International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.4 (2014), pp.103-108
http://dx.doi.org/10.14257/ijmue.2014.9.4.11

Human Movement Searching in Video Streams Using Shape Sequence

Min-seok Choi

Dept. of Management Information Systems, Sahmyook University
mschoi@syu.ac.kr

Abstract

Movement information on objects in videos can be used to characterize the content of a scene. This paper proposes a novel shape-based movement-searching algorithm that can effectively search human movements in video streams. Movement information is represented a sequence of boundary shapes of human subjects extracted from input video frames. Information on individual shapes in the sequence is converted into a 1D shape feature by using a shape descriptor. Human movements can be identified in a sequence of shape features extracted from the video, as in the case of searching for words in a long block of text. A comparison of the performance of the proposed algorithm with that of other methods shows that the proposed method can effectively describe human movement information and be useful for movement retrieval applications.

Keywords: *movement search, movement retrieval, shape sequence*

1. Introduction

In video analysis and retrieval tasks, motion information is more important than any other feature [1, 2], and therefore many methods have been proposed to analyze and retrieve video content by using motion information. Motion analysis techniques start by examining the motion of the camera and the trajectory of an object and have recently focused on labeling actions on the screen by analyzing of the movement of objects [3-5].

Several motion descriptors have been proposed and adopted in MPEG-7, the international standard for the interface describing multimedia content [6]. Most such methods focus more on analyzing the direction or trajectory of some motion than on examining the movement of the object itself. However, the movement of an object in a video clip often plays an important role in characterizing the content of the clip. Therefore, a shape variation descriptor for describing the movement of an object in continuous video frames has been proposed and adopted [7]. The motion segmentation step dividing an input video stream into a set of homogeneous clips including individual movements is an important procedure in these methods. However, many real-world video applications are designed to search for shots or scenes wanted by users not in segmented clips but in whole videos. Therefore, a fast and accurate movement search method for large and continuous video streams, such as the one shown in Figure 1, is needed.

ISSN: 1975-0080 IJMUE

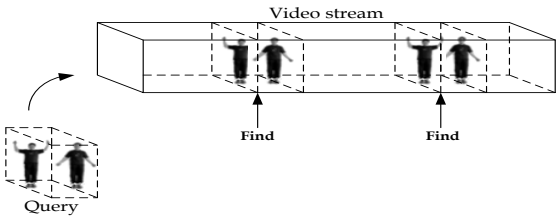

Figure 1. Searching for similar movements in a video

This paper proposes a method for describing shape sequence and searching human movements in video streams. Section 2 provides a brief overview of previous studies, and Section 3 presents the proposed method. Section 4 compares the performance of the proposed method with that of other methods, and Section 5 concludes with a summary.

2. Previous studies of movements

A motion analysis of image sequences or videos is an important element of computer vision [3]. Many studies have examined ways to recognize human movements and can be divided into three groups according to their approaches [4]. The first group reflects a model-based approach in which movement information is described as the pose of the individual by using a 3D model at each time point and the movement is recognized by fitting the 3D model. In contrast to the model-based approach, other methods make use of only the 2D appearance of some given action. A movement is described as a sequence of 2D shapes of an object, and many methods use normalized images of objects segmented from the background. The motion-based approach characterizes the motion itself without referring to the static pose of the body. This section presents three appearance-based methods.

2.1. Temporal templates

To represent and recognize human movements, a 2D temporal template has been proposed [4] and widely used for analyzing and recognizing movements [8, 9]. The temporal template consists of two types of images. A binary motion-energy image (MEI) represents a region where a given motion occurs in the image sequence, and a motion-history image (MHI) is a scalar-valued image in which intensity is a function of the recency of the motion.

An MEI is a binary cumulative motion image indicating the region of motion, whereas an MHI is used to describe the type of motion. Seven Hu moments are used to statistically describe these two images.

2.2. Shape variation descriptor

The purpose of the shape variation descriptor is to retrieve similar movements. To classify a range of complex movements into groups of similar moments in a broad sense, 3D information on movements is represented as simple 2D images by using a shape variation map. This map is a 2D vector image in which the vector value of each point is a function of the given motion [7] and consists of two parts: a low variation map (LVM) and a high variation map (HVM). The LVM describes the shape of the low activity region. The HVM describes the shape of the high activity region during a given movement. The HVM is an inverted image of the LVM except for the background.

The shape variation map cannot be used directly as a feature because it is not invariant to size variation and rotation. In addition, the size of the map is too large to be of any use for the feature. Because shape variation maps are in the form of a gray image, a minor modification

of a region-based shape descriptor is used to extract the feature. The ART descriptor [10], which is scale and rotation invariant, is used to describe the map's shape information.

2.3. Shape sequence descriptor

The shape variation descriptor cannot distinguish individual movements in detail because the shape variation map does not keep track of sequential information on a movement. The proposed shape sequence descriptor represents exact motion information, including both the shape variation and sequence of a given movement [5].

To represent the variation and sequence of the shape of an object, shape features extracted from the object in each frame are aligned along the temporal direction. Shape information on an object in each frame is converted into 1D shape features by using a shape descriptor, and then these features are arrayed in time order to make a 2D feature representing the sequence of shape variations. The shape descriptor used to extract a 1D feature from the object shape is the ART descriptor.

Even when the shape sequence represents a 3D movement into a compressed 2D array, a more compact and efficient feature is needed to construct a movement retrieval or recognition system. For this reason, each column of the ART coefficient in the shape sequence can be transformed into the frequency domain along the time axis to separate common and distinct characteristics into groups of similar movements. Then common characteristics in low-frequency components are selected as a feature to classify each group of movements, that is, the shape sequence descriptor.

3. Movement Searching in a Video Stream

A motion segmentation method for dividing an input video into a set of homogeneous clips is an important step in the movement description methods discussed in the previous session. Shape variation and sequence descriptors can be extracted only from a homogeneous video clip. That is, a movement search using these descriptors can be performed only for video clips divided according to motion segmentation. This section discusses the search methods for finding starting positions of shots or scenes with similar movements in a continuous video stream by using shape variation descriptor and shape sequence.

3.1. Shape variation search

Shape variation descriptors using the shape variation map can be extracted from any intervals or windows in the whole video stream. In addition, because the shape variation map can be constructed continuously from the beginning of a video stream to the end, searching for a particular movement is possible by using a shape variation descriptor. Here the shape variation map (both the HVM and the LVM) can be updated recursively for computational efficiency [11].

The LVM can be calculated using the following recursive equation:

$$LVM(i) = LVM(i-1) + \frac{F(i+d) - F(i-1)}{d} \quad (1)$$

where $LVM(i)$ and $F(i)$ denote the LVM and the segmented image at the i-*th* frame, respectively, and d is the duration of a movement. The HVM can be obtained directly from the LVM.

After the shape variation map is calculated, the ART descriptor is extracted from each map, and the similarity distance between the query movement and the movement in the

window is calculated. Then the window is moved to the next frame, and the process is repeated. In this manner, the similarity distance of each frame in the whole video can be calculated, and positions of shots including similar movements can be selected using these distances.

3.2. Shape sequence search

If an object is separated from the background in each video frame through object segmentation, then it is possible to extract shape information on the object region from each frame. Then a 2D shape descriptor sequence can be obtained from a video. As shown in Figure 2, it is possible to search for a query sequence in the generated sequence (*e.g.*, searching for a word in a block of text).

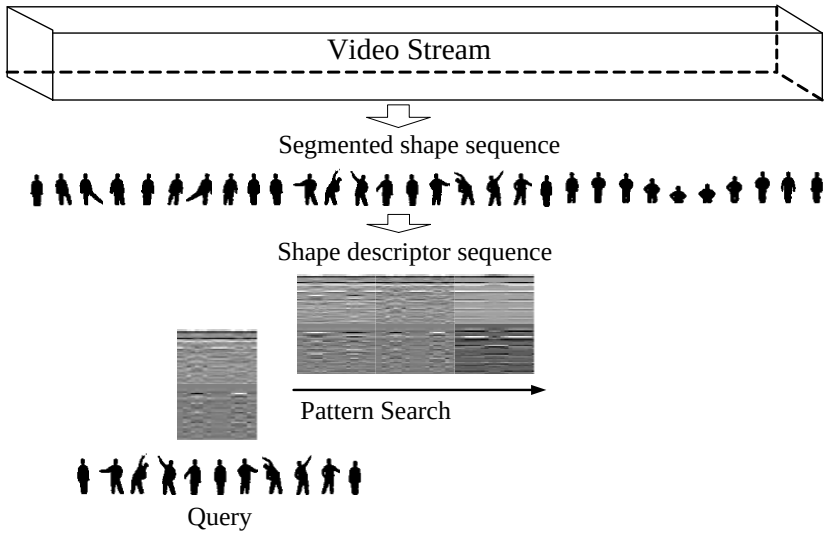

Figure 2. Outline of the shape sequence search

The speed variation of a movement under various conditions is a difficult problem in movement searching. A large search window and minimum-error matching are adopted to reduce the effect of speed variations. To find similar movements in a video stream, the length of the search window is set to be twice the query sequence, and minimum-error matching is performed by overlapping half of the window, as shown in Figure 3.

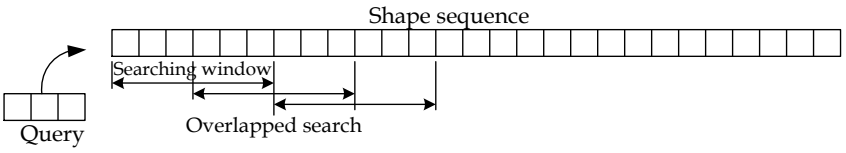

Figure 3. Sequence search scheme

There are two types of matching schemes in sequence searching as shown in Figure 4. One type is a non-ordered matching method that does not consider the order of matched sequences, such as the shape variation descriptor. The other is an ordered matching method considering this order to find exact matched sequences, such as the shape sequence descriptor.

Figure 4. (a) Non-ordered and (b) ordered sequence matching methods

If the order of the sequence is not considered, then the matching problem can be considered as a general minimum-cost problem. Then bipartite graph matching can be applied to reduce the matching complexity of finding the minimum error. The present paper uses the more efficient algorithm in [12] for the experiment. In the case of ordered sequence matching, the greedy algorithm is adopted. This algorithm always makes the best-looking choice at the moment. That is, the algorithm makes a locally optimal choice under the assumption that it would lead to a globally optimal solution and thus does not always yield a genuinely optimal solution. However, the algorithm is simple and efficient and works well for a wide range of problems [13].

4. Experimental results

The test video stream is made by combining 110 segmented video clips depicting 22 types of movements by five human subjects. The number of frames in each movement varies from 20 to 80 according to the type of movement and the characteristic of each human subject.

To evaluate the accuracy of each search method, retrieval performance is compared using the NMRR (normalized modified retrieval rank) [6]. Because the NMRR can take a value between 0 (indicating the whole ground truth found) and 1 (indicating nothing found), the lower the value, the better the retrieval performance is.

The average NMRR of all movements in 22 groups are used to evaluate retrieval performance. The performance of three methods, namely shape variation, non-ordered sequence, and ordered sequence search methods, is compared. Table 1 summarizes the overall performance of each method.

Table 1. Retrieval performance of each search method (NMRR)

Method	NMRR
Temporal template	0.353
Shape variation search	0.263
Shape sequence search (non-ordered)	0.112
Shape sequence search (ordered)	0.230

According to the results, the shape variation map does not reflect the speed variation of a movement, which indicates that the shape sequence search method outperforms the shape variation search method. On the other hand, the shape sequence search method solves the speed variation problem by using a wide search window and minimizing the matching error.

In the sequence search method, the matching scheme using the bipartite matching algorithm outperforms the greedy algorithm. As mentioned earlier, the greedy algorithm does not always produce optimal solutions. However, bipartite matching cannot be applied to search for exact movements in sequential order

5. Conclusions

Simple and efficient schemes for searching and matching human movements in a continuous video stream are proposed. Human movements in videos can be found simply, as in the case of searching for words in a block of text with no movement segmentation, by using shape sequence. In addition, some fast algorithms for calculating the similarity distance between two movements of different durations or frame rates are proposed. The experiment shows promising results for retrieving similar movements in continuous video streams in comparison to the shape variation descriptor.

References

[1] S. F. Chang, W. Chen, H. Meng, H. Sundaram and D. Zhong, "A Fully Automated Content-Based Video Search Engine Supporting Multi-Objects Spatio-Temporal Queries", IEEE Transaction on Circuits and Systems for Video Technology, vol. 8, no. 5, **(1998)**, pp. 602-615.
[2] Y. P. Tan, S. R. Kulkarni and P. J. Ramadge, "Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation", IEEE Transaction on Circuits and Systems for Video Technology, vol. 10, no. 1, **(2000)**, pp. 133-146.
[3] J. Aggarwal and Q. Cai, "Human Motion Analysis: A review", Computer Vision and Image Understanding, vol. 73, no. 3, **(1999)**, pp. 428-440.
[4] A. F. Bobick and J. W. Davis, "The Recognition of Human Movement Using Temporal Templates", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, **(2001)**, pp. 257-267.
[5] M. Choi and S. Choi, "An Efficient Method for Human Movement Retrieval and Recognition Applications", International Journal of Advancements in Computing Technology, vol. 5, no. 12, **(2013)**, pp. 461-469.
[6] B. S. Manjunath, P. Salembier and T. Sikora, "Introduction to MPEG-7: multimedia content description interface", John Wiley & Sons, West Sussex, England, **(2002)**.
[7] M. S. Choi and W. Y. Kim, "The Description and Retrieval of a Sequence of Moving Objects using Shape Variation Map", Patten Recognition Letters, vol. 25, issue 12, **(2004)**, pp. 1369-1375.
[8] A. Iosifidis, A. Tefas, N. Nikolaidis and I. Pitas, "Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis", Computer Vision and Image Understanding, vol. 116, no. 3, **(2012)**, pp. 347-360.
[9] M. A. R. Ahad, J. K. Tan, H. Kim and S. Ishikawa, "Motion history image: its variants and applications", Machine Vision and Applications, vol. 23, no. 2, **(2012)**, pp. 255-281.
[10] J. Ricard, D. Coeurjolly and A. Baskurt, "Generalizations of angular radial transform for 2D and 3D shape retrieval", Patten Recognition Letters, vol. 26, issue 14, **(2005)**, pp. 2174-2186.
[11] M.-s. Choi, "An Efficient Search Method for Human Movement in Video Steam", Proceedings International Workshop, HCI 2013, Multimedia 2013, **(2013)** December 11-13, Jeju Island, Korea, pp. 71-74.
[12] R. Jonker and A. Volgenant, "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems", Computing, vol. 38, issue 4 **(1987)**, pp. 325-340.
[13] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms – second edition", The MIT Press, London, UK **(2001)**.

Author

Min-seok Choi

He is an assistant professor in the department of Management Information Systems at Sahmyook University, Korea. He received his B.S. and Ph.D. in Electronic Engineering from Hanyang University in 1998 and 2004, respectively. His research interests include machine vision and content-based image and video retrieval.

