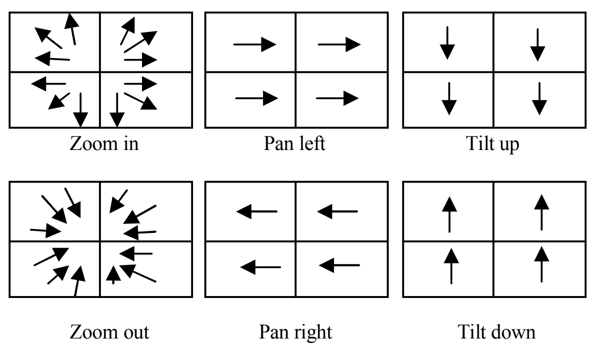


Figure 5 Sketch maps of orientation angles for FMVs for different camera motions.

Undeniably, under the real situation, FMVs are not so organized as in Figure5. There must be a permissible range to determine each camera motion. Accordingly, the following steps are implemented.

First of all, for each group, the orientation angles of FMVs are calculated. Then, we partition all of the orientation angles of FMVs in one group into eight regions showed in Figure6. Camera motion type could be determined through analyzing all the angles. For example, if all the angles are in area 1 and 8, we can consider this camera motion as pan left.

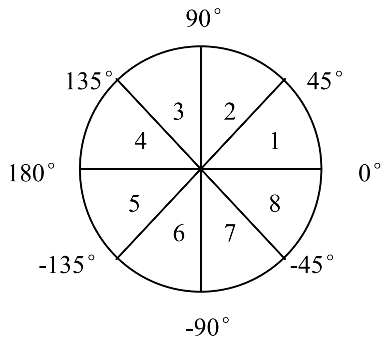
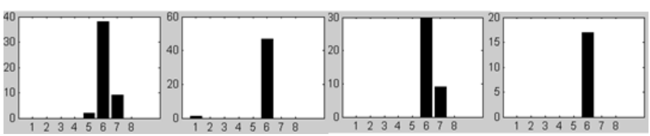


Figure 6 Orientation angles' partition

Consequently, camera motion could be determined by the distribution situation of angles in these eight areas. Considering the real situation, the determination method should be designed with inclusiveness. For example, if all the angles are in area 1 or 8, then this camera motion could be determined as pan left.

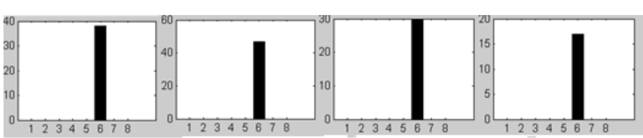
Count the number of FMVs in one group appearing in each partition to form a histogram. Therefore, for FMVs group G_1 , G_2 , G_3 and G_4 , four histograms $H_1[a]$, $H_2[a]$, $H_3[a]$ and $H_4[a]$ are generated as in theFigure7, where a stands for the eight angle areas from 1 to 8.



(a) $H_1[a]$ (b) $H_2[a]$ (c) $H_3[a]$ (d) $H_4[a]$

Figure 7 Histograms generated for the four FMVs groups.

To prune off the noises, data is processed with some statistics methods. The processed histograms are shown in the figure below.

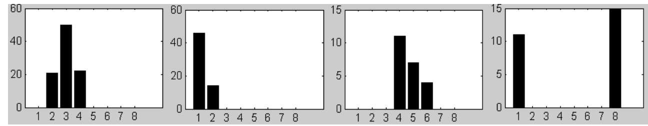


(a) $H_1[a]$ (b) $H_2[a]$ (c) $H_3[a]$ (d) $H_4[a]$

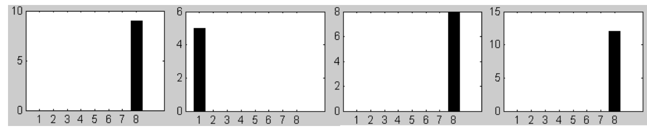
Figure 8 Histograms after trimming

From figure 8, as all of the four histograms belong to the 6th area, we can tell that the camera motion is tilt up.

Other examples of improved histograms are shown in Figure9.



(a) Histograms for a zoom motion



(b)Histograms for a pan motion

Figure 9 Examples of FMV histograms

Consequently, the camera motion types can be determined according to the following table:

TABLE I. DETERMINATION RULES

Conditions of satisfaction	**Camera motion types**
3 or more of the four histograms meet the condition: $H[2] > 0$ *or* $H[3] > 0$	Tilt down
3 or more of the four histograms meet the condition: $H[6] > 0$ *or* $H[7] > 0$	Tilt up
3 or more of the four histograms meet the condition: $H[4] > 0$ *or* $H[5] > 0$	Pan right
3 or more of the four histograms meet the condition: $H[1] > 0$ *or* $H[8] > 0$	Pan left
$H_1[7] > 0$ *or* $H_1[8] > 0$ and ($H_2[5] > 0$ *or* $H_2[6] > 0$) and ($H_3[3] > 0$ *or* $H_3[4] > 0$) and ($H_4[1] > 0$ *or* $H_4[2] > 0$)	Zoom in
$H_1[3] > 0$ *or* $H_1[4] > 0$ and ($H_2[1] > 0$ *or* $H_2[2] > 0$) and ($H_3[5] > 0$ *or* $H_3[6] > 0$) and ($H_4[7] > 0$ *or* $H_4[8] > 0$)	Zoom out

III. EXPERIMENTAL RESULTS

According to the process above, a conversation scene can be analyzed.

To demonstrate the effectiveness of proposed camera motion detection method in this paper, we select 32 famous conversation scenes from four movies for experiments. The results are shown in the table below.

TABLE II. EXPERIMENTAL RESULTS

Movie title	Movie 1 (8 scenes)	Movie 2 (8 scenes)	Movie 3 (8 scenes)	Movie 4 (8 scenes)
Zoom in detected/ ground	4/4	4/4	4/3	2/3
Zoom out detected/ ground	2/2	1/1	2/2	1/1
Tilt up detected / ground	2/2	2/2	2/3	0/0
Tilt down detected / ground	1/1	0/0	0/0	2/2
Pan left detected /ground	2/2	1/1	2/2	3/3
Pan right detected / ground	3/3	3/3	1/2	4/4
Evaluation for proposed method (32 scenes)				
Precision	98.0%			
Recall	96.3%			

In this table, Movie1, Movie 2, Movie 3 and Movie 4 stands for "Gone with the wind", "Forest Gump", "Shawshank's redemption" and "Crouching Tiger Hidden Dragon", respectively. Proposed camera motion detection system is compared with a template matching based method [10] and a threshold based method [11] in the table below.

TABLE III. RESULTS COMPARISIONS

Method	Average Precision
Template matching based camera motion detection	87.4%
Threshold based camera motion detction	96.4%
Proposed feature based method	98.0%

IV. CONCLUSION

In this paper, a camera motion detection method is proposed for conversation scenes in movies. As conversation scene is one of the most essential elements in a movie, this work is of significance in movie analysis and editing. Moreover, camera motion could be used to construct a visual attention model. Unlike other works, we apply a more feature matching method improved with RANSAC algorithm, to detect camera motions. The results show that proposed method is able to detect camera motion types with a high accuracy rate.

REFERENCES

[1] Yu-Fei Ma, Xian-Sheng Hua, Lie Lu and Hong-Jiang Zhang, " A generic framework of user attention model and its application in video summarization," IEEE Transactions on Multimedia, vol. 7, no. 5, pp. 907-919, October 2005.

[2] Adriana Dumitras¸ and Barry G. Haskell,"A look-ahead method for pan and zoom detection in sequences using block based motion vectors in polar coordinates", IEEE, 2004.

[3] Luca Superiori, Markus Rupp, "Detection of pan and zoom in soccer sequences based on H.264/AVC motion information," wiamis, pp.41-44, 2009 10th Workshop on Image Analysis for Multimedia Interactive Services, 2009

[4] Lowe, David G. "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision.2.pp.1150–1157, 1999.

[5] Hyunsup Yoon , Hwan-Ik Chung ,Hernsoo Hahn , "SURF algorithm with color and global characteristics", ICCAS-SICE ,pp. 183 - 187,2009

[6] Zhiyong Ye , Yijian Pei , Jihong Shi, "An Improved Algorithm for Harris Corner Detection", CISP '09. 2nd International Congress on, pp1-4,2009.

[7] Martin A. Fischler and Robert C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". Comm. of the ACM 24, pp.381–395, 1981.

[8] Sunglok Choi;,Taemin Kim;,Wonpil Yu, "Robust video stabilization to outlier motion using adaptive RANSAC", Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on,pp. 1897 – 1902,2009

[9] Kexiang Liu;,Jie Qian, Ruike Yang, "Block Matching Algorithm based on RANSAC Algorithm", Image Analysis and Signal Processing (IASP), 2010 International Conference on, pp. 223 – 227,2010.

[10] Dong-Jun Lan; Yu-Fei Ma ; Hong-Jiang Zhang, "A Systemic Framework Of Camera Motion Analysis For Home Video", IEEE Image processing,2003, pp. 282-92 vol.1,2003

[11] Naito, M.; Matsumoto, k.; Sugaya F.; "Camera Motion Detection Using Video Mosaicing", Multimedia and Expo, 2006 IEEE International Conference on 1741-1744, 2006