Various Camera Motion Type Estimation of Animation Sequences

Hailong Jiang, Guoxu Liu, Jae Ho Kim
Pusan National University, Department of Electronics Eng.
E-mail: jhkim@pusan.ac.kr

Abstract

Camera Motion Estimation (CME) is an important tool for Visual Story Telling (VST) of animations. However, up to now, most of the CME methods have focused on the 6 types of camera motion (CM), and they are far from being satisfactory for VST analysis of animations. In this paper, a new scheme of CME adopting the minimum sum of squared difference of the block motion vector angles is proposed.

1. Introduction

Camera Motion Estimation (CME) is an important tool in many areas such as contents production analysis, video indexing, structure from motion, and object tracking. Besides, CME is also an indispensable component in video coding.

In recent twenty years, there have been many studies on CME. Camera motions can be derived from feature points between consecutive frames. As a conventional method for camera motion detection, detection of motion vectors using SIFT [1] (Scale-Invariant Feature Transform) is widely used. Another feature based method - MSD (Multi-Scale Descriptor) was presented by Changhun Sung [2]. Besides, CME based on a parametric model was also used in many papers [3].

In this paper, a new approach is presented working effectively for estimating 18 CM types and eliminating the problem of rapidly moving character (object).

An experiment is conducted to verify the efficiency of the proposed method. Video data used for the experiment is extracted from the animations 'Frozen' (2013) and 'Big Hero' (2014), respectively. S. G. Ma's method[2] is used for the comparative study because unwanted motion vectors due to the moving objects (characters) can be removed.

2. Theoretical Background

2.1. Types of Camera Motion

In CCTV, the PTZ camera is essential to capture the necessary information. It has pan, tilt and zoom capabilities as shown in Fig. 1.

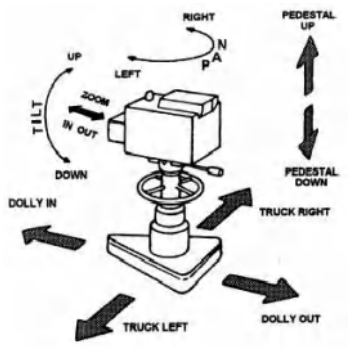

Fig. 1. Representation of pan, tilt, and zoom.

The authors have found that most previous research were carried out for 6 camera motions such as pan right/left, tilt up/down, and zoom in/out [4]. There were few studies on the detailed camera motions important in the VST of recent animations.

2.2. Block Matching Algorithm

The block matching algorithm [5] is a method of finding the most similar blocks between consecutive frames. In this paper, it is adopted to find the motion vectors for camera motion estimation. The N×N size reference block, $r(i,j)$, is in the current frame. BMA is to find the most similar N×N size block in the previous frame. The motion of the block having the

smallest cost function is determined as the motion vector. If the search range is p, the search is performed in the search area of N+2p size as shown in Fig. 2.

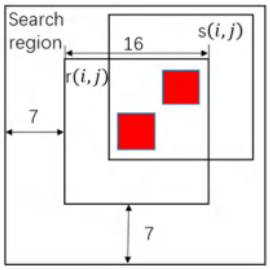

Fig. 2. The search area is shown. The macro block size is 16*16 pixels and the searching parameter, p, is 7.

2.3. Global Motion Filter

Y. F. Ma proposed the global motion filter [6] to extract actual object motion from the mixed motion of object and camera. In this paper, the global motion filter is applied to distinguish background and objects of a video clip.

The principle is stated as follows, shown in Fig. 3. After getting the motion vectors using BMA,

the magnitude and angle of MV are calculated using its x and y components. Then, the entropy of MV's angle can be calculated and used to determine whether the block is corresponding to the camera motion or object motion.

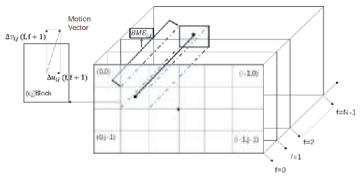

Fig. 3. The $Block(i,\ j)$ and $BME_{i,j}$ in the global motion filtering.

3. The Proposed CME Approach

3.1. The Ideal Motion Vector Angles for Various Camera Motion

Normally, 4 camera motions (Pan Left, Pan Right, Tilt Up, and Tilt Down) are detected. [1,2] In this paper, 12 more directions are added in order to acquire the details of camera motions, as show in Fig. 4.

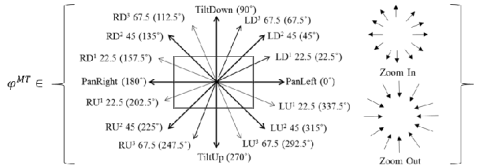

Fig. 4 Types of camera motions to be distinguished in the proposed method.

Each block's motion vector depends on camera motion and the position of the block in the frame. In addition, the motion vectors are different due to the resolution of the image. Thus, ideal motion vectors are calculated considering the above factors. Note that the width and height of a frame are assumed as W and H.

Eq. (1) to Eq. (8) are the ideal MV angles for PanLeft, PanRight, TiltUp, TiltDown, LD^1~LD^3, LU^1~LU^3, RD^1~RD^3, and RU^1~RU^3, respectively.

$$\varphi_{i,j}^{PanLeft}=0^{\circ}, \quad (1)$$

$$\varphi_{i,j}^{PanRight}=0^{\circ}, \quad (2)$$

$$\varphi_{i,j}^{TiltDown}=0^{\circ}, \quad (3)$$

$$\varphi_{i,j}^{TiltUp}=0^{\circ}, \quad (4)$$

$$\varphi_{i,j}^{LD^1}=22.5^{\circ},\ \varphi_{i,j}^{LD^2}=45^{\circ},\ \varphi_{i,j}^{LD^3}=67.5^{\circ}, \quad (5)$$

$$\varphi_{i,j}^{LU^1}=337.5^{\circ},\ \varphi_{i,j}^{LU^2}=315^{\circ},\ \varphi_{i,j}^{LU^3}=292.5^{\circ}, \quad (6)$$

$$\varphi_{i,j}^{RD^1}=157.5^{\circ},\ \varphi_{i,j}^{RD^2}=135^{\circ},\ \varphi_{i,j}^{RD^3}=112.5^{\circ}, \quad (7)$$

$$\varphi_{i,j}^{RU^1}=202.5^{\circ},\ \varphi_{i,j}^{RU^2}=225^{\circ},\ \varphi_{i,j}^{RU^3}=247.5^{\circ}, \quad (8)$$

For the case of zoom-in and zoom-out, the motion vectors are different according to the locations of blocks. Now, let's assume that the zoom center is the same as the screen center (W/2, H/2).

Let's define a and b as functions of variable i and j as follows:

$$a(i,j)=16*i+8-W/2, \quad (9)$$

$$b(i,j)=-16*j-8+H/2. \quad (10)$$

where $(a(i,j), b(i,j))$ are the center of each block. Then the ideal angles of ZoomIn and ZoomOut for each block can be described as follows:

$$\varphi_{i,j}^{ZoomIn} = \begin{cases} atan2(a(i,j), b(i,j)) * 180/pi, & if\ b(i,j) \geq 0 \\ atan2(a(i,j), b(i,j)) * 180/pi + 360°, & if\ b(i,j) < 0 \end{cases} \quad (11)$$

$$\varphi_{i,j}^{ZoomOut} = \begin{cases} atan2(a(i,j), b(i,j)) * 180/pi + 180°, & if\ b(i,j) \geq 0 \\ atan2(a(i,j), b(i,j)) * 180/pi + 180°, & if\ b(i,j) < 0 \end{cases} \quad (12)$$

In the proposed method, the Minimum Sum of Squared Difference (MSSD) criterion is used to decide the camera motion type.

Note that the fast moving object (usually animation characters) has already been removed by the global motion filter. Then, the angular difference between the angle of the estimated MV, $\theta_{i,j}(f,f+1)$, and the ideal vector angles, $\varphi_{i,j}^{MT}$, of 18 camera motion types can be denoted as follows:

$$Diff_{i,j}^{MT}(f,f+1) = \theta_{i,j}(f,f+1) - \varphi_{i,j}^{MT},\ Diff_{i,j}^{MT}(f,f+1) \in [0, 360], \quad (13)$$

Note again that the ideal angles of zoom-in and -out is block position (i, j) dependent.

The error energy, $e^{MT}(f,f+1)$, corresponding to all camera motion types are obtained as Eq. (14).

$$e_{i,j}^{MT}(f,f+1) = Diff_{i,j}^{MT}(f,f+1)^2, \quad (14)$$

The sum of $e_{i,j}^{MT}$, E^{MT}, is as follows.

, (15)

where I×J and F represents the number of macro blocks in a frame and the number of frames in a video clip.

The final decision of Camera Motion type is as follows:

$$Decision = (mt \mid mt \in MT, E^{mt} = the\ minmum\ of\ \{E^{MT}\}) \quad (16)$$

where MT is the total camera motion type set and mt is an index of MT.

4. Experimental Results

4.1 Experimental Data

Movie animations are used as experimental data to test the proposed method. Successful animations released during the year 2013 - 2016, are selected. In this paper, 'Frozen' (2013) and 'Big Hero' (2014) which got the 71th and 72th Golden Globe Awards, respectively, were selected.

In the experiment, all shots having constant speed and direction for 1 second or more were selected for the experiment. The mixed shots (including 2 or more CMs) were excluded. The number of shots selected is 138 and 140 for Frozen and Big Hero, respectively.

4.2. Camera Motion Prediction with The Proposed Algorithm

Based on the experimental data in Sec. 4.1, the proposed method is applied to 138 clips extracted from 'Frozen' and 140 data of 'Big Hero', and one examples of the result is presented below.

Figure 5 shows the experiment data extracted from 'Frozen'. In this case, the motion type of PanLeft is correctly found. As time passes, the characters move towards to the camera, while the background moves to the right. Table 1 shows E^{MT} and the value corresponding to Pan Left is the minimum.

Fig. 5. An example of good result in the experiment (PanLeft).

Table 1. Calculated E^{MT} from the video of Fig. 5

	Pan-right	Pan-Left	Tilt-Up	Tilt-Down	RU	LU	RD	LD	RU
E^{MT}	1.35E+05	6.90E+03	7.27E+04	6.66E+04	1.25E+05	2.18E+04	1.17E+05	1.91E+04	1.07E+05
	LU	RD	LD	RU	LU	RD	LD	Zoom-In	Zoom-Out
E^{MT}	3.86E+04	1.00E+05	3.44E+04	9.00E+04	5.56E+04	8.33E+04	5.03E+04	7.57E+04	6.50E+04

4.4. Statistical Evaluation

In the analysis of the experimental results, Precision, Recall and Accuracy are used to verify effectiveness of the newly proposed method in this paper. They can be calculated with definitions (TP, TN, FP, FN) in Table 2.

Table 2. Statistical analysis of the predicted results and the actual ones.

		True Class	
		Positive	*Negative*
Predicted Class	*Positive*	True Positive *Count(TP)*	False Positive *Count(FP)*
	Negative	False Negative *Count(FN)*	True Negative *Count(TN)*

To evaluate the effectiveness of this paper, the proposed method is compared with S.G. Ma's method [2].

The proposed shows 91.56% precision, 92.59% recall, and 99.03% accuracy for the average of 16 CMs. These are 8.36%, 8.89%, and 6.63% higher in precision, recall, and accuracy compared to Ma's method. The proposed method shows 92.05% precision, 86.1% recall, and 97.3% accuracy for zoom. These are 5.15%, 7.7%, and 0.6% higher in zoom compared to Ma's method. The results are shown in table 3 as below.

Table 3. Comparison with S. G. Ma's method.

	Precision		Recall		Accuracy	
Method / *CMT*	*The Proposed*	*Ma's Method*	*The Proposed*	*Ma's Method*	*The Proposed*	*Ma's Method*
Average of 16 CMs	91.56%	-	92.59%	-	99.03%	-
Average of Pan and Tilt	-	83.2%	-	83.7%	-	92.4%
Zoom	92.05%	86.9%	86.1%	78.4%	97.3%	96.7%

5. Conclusions

Since CM has become more and more important in 3-D animation production, and has great effect in Visual Story Telling (VST) as well as other areas. A new method is proposed in this paper to estimate 18 types of camera motion (Pan, Tilt, Zoom, and various angles). The aims of this paper are: 1) Detecting 18 types of CM for VST analysis of the animation and the movie and 2) moving character removal for correctly detecting the CMs.

Even though CM types are increased from 6 to 18, the proposed method is superior in all aspects of precision, recall, and accuracy compared with S. G. Ma's method. The experimental results verify the effectiveness of the proposed method. This can be used for video indexing and VST analysis for the production of movie and animation.

References

[1] S.G. Ma & W.Q. Wang, "Effective camera motion analysis approach," in IEEE International Conf. on Networking, Sensing and Control (ICNSC'10), pp. 111-116 (2010).

[2] C. Sung, & M. J. Chung, "Multi-scale descriptor for robust and fast camera motion estimation," IEEE Signal Process. Lett., 20(7), 725-728, (2003).

[3] P. Chang, M. Han & Y. Gong, "Extract highlights from baseball game video with hidden markov models," in Proc. IEEE Int. Conf. on Image Process. (ICIP'02), pp. 609-612 (2002).

[4] N. Nguyen & D. Laurendeau, "A robust method for camera motion estimation in movies based on optical flow," Int. J. Intelligent Systems Technologies and Applications, 9(3-4), 228-238 (2010).

[5] A. Barjatya, "Block matching algorithms for motion estimation," IEEE Trans. Evol. Comput. 8(3), 225-239 (2004).

[6] Y. F. Ma & H. J. Zhang, "A new perceived motion based shot content representation," in Proc. IEEE Int. Conf. on Image Process. (ICIP), 3, pp. 426-429 (2001).