MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

Synchronization between Image, Sound and Scenario Document using DP Matching

Yoshitomo Yaginuma, Makoto Kageyama and Masao Sakauchi
Institute of Industrial Science, University of Tokyo
Roppongi 7-22-1, Minato-ku, Tokyo 106, Japan

Abstract

In the management of TV drama, there will be several kinds of requests such as 'Show the image and sound which are correspond to this scene in the scenario document' or 'I want to see the scenes where someone is speaking'. In order to answer to such kinds of requests, the synchronization between images, sound and scenario document have to be carried out. In this paper, a synchronization method between multiple media using DP matching is proposed. In order to carry out inter-media synchronization, in the proposed method, several numbers of patterns of common concepts, which are not depend on the kinds of the media, are extracted from each media and the synchronization between these patterns is carried out using DP matching. The results of simulation experiments reveal the effectiveness of the proposed method.

1 Introduction

Several kinds of studies on color image analysis such as automatic keyword extraction, automatic classification of color images and automatic video indexing have been carried out [1][2] [3][4] [5][6] [7] . However, the recognition of color images by computers is not sufficient to understand all required objects or information.

Meanwhile, in some applications of multimedia field, TV images can give a lot of useful information. The TV drama is composed of image, sound and scenario document. Therefore, higher level recognition which can not be achieved by the image analysis will be carried out by integrating image analysis, sound analysis and document analysis.

In the management of TV drama, there will be several kinds of requests such as 'Show the image and sound which are correspond to *this* scene in the scenario document' or 'I want to see the scenes where *someone* is speaking'. In order to answer to such kinds of requests, the synchronization between images, sound and scenario document have to be carried out [8][9][10].

In this paper, a synchronization method between multiple media using DP matching is proposed. In order to carry out inter-media synchronization, in the proposed method, several numbers of patterns of common concepts, which are not depend on the kinds of the media, are extracted from each media and the synchronization between these patterns is carried out using DP matching.

In section **2**, the inter-media synchronization method using DP matching is proposed. In section **3**, in order to reveal the effectiveness of the proposed method, the results of simulation experiments are shown.

2 Inter-Media Synchronization using DP Matching

The scenario document contains information such as locations, names of speakers and the words of the speakers. Therefore, if the synchronization between image, sound and scenario document can be carried out, semantic retrieval of TV drama can be carried out using such kinds of information. However, the synchronization between image, sound and scenario document is usually difficult because the scenario document does not have the information of time explicitly.

In order to overcome this problem, in this paper, inter-media synchronization method based on the DP matching is proposed.

In order to carry out inter-media synchronization, in the proposed method, at first, several numbers of patterns of common concepts, such as pattern of scene change which represents the discontinuity of scenes, pattern of voice which represents the existence of voice and so on are extracted from multiple media such

as image, sound and scenario document as shown in Fig.1.

The patterns are not depend on the kinds of the media, and the synchronization between these patterns is carried out based on the DP matching Algorithm.

In the following subsections, the extraction of the patterns of the common concepts and the synchronization algorithm are discussed.

2.1 Extraction of Patterns of Common Concepts

In order to carry out inter-media synchronization, in the proposed method, at first, several numbers of patterns of common concepts, which are not depend on the kinds of the media, are extracted from multiple media such as image, sound and scenario document.

As the patterns of common concepts, we can utilize the pattern of scene change which represents the discontinuity of scenes, the pattern of voice which represents the existence of voice and so on.

Here, the extraction of the pattern of scene change from image and scenario document, and the extraction of the pattern of voice from sound and scenario document are discussed.

Pattern of Voice from Sound

The pattern of voice is extracted from sound media by extracting the loudness of the sound(Fig.2(a)). The sound of TV drama is usually well conditioned. Therefore, the voice of TV drama is usually clearly heard. If the loudness of sound is enough large, the region in the sound is regarded to be the region where voice exists as shown in Fig.2(a).

Pattern of Voice from Scenario Document

The pattern of voice is also extracted from scenario document using the number of characters in the words(Fig.2(b)). In the scenario document, the length of the voice is estimated by (the number of characters in the words) × (average time to speak one character). The length of the pause between the words is estimated by ((entire length of the drama) - (length to speak entire words))/(the number of words).

Pattern of Scene Change from Image

In order to reduce the size of image sequence, the extraction of scene change is widely carried out. In the proposed method, the pattern of scene change is utilized as one of the patterns which are utilized to carry out synchronization.

The extraction of scene change is usually carried out by comparing the color distributions of images. If the color distribution of one image is enough different from that of the previous image, the image is regarded to be a scene change.

To evaluate the difference between two color distributions, the χ^2 defined as below is calculated for each image.

$$\chi^2 = \sum_{i=0}^{n} \frac{(H_2(i) - H_1(i))^2}{H_1(i)} \quad (1)$$

Here, $H_1(i)$ and $H_2(i)$ are the color histograms of images, and n is the number of steps of the color histograms. (In the following experiments, at first, the RGB of the image was converted to HVC color space[11]. Then, the color histogram was made from H.)

Usually, if the χ^2 is enough large, the image is regarded to be the scene change. In the proposed method, the pattern of the χ^2 is utilized as a pattern of scene change extracted from the image(Fig.3(a)).

Pattern of Scene Change from Scenario Document

The pattern of scene change is also extracted from scenario document. In order to extract the pattern of scene change from scenario document, the number of characters in the words is utilized(Fig.3(b)). At first, the number of characters of entire words in each scene is extracted from the scenario document. To estimate the length of each scene, the entire length of the drama is distributed in proportion to the number of characters of entire words in each scene.

In this way, several numbers of patterns of common concepts, which can be extracted from at least 2 media, are extracted from each media.

2.2 Algorithm of Synchronization

The synchronization between the patterns of common concepts is carried out using DP matching. The DP matching algorithm carries out pattern matching between 2 patterns by nonlinearly expanding and contracting the patterns. The distance(D) between two patterns is defined as follows.

$$D = \Sigma_{(for\ corresponding\ a_i\ and\ b_j)}\ d(a_i, b_j) \cdot w_{ij} / \Sigma w_{ij} \quad (2)$$

Here, $d(a_i, b_j)$ is the distance between the patterns a_i and b_j, and w_{ij} is the weight of the distance. The correspondence whose distance D becomes minimal is chosen as a correspondence between two patterns. In the following experiment, as a weight w_{ij}, city block distance is utilized and as the distance between two patterns a_i and b_j, the absolute of the difference of

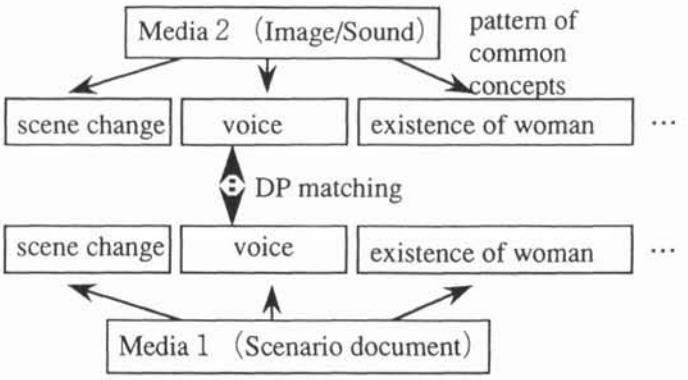
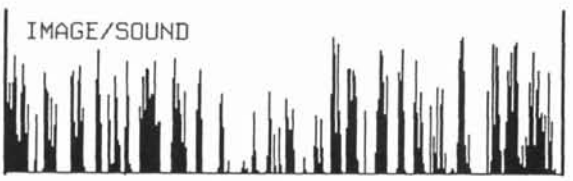


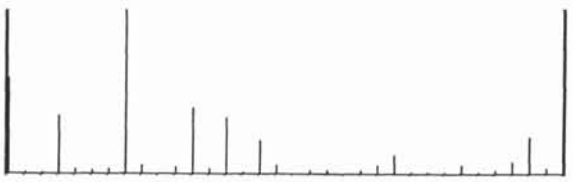
Fig.1 synchronization between multiple media



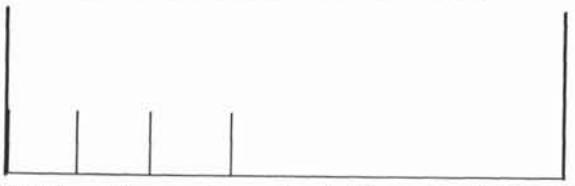
(a) Pattern of voice extracted from sound

(b) Pattern of voice extracted from sceanrio document

Fig.2 Pattern of voice



(a) Pattern of scene change extracted from images



(b) Pattern of scene change extracted from sceanrio document

Fig.3 Pattern of scene change

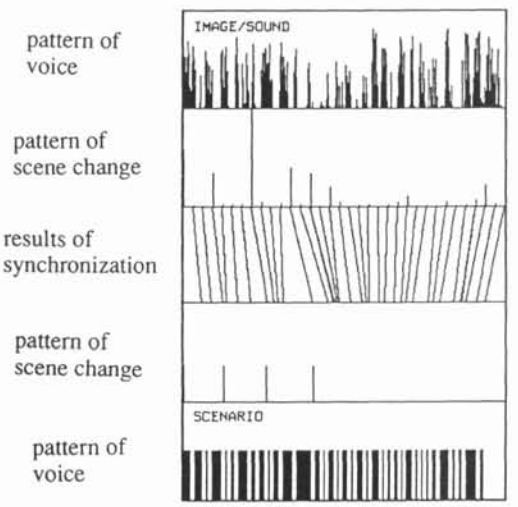


Fig.4 Results of synchronization between images, sound and scenario document

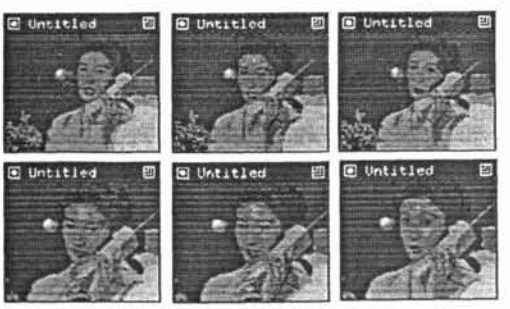


Fig.5 Extraction of specified person

two patterns is utilized ($d(a_i, b_j) = abs(a_i - b_j)$). The distance D can be calculated in a high speed utilizing dynamic programming algorithm.

3 Experiments

To show the effectiveness of the proposed method, experiments were carried out using TV drama. In the experiments, the synchronization of image, sound and scenario document was carried out using two patterns of common concepts; pattern of scene change and pattern of voice.

The experimental result of the synchronization is shown in Fig.4. The TV drama used in the experiment contains 4 scenes described in the scenario document. In this experiment, the synchronization between scene change patterns was perfect. In other words, the image could be decomposed into 4 logical scenes described in the scenario document.

If the synchronization between image, sound and scenario document is carried out, the image can be retrieved using the key words described in the scenario document. For example, the names of speakers are described in the scenario document. Therefore, we can retrieve images which contain specified person using the results of synchronization. The results of the retrieval of a specified person is shown in Fig.5. Fig.5 shows the images which contain a specified person whose name is *Yumi*. The correctness of the synchronization between the sound and the words in the scenario document was about 70%.

4 Conclusion

In this paper, inter-media synchronization method using DP matching was proposed. In order to carry out inter-media synchronization, in the proposed method, several numbers of patterns of common concepts are extracted and the synchronization between the patterns is carried out using DP matching.

To show the effectiveness of the proposed method, experiments were carried out using TV drama. The results of the experiments revealed that the proposed method could successfully carry out synchronization between image, sound and scenario document. It was also shown that the proposed method could successfully carry out the extraction of the specified person.

It remains as a future work to develop a TV drama database which can answer to various requests from users based on the inter-media synchronization.

References

[1] M.Sakauchi and J.Yamane: "Realization of fully automated keyword extraction in image database system", SPIE International Symposium on Optical Applied Science and Engineering (1992)

[2] Y.Gong and M.Sakauchi: "A Method for Color Moving Image Classification Using the Color and Motion Features of Moving Images", ICARCV'92 (1992)

[3] H.Ueda, T.Miyatake and S.Yoshizawa: "IMPACT: An Interactive Natural-Motion-Picture Dedicated Multimedia Authoring System", CHI'91, pp.343-350 (1991)

[4] A.Nagasaka and Y.Tanaka: "Automatic Video Indexing and Full-Video Search for Object Appearances", Transactions of Information Processing Society of Japan, 33, 4, pp.543-550 (1992)

[5] A.Gupta, T.Weymouth and R.Jain: "Semantic Queries with Pictures: the VIMSYS model", 17th International Conference on Very Large Data Bases, pp.69-79 (1991)

[6] Thomas G.Aguierre Smith and Natalio C.Pincever: "Parsing Movies in Context", Proceedings of Summer 1991 Usenix Conference (1991)

[7] HongJian Zhang, Chien Young Low, Yihong Gong, and S.W.Smoliar: "Video Parsing using Compressed Data", SPIE Image and Video Processing II, 2182, pp.142-149 (1994)

[8] G.D.Drapeau: "Synchronization in the Maestro Multimedia Authoring Environment", ACM Multimedia 93, pp.331-340 (1993)

[9] Mitsutoshi Iino, Young Francis Day and Arif Ghafoor: "An Object-Oriented Model for Spatio-Temporal Synchronization of Multimedia Information", International Conference on Multimedia Computing and System 94 (1994)

[10] T.D.C.Little and A.Ghafoor: "Synchronization and Storage Models for Multimedia Objects", IEEE Journal on Selected Areas in Communications, 8, 3, pp.413-427 (1990)

[11] M.Miyahara and Y.Yoshida : "Mathematical Transform of (R,G,B) Color Data to Munsell (H,V,C) Color Data", SPIE's Visual Communications and Image Processing '88, 1001, 118, pp.650-657 (1988)