VIDEO HYPERMEDIA AUTHORING FOR VIDEO DELIVERING SERVICE BASED ON A COMBINATION OF AUTOMATIC TRACKING AND MANUAL EDITING

Hironobu Abe[1,2], *Junshiro Kanda*[1], *Koji Wakimoto*[1], *Hiroshi Shigeno*[2] *and Ken-ichi Okada*[2]
[1]:Mitsubishi Electric Corporation, {hironobu, junshiro, waki}@isl.melco.co.jp
[2]:Keio University, {abe, shigeno, okada}@mos.ics.keio.ac.jp

ABSTRACT

This paper proposes an effective anchor setting method for use in video hypermedia systems for broadband video delivering service based on a combination of automatic tracking using motion picture analysis and manual editing. The main advantages are (1) improving tracking accuracy by using bi-directional tracking, (2) effective anchor editing process by deleting intermediate frames when using the proposed anchor setting method. We present experiments using sport videos. As evaluation results of these experiments we and proved the effectiveness of the method.

1. INTRODUCTION

Recently, video delivering services over broadband networks are expanding rapidly. In this paper, we describe an application of our video hypermedia architecture to provide an enhanced video delivering service. For example, by applying this architecture to an on-demand video delivering service consisting of sports video highlights, we can intuitively retrieve a player's profile and records by directly pointing to the specific player on the video image. One important requirement for practical applications of this service is the simplification of the authoring process of video hypermedia after completion of the shooting and editing of video materials.

Various studies on video hypermedia enabling the retrieval of related information by selecting an object from the video, have been conducted actively [1]-[4]. However studies about efficient anchor setting methods during the authoring process of video hypermedia have not been actively pursued. By using current computer performance and video analyzing technologies, it is impossible to track objects automatically as the user would wish. Tracking errors will definitely occur. So it is important to combine automatic tracking with manual editing to increase the reliability of the authoring process of this system. According to Takada's approach, the automatic tracking process may be interrupted repeatedly during the process depending on the kind of target video data, because tracking errors must be corrected manually when a tracking error occurs [5].

Up to now, we have been conducting studies on a framework to handle video data easily on computers by structuring it [6]-[8]. In this paper, we propose an efficient anchor setting method of objects in video. The proposed method consists of tracking objects automatically using motion picture analysis and correcting tracking results manually if necessary. The advantages include (1) improving the accuracy of tracking using bi-directional tracking, (2) effective anchor editing by deleting intermediate frames. We describe experimental results of the anchor setting process used in sport videos to prove its effectiveness.

2. PROPOSAL OF ANCHOR SETTING METHOD

The basic policy of our approach consists of (1) upgrading effectiveness of the automatic tracking function of the object region using motion picture analysis, (2) efficient manual correction of the tracking result, in case the automatic tracking result was against the user's intention.

The objective of this approach is make more efficient the painful process of anchor setting by combining automatic tracking using motion picture analysis and manual editing.

2.1. Anchor model

Figure.1 shows the proposed anchor model in this paper.

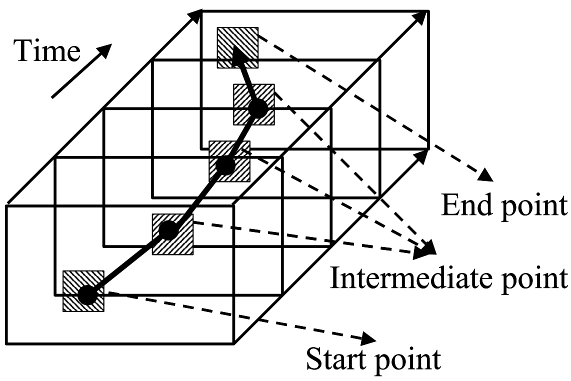

Figure.1 Anchor model

In our model, an anchor assigned to an object in video does not necessarily have anchor data in every frame, it has anchor data in the frames we call start point, end point and intermediate points. Frames which have no anchor data make an estimate based on anteroposterior frames having anchor data using interpolation, as shown in Figure.1.

In this paper, we define the frames which have anchor data such as start frame, end frame and intermediate frame as key frames, and frames which have no anchor information as normal frame. Anchor data of a normal frame can be estimated from anteroposterior key frames using interpolation.

2.2. Anchor setting method

We propose an setting method of the anchor defined in the previous section. Basically, it is a method combining automatic tracking using motion picture analysis and correcting the tracking error manually, as shown below.

(1) Step 1: Setting start point and end point:

User sets the position and size of the anchor at the start frame and the end frame. This manipulation is done by seeking the video in the start frame and the end frame and drawing an oblong anchor on the target object in the video.

(2) Step 2: Automatic object tracking using motion picture analysis:

After Step 1, if the user chooses automatic anchor tracking function, the system analyzes the video automatically and sets intermediate key frames to the object in motion as the result of tracking.

(3) Step 3: Correcting tracking error:

User checks the automatic tracking result and corrects the anchor if necessary. Correcting process consists of inserting intermediate key frames to the appropriate position, deleting wrong key frames, or editing the size or position of the selected anchor.

A flow of the automatic anchor setting process using motion picture analysis is shown below.

(1) Step 1: Forward tracking:

The system assigns the start frame's anchor data set by the user as a first template, and tracks the object by repeating pattern matching process until the end frame every five frames (five is the default value of the tracking frame interval.). According to this process, the system sets intermediate points every five frames.

The process uses normalized correlation shown in Formula (1) and calculates score S in Formula (2). If the score S goes over the threshold, the process figures that the matching result is good.

$$r = \frac{N\sum IM - \sum I \times \sum M}{\sqrt{\left[N\sum I^2 - \left(\sum I\right)^2\right]\left[N\sum M^2 - \left(\sum M\right)^2\right]}} \qquad (1)$$

$$S = \max(r^2, 0) \times 100 \qquad (2)$$

N : number of pixels of the model

I : value of pixels of the model

M : value of pixels of the target

(2) Step.2: Checking the result of tracking:

If the anchor position result in the end frame generated by forward tracking matches the anchor position of the end frame set by the user, go to Step.4, if not go to Step.3.

(3) Step.3: Backward tracking:

The system tracks the object in the reverse direction using the anchor of the end frame set by user as a first template. During this backward tracking process, if it reaches the locus of Step.1, the system stops backward tracking process. However if the backward tracking process does not reach the locus of Step.1, the system stops tracking at the position of 50% of the whole section.

(4) Step.4: Deleting intermediate points:

Step.4 deletes redundant intermediate points for approaching the locus by Step.1-3 to as less intermediate points as possible. In particular, we select a 3-D extended method of dividing linear approximation method using at the linear approximation of line figure.

Features of this setting method are shown below.

(1) Using bi-directional tracking result:

Forward and backward tracking processes are enabled by the user by setting the duration of the anchor, and the start and end frames at the same time. In the case of tracking using pattern matching, the remaining tracking process will be wasted if it misses the object during the tracking process. But this method is able to improve the accuracy of tracking by using the bi-directional tracking results.

(2) Deleting intermediate points to make error tracking correction more efficient:

The system inserts intermediate points every predetermined frame interval by setting the intermediate points based on the locus of the tracking result. In this case, if the object moves linearly, anchor data will be redundant and more effort to correct the tracking error is required. So this method can reduce the correcting effort drastically by deleting redundant intermediate points inserted every several frames.

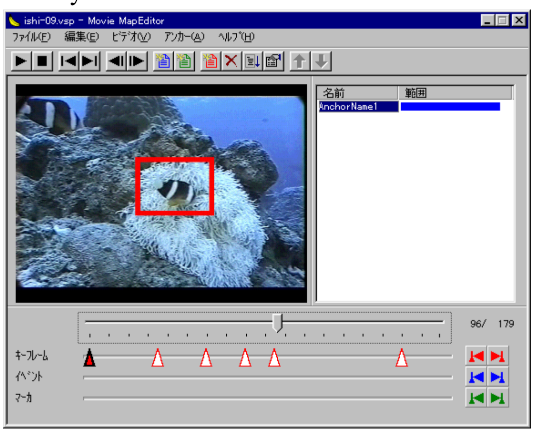

Figure.2 A screen image of an anchor setting software

2.3. Implementation

We implemented an anchor setting software based on proposed method. Figure.2 shows the screen image of this software.

3. EXPERIMENTATION

3.1. Overview

We experimented on the following items to check the efficiency of the proposed anchor setting method.
(1) Evaluation of tracking algorithm
(2) Evaluation of deleting intermediate points

3.2. Data for evaluation

We chose sports video as data for evaluation to focus in applications of the proposed method in this paper to the authoring process of video delivering service in broadband network. For example by applying video hypermedia architecture to the sports video on-demand delivering service, users can get profiles or statistics of the selected player by directly clicking the anchor on the player in the video.

We chose 110 shots, which can be used for video hypermedia service from eleven types of sport programs as video data for evaluation. And we chose 110 objects, used as anchor from selected shots. The object selection criterion consists of choosing the centered object of a shot, such as a player holding the ball in case of a ball sport.

3.3. Experimentation on tracking algorithm

We set the start and end positions at the same position to prepare all shots and executed the automatic tracking process. In this case we experimented both, executing backward tracking and without it. In this experiment, we executed the tracking process every five frames without deleting intermediate points and evaluated the results.

Evaluation process consists of setting the correct anchor region by hand at first, and calculating the overlapped area between anchor region of every key frames after object tracking process and correct anchor region.

We defined the accuracy of every key frame in Formula (3) as an indicator of the accuracy of anchor region of acquired key frames, and calculated the accuracy every key frames.

$$R = \frac{S_b}{S_a} \tag{3}$$

S_b : Overlapped area between anchor region of acquired key frame and correct anchor area

S_a : Area of anchor region of acquired key frame

3.4. Experimentation on deleting intermediate points

We set the start and end positions at the same position to prepare all shots and executed the automatic tracking process. In this case we experimented both, deleting the intermediate points, and keeping them. Backward automatic tracking process was enabled.

The evaluation process is the same as in the case of the experimentation on tracking algorithm. First setting the correct anchor region by hand, and then calculating the overlapped area between the acquired anchor region of every key frame and the correct anchor region.

4. EVALUATION

4.1. Evaluation of tracking algorithm

After the automatic object tracking process, 3645 key frames (average 33.1 key frames/shot) were set by the system.

We calculated the accuracy of every sports type. Table.1 shows the results. According to Table.1, the average accuracy of forward tracking was 0.666 and the average accuracy when combining forward and backward tracking was 0.744.

By combining backward tracking process to the tracking algorithm, the average accuracy was improved 11.7% compared with tracking only in the forward direction.

Table.1 Results of experiment on automatic tracking

Sports Type	Accuracy (forward)	Accuracy (backward)
Scuba diving	0.658	0.772
Tennis	0.660	0.675
Speed skating	0.695	0.799
Football	0.617	0.709
American football	0.758	0.775
Rally	0.701	0.785
Ski jumping	0.647	0.684
Basketball	0.621	0.673
Figure skating	0.641	0.796
Ice hockey	0.704	0.771
Baseball	0.626	0.742
Average	0.666	0.744

Therefore, if the system supports only forward tracking function, the remaining tracking process will be wasted if the system misses the object during the tracking process. We confirmed accuracy improvements of the automatic object tracking function by combining forward and backward tracking used in this case.

However the accuracy of backward tracking for fourteen objects (12.7% of all) was less than that of forward tracking. We analyzed shots which included such objects and found the analogies below.
(1) Missing target object at the end of the forward tracking.
(2) Tracking the background at the border overlapped point of the target object and a part of the background, because target object was small.
(3) Tracking the background at the transformed point of the shape of the target object.

4.2. Evaluation of deleting intermediate points

After the automatic object tracking process, 3645 key frames (average 33.1 key frames/shot) were set by the system. After deleting the intermediate points, they were reduced to 1032 key frames (average 9.38 frames/shot). 71.7% of the key frames were deleted by using this function.

We set the target value of accuracy of the experimentation results to 0.6, 0.7, 0.8 and 0.9, and counted key frames which did not accomplish the target value, calling them correction required frame (CRF). Figure.3 shows the results.

According to Figure.3, if we set the target value of accuracy to 0.9, the average accuracy of key frames by automatic tracking process was 0.744, so it requires to correct an average of 4.653 CRFs per second by hand to accomplish the target value 0.9. By deleting intermediate key frames as described in the previous process, the number of average CRFs is reduced to 1.150 per second.

If we assume that the average work rate of correcting anchor per frame stays constant, we can make the anchor correcting process almost 75% more effective by deleting intermediate key frames, because CRFs per second were reduced from 4.653 to 1.150 by this function.

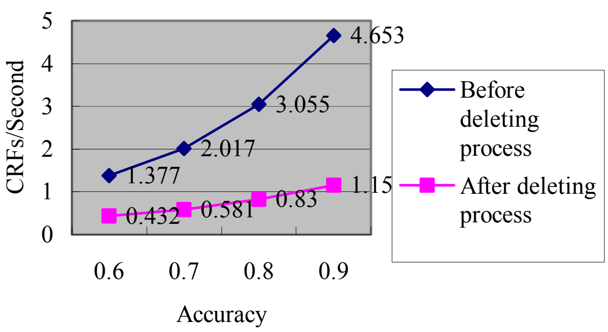

Figure.3 Results of experimentation on deleting intermediate frames

5. CONCLUSION

In this paper, we proposed a method of combining automatic object tracking using motion picture analysis and manual editing as an efficient anchor setting method, which can be used as a start point of a link in a video hypermedia system, and experimented on the proposed method by using sports video for evaluating its effectiveness.

Based on the experimentation results, we confirmed the features of this method (1) improvement of the tracking accuracy by using bi-directional tracking, (2) effective anchor editing by deleting intermediate key frames.

For future work we are considering a more effective anchor setting method and applying the proposed method to the authoring process of broadband video delivering services.

6. REFERENCES

[1] V. M. Bove, Jr., J. Dakss, E. Chalom and S. Agamanolis, “Hyperlinked television research at the MIT Media Laboratory,” *IBM Systems Journal*, Vol. 39, No. 3-4, pp.470-478, 2000.

[2] N. Sawhney, D. Balcom and I. Smith, “HyperCafe: Narrative and Aesthetic Properties of Hypervideo,” *Proceedings of Hypertext’96*, pp.1-10, 1996.

[3] L. Hardman, D. C. A. Bulterman and G. van Rossum, “The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model,” *Communications of the ACM*, Vol. 37, No. 2, pp.50-62, 1994.

[4] K. Tanaka and Y. Tanaka, “Hypermovie Architecture and its Applications,” *Information Processing Society of Japan Special Interest Group of Information Media*, 96-IM-28-16, pp.73-77, 1996. (in Japanese)

[5] T. Takada, K. Sato, S. Aoyagi, T. Hirotsu and R. Onai, “Cmew: Integrating Continuous Media with the Web,” *Japan WWW Conference’97*, 1997. (in Japanese)

[6] H. Abe and K. Wakimoto, “Architecture and Implementation of a Video Object Retrieval Environment,” *2nd IASTED/ISMM International Conference DISTRIBUTED MULTIMEDIA SYSTEMS AND APPLICATIONS*, pp.43-49, 1995.

[7] H. Abe and K. Wakimoto, “Content based Management of Video in a Multimedia Authoring Environment,” *Journal of Multimedia Tools And Applications*, Vol. 2, No. 3, pp.199-214, 1996.

[8] J. Kanda, K. Wakimoto, H. Abe and S. Tanaka, “Video Hypermedia authoring using automatic object tracking,” *IS&T/SPIE’s 10th International Symposium Electronic Imaging: Science and Technology*, pp.108-115, 1998.