A Study on a Motion Recognition from Moving Images with Camera Works

Shin-ichi Murakami and Tomohiko Shindoh

Graduate School of Engineering, Tokyo Denki Umiversity
2-2 Nishiki-cho Kanda Chiyoda-ku Tokyo 101 Japan
TEL&FAX:+81-3-5280-3334, e-mail:murakami@c.dendai.ac.jp

Abstract

This paper describes an automatic recognition method of contents in moving images. The recognition process is carried out by the following two steps. At first, camera works in moving images are analyzed and moving objects are extracted from the moving images. Next, the motion of the object is recognized by pre-procured knowledge.

These techniques will be applied to a construction of an efficient image database.

Keyword : Moving Image , Camera Work , Object Extraction , Motion Recognition

1. Introduction

Along with the progress of processing speeds and memory capacities of computers, moving images have become handled with computers. In fact, several companies are trying to construct some image databases which can deal with moving images such as motion films, TV-programs and so on by fully digital forms[1]~[2].

However, it is difficult for users to get their objective images efficiently from moving image databases because moving images are composed of a large number of time sequential image frames. In order to overcome this difficulty, it can be considered to index the moving images based on their contents. But an indexing of image frames by hands is a laborious task especially for an image database with a large number of image frames. Then, the development of an automatic recognition technique of image contents is required and it will give us a powerful tool for a construction of an efficient information retrieval system of moving images.

On the other hand, ordinary moving images often contain camera works such as panning, zooming, tilting and so on. These camera effects also give us another difficulty for the recognition of contents of moving images, because the back ground scenes of the moving images with camera works are also in motion.

In order to extract an object motion from moving images, several methods such as optical flow method etc. have been studied [3]. However, optical flow method requires a lot of computation time and large memory area for computers.

This paper describes a relatively simple camera work analysis method for moving images and how to extract an object from them and to recognize the object motions.

2. Preprocessing

2.1 Assumptions

Ordinary moving images are composed of several shots each of which is usually picked up by a single camera in a consecutive time sequence. And a connecting point between consecutive two shots is called as a cut point. One shot usually contains at most one or two camera works. Then, we partition the moving images into some sets of shots in order to recognize the camera works of the moving images. Namely, we firstly find the cut points of moving images. And next, we apply the recognition process of contents to each of shots partitioned by the cut points.

Here, we assume that the moving images considered in this paper satisfy the following assumptions.

Assumptions:

(1) Every frame in one shot contains objects with clear edges.

(2) Every shot contains at most one camera work.

2.2 Detection of Cut Point

Since a shot is picked up by a single camera for the time sequence, image contents in one shot do not change quickly. Therefore, in order to find a cut point, we extract a frame with quick changes of colors in moving images. In this case, we divide an image frame into several number of blocks shown in Figure 1.

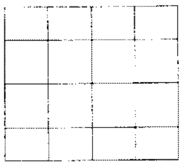


Figure 1 Division of an Image Frame

Then, the color differences between the consecutive two frames are observed by block by block so that we can find a cut point yielded by camera changes in a same scene. Namely the color difference degree is defined by the following equation.

$$\chi^2 = \frac{1}{n}\sum_{l}^{n}\sum_{j}^{m}\frac{\left(c_{i+1,j} - c_{i,j}\right)^2}{c_{i,j}} \qquad (1)$$

where n is the number of blocks

m is the number of pixels in a block,

$c_{i,j}$ is the color value of pixel j in frame i.

Using the color difference degree (1), the cut point can be detected by the following procedure.

Procedure 1:

(1) Each frame is divided into blocks shown in Figure 1.

(2) The color difference degree (1) for every consecutive two frames is calculated.

(3) If the color difference degree is greater than a threshold, it is inferred that a cut point exists between the consecutive two frames.

3. Analysis of Camera Works

3.1 Kinds of Camera Works

Prior to the recognition of objects in moving images, camera works are extracted from the moving images. Here, we consider to detect the following 6 kinds of camera works shown in table 1.

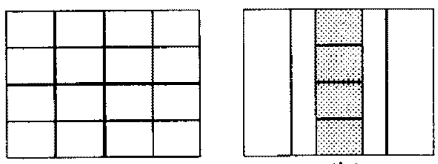
Table 1 Kinds of Camera Works

Camera works	Object motion in the image
Panning	Right or left shifting
Tilting	Up or down shifting
Zooming	Closing up or going far
Rotation	Right or Left circulating
Panning & Zooming	Right or left shifting & closing up or down shifting
Pursuing panning of objects	The object is set in the center and the background is shifted to right or left.

3.2 Detection of Motion Vector

As mentioned before, the shapes and positions of objects in one shot do not change quickly. Then, the motion vector in an image frame can be detected by checking the correlation of the shapes and positions of objects in the consecutive two frames.

In moving images with camera works, every part of an image frame shows a different direction of motion depending on the camera work. Namely, for a panning action, all parts of an image frame shift to the right or to the left and for a zooming up action, the right part shifts to the right and the left part shift to the left. Then, if we analyze the motions of the divided parts in an image frame, we can easily find the camera work of the image. Based on this idea, we divide an image frame into several blocks as shown in Figure 2. Here, the finding process of the motion direction for each block is effectively done if the image frame contains clear edges. Then, every image frame is transformed to an edge image prior to the finding process of motion directions of the object.



(a) (b)

Figure 2 Division of an Image Frame

The procedure for finding the motion vector in image frames is shown bellow.

Procedure2:

(1) Noise components in every frame are reduced by a noise filtering technique.

(2) Every image frame is transformed to an edge image.

(3) Every edge image is divided into blocks shown in Figure 2 (a) and (b). Here, the blocks in Figure 2 (a) are for the detection of background motions and the blocks in Figure 2 (b) are for the detection of object motions.

(4) Each block is matched to the neighbored blocks in the next consecutive frames by shifting it pixel by pixel. The difference degree of the matching is calculated by a standard deviation of density differences between the two blocks shown in the following equation.

$$T=\sum_{j}^{m}\sum_{k}^{m}(c_{i+1,j}-c_{i,k})^2 \qquad (2)$$

,where m is the number of pixels in a block, and $c_{i,j}$ is the color density of pixel j in frame i.

(5) The motion vector in each block is found as the direction between the two blocks which minimizes the difference degree of the above equation (2).

3.3 Detection of Camera Works

As mentioned before, every camera work has its own motions of every part of an image frame. Then, a camera work is detected by the following procedure using the motion vectors obtained by *Procedure 2*.

Procedure3:

(1) Divide each of the consecutive two frames into blocks shown in Figure 2 (a) and (b).

(2) Find the motion vectors in every block by *Procedure 2*.

(3) Transform the motion vector in every block in Figure 2 to the motion vector in every block in Figure 3. The components of the motion vector in every block of Figure 3 are calculated by averaging the components of the motion vectors in the corresponding blocks in Figure 2.

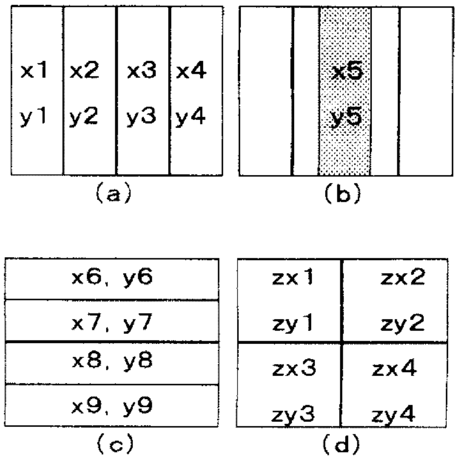


Figure 3 Transformed Block

(4) Panning is detected by checking the following two conditions for all blocks in Figure 3 (a) and (c).

(a) All the y-components are nearly equal to zeros.

(b) All the x-components are constant.

(5) Tilting is detected by checking the following two conditions for all blocks in Figure 3 (a) and (c).

(a) All the x-components are nearly equal to zeros.

(b) All the y-components are constant.

(6) Zooming is detected by checking the following two conditions for all blocks in Figure 3 (d).

(a) Zoom in: $zx_1, zx_3 < 0$, $zx_2, zx_4 > 0$ for x-coordinates and $zy_1, zy_2 > 0$, $zy_3, zy_4 < 0$ for y-coordinates.

(b) Zoom out: $zx_1, zx_3 > 0$, $zx_2, zx_4 < 0$ for x-coordinates and $zy_1, zy_2 < 0$, $zy_3, zy_4 > 0$ for y-coordinates.

(7) Rotation is detected by checking the following two conditions for every block in Figure 3 (a).

(a) Right circulation: $y_1 > y_2 > y_3 > y_4$

(b) Left circulation: $y_1 < y_2 < y_3 < y_4$

(8) Panning and zooming is detected by the following process.

(a) x-components $x_i (i = 1\sim4)$ in Figure 3 (a) are transformed by the following equation.

$$x_i = x_i - \frac{|x_2 + x_3|}{2} \quad (i = 1\sim4) \qquad (3)$$

(b) The transformed image is checked whether it is a zoomed image or not, by the procedures in step (6) mentioned above.

(9) Pursuing panning of object is detected by checking the following three conditions for Figure 3 (a) and (c).

(a) All the y-components are nearly equal to zeros.

(b) x-components x_1, x_4 is constant.

(c) x-component x_5 is nearly equal to 0.

4. Recognition of Object

4.1 Reduction of camera work effects

In order to extract the motion of an object from moving images with camera works, it is necessary to reduce the camera work effects from the original images. This process is carried out by using *Procedure 3* as follows.

Procedure4:

(1) Reduction of panning effects:
All the blocks with panning components are reduced by the amounts of panning components obtained by *Procedure 3*.

(2) Reduction of tilting effects:
All the blocks with tilting components are reduced by the amounts of tilting components obtained by *Procedure 3*.

(3) Reduction of zooming effects:
All the blocks with zooming components are enlarged or shrunk by the amounts of zooming components obtained by *Procedure 3*.

(4) Reduction of rotation effects:
All the blocks with rotation components are rotated inversely by the amounts of rotation components obtained by *Procedure 3*.

4.2 Extraction of Moving Object

Since a moving object can be extracted based on the shift quantity of its position between the consecutive two frames, the object extraction is carried out by the following procedure shown in Figure 4.

Procedure5:

(1) The effects of camera works are reduced by using *Procedure 4*.

(2) The frame difference is calculated from a consecutive two frames (Figure 4 (b)).

(3) The edge image is extracted by differentiating the corrected image obtained in step (1) (Figure 4 (c)).

(4) The logical product is calculated between the two frame difference images extracted by step (2) and the edge image extracted by step (3) (Figure 4 (d)).

(5) The object is extracted as a logical product of consecutive two images determined by step (4) (Figure 4 (e)).

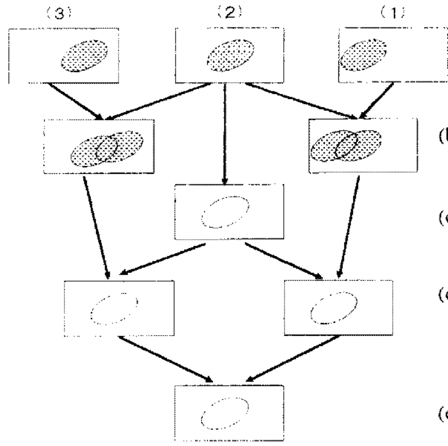


Figure 4 The Method of Extracting Object

5. Analysis of Object Motion

By *Procedure 5*, we can get object images only containing moving objects. Then, the motions of the objects are analyzed through those images.

Here, we assume that the object image contains only one person. And furthermore, we also assume that the general motion of an object is known in advance, for example "a walking motion". In this case, the purpose of the motion analysis is to obtain the shift quantity of the person and the motion pitches of his arms and legs.

The procedure for the motion analysis is shown below.

Procedure 6:

(1) In general, the position of an object in each frame shifts frame by frame under the influence of the camera motion. In order to set the basic position of the object, the fundamental frame is selected. Ordinary, the first frame in the shot is selected as the fundamental frame.

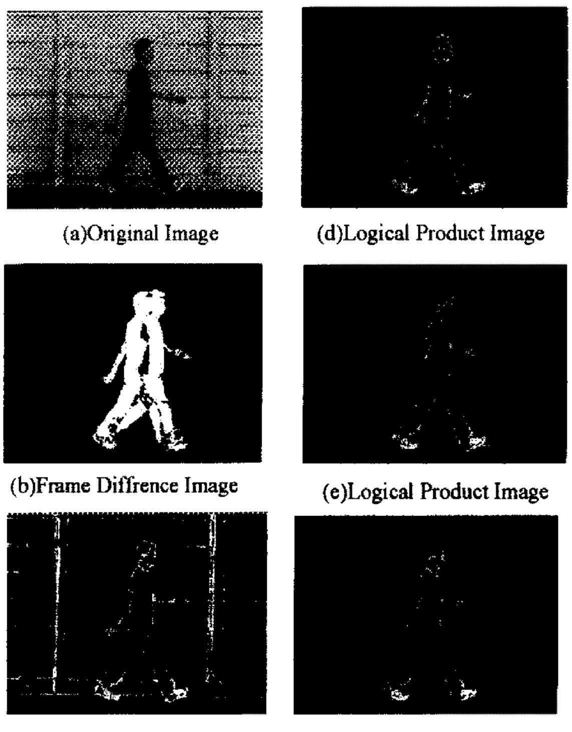
(2) The characteristic points of a human body are extracted. For example, the top

of head, the center of loin, the tips of both hands and legs are selected as the characteristic points.
(3) Since the position of the loin can be considered as the center of the body, the shifting speed of his body is calculated from the position of his loin.
(4) The motion pitches of hands and legs can be calculated from the shift quantities of their positions in each frame.

6. Experiments

In order to ascertain the effectiveness of the proposed algorithm, some experiments of extracting a person and recognizing his walking motions are carried out.

The extracted images of a walking person are shown in Figure 5. Every picture corresponds to every level in Figure 4. We can see the shapes of a walking person is effectively extracted.



(a)Original Image (d)Logical Product Image

(b)Frame Diffrence Image (e)Logical Product Image

(c)Edge Image (f)Object Image

Figure 5 Image given by Extraction of Object

As for the recognition of motions of a walking person, we extracted the tip positions of his arms and legs. The vertical line in Figure 6 shows the central axis of his body and the two sets of horizontal lines show the trajectories of the tip points of his arms and legs, respectively. The central two dots show the positions of his armpit and loin. From Figure 6, we can calculate the motion pitches of his arms and legs.

Figure 6 The Movement of Arms and Legs

Table 2 shows the recognition results for his walking motion.

Table 2 Motion Parameter

Shift of motion	Pitch of arms	Pitch of legs
15	104	73

(Unit : picture elements/sec)

7. Conclusion

We have proposed an object extraction method from moving images with camera works such as panning, zooming, tilting and so on, and also proposed a motion recognition method of the objects. And we have shown the effectiveness of our algorithm through some experiments.

However, in the present experiment, the object motion is restricted to a simple walking motion. Then, we would like to apply this algorithm to more complicated motions such as running, jumping, wrestling and so on, hereafter. But in that case, much more complicated characteristic points, for example, shoulders, elbows, knees, etc. will be required.

Furthermore, we will also want to extend our algorithm to the recognition of emotions by analyzing motion parameters.

References

[1] Akutsu, Tonomura, Hashimoto, Ohba, "Camera Operation Extraction from Video Data", IEICE, vol.J75-D-II, no.2, pp.226-235, 1992

[2] Nagasaka, Tanaka, "Automatic Video Indexing and Fill Video Search for Object Appearances"

IPS. vol.34. no.4, pp543-550. 1992
[3] Ohta, "Structure Motion with Confidence Measure and Its Application for Moving Object Detection" ,IEICE. vol.J76-D-Ⅱ. no.8. pp.1562-1571. 1993
[4] Nagao, Miyatake, Fujita, Ueda, "Moving Object Detection by Time-Correlation-Based Background Judgement Method" ,IEICE. vol.J79-D-Ⅱ, no.4. pp.568-576,1996
[5] Asaoka, Yokoya, Takemura, Yamazawa, "Motion Detection from Image Sequences with a Moving Camera Using Robust Statistics", IEICE. PRMU96-145, pp.33-40, 1997
[6] Yoneyama, Nakajima, Ujihara, "Detection of Moving Objects from MPEG Video Stream" ,IEICE, PRMU96-101, pp.63-70, 1996
[7] Taniguchi, Tonomura, Hamada, "A Method for Detection Shot Changes and Its Application to Access Interfaces toVideo" ,IEICE, vol.J79-D-Ⅱ. no.4, pp.538-546, 1996
[8] Maki, Haseyama, Kitajima, "A Note of Cut Detection for Images Sequences Using Pattern Recognition", IEICE, IE96-129, pp.47-52, 1997
[9] Ueda, Miyatake, Yoshizawa, "A Proposal of an Interactive Motion Picture Editing Scheme Assisted by Media Recognition Technology". IEICE, vol.J75-D-Ⅱ. no.2, pp216-225, 1992
[10] Shindoh, Murakami, "A Study on Automatic Analysis of Picture Expression", ITE, 1996

Remarks: Experimental results will be demonstrated by VTR.