Motion Capture Employing an Uncalibrated Camera

J K Tan, S Ishikawa, K Kouno, H Ohbuchi and H S Kim

Department of Control Engineering, Kyushu Institute of Technology, Kitakyushu, Japan
(E-mail: {etheltan, ishikawa}@ss10.cntl.kyutech.ac.jp)

Abstract: This paper describes a novel optical motion capture technique. Motion capture is a technique for producing a 3-D model of a human motion or action. Unlike the existent methods of motion capture, the presented technique neither performs camera calibration nor employs markers. Instead it makes use of a motion database, and, by retrieving the database, it recognizes the unknown motion of an observed human and it reproduces the motion by use of a digital human model. This new motion capture system may provide a simple and easy-to-use motion capture system. Performance of the proposed technique is shown experimentally.

Keywords: Motion capture, database, eigenspace, pattern recognition, digital human.

I. INTRODUCTION

Motion capture is a technique for creating a 3-D model of a human motion or action using various sensors and a computer. Among them, the motion capture employing cameras called optical motion capture is the most popular, since it doesn't give a person constraints on his/her motion unlike the system employing magnetic sensors or mechanical instruments. It is, however, definitely necessary to calibrate the used cameras before video capture [1]. This gives difficulty in modeling transient or unexpected events such as the sudden performance of street dance or a certain accident where there is no occasion on camera calibration. The existent optical motion capture also employs markers by which 3-D shape is calculated. This is again a drawback for the 3-D modeling of transient human actions or events. Thus the existent motion capture techniques cannot extend their own applicability to a variety of human motions and actions in our life.

Unlike the existent techniques [2,3,4,5], the technique presented in this paper neither performs camera calibration nor employs markers. Instead it makes use of a motion database, and, by retrieving the database, it recognizes the motion of an observed human and reproduces the motion by use of a digital human model. In this way, a new motion capture system is realized which may provide a simple and easy-to-use motion capture system.

II. MOTION RECOGNITION BY A MOTION DATABASE

In a 3-D space, we observe a human motion from various orientations, and yet we understand what the motion represents. This should also be realized in robot vision. In the proposed system, a human motion is recognized by a camera irrespective of its orientations of observation. This is realized by taking video images of the motion interested from multiple directions and registering the images in an eigenspace as a series of points each of which represents an image frame [6]. An example of motion representation in an eigenspace is given in **Fig. 1**. In Fig. 1, (a) is an original motion of 'Lifting up a box'; (b) is part of its motion sequence shown in differentiated images which we employ instead of the original gray value images; and (c) the motion representation in an eigenspace. As seen in the figure, the motion concerned is described by a set of point sequence. This eigenspace corresponds to a single camera orientation.

Alternatively, a motion sequence can be compressed into a single image such as motion history images [7], directional motion history images [8] or JK motion representation [9], and is registered into the eigenspace.

A motion database is a set of the eigenspaces. A single eigenspace corresponds to a single camera orientation and it contains multiple points representing multiple motions. If a motion database is denoted by a set B and the eigenspace corresponding to the rth camera orientation by $ES(r)$, B is written as

$$B = \{ES(r) \mid r = 1,2,...,R\} \qquad (1)$$

Here R is the number of cameras employed or equivalently the number of the orientations of observation. It should, however, be noted that all the eigenspaces must be defined by a common data set in order that they have common eigenvectors. Namely, once we have obtained the video images of M motions of S persons taken from R orientations, provided that each video contains F image frames, A=$FMSR$ image frames are employed for making a covariance matrix from which a set of eigenvectors are calculated. These eigenvectors define R eigenspaces and the respective eigenspaces only contain the motion data observed from respective orientations. The image of a motion database is given in **Fig. 2**.

Once an unknown motion video is given, the frame sequence is differentiated and transformed into a series of points, and their match is calculated with every point series in the motion database, i.e., their match with every point series in the respective eigenspaces in the motion database. The best match point series having the largest correlation value above a predefined threshold is recognized as the motion the unknown video represents.

This matching strategy suggests parallel processing among the R eigenspaces in order to find the best match within a reasonable process time. Otherwise the process time may increase impractically according to the increase of the observing orientations. But we employ a single PC to realize a low cost system and the high speed matching is realized by the employment of a structured eigenspace [10], in which an eigenspace is described in terms of a B-tree.

III. MOTION REPRODUCTION BY HARDWARE

Normally, in the existent motion capture systems, the obtained 3-D motion data is fed to an avatar, a 3-D graphic human model in a computer, to reproduce the acquired human motion and we can observe the avatar's motion in a display. The avatar can be called a digital human model. But in our motion capture system, the obtained motion is represented by a human-figured robot. Recently human-type robots such as ASHIMO [11] have been developed rapidly and inexpensive robots have become commercially available. We employ such hardware to reproduce the motion recognized by the proposed system. This representation is much more understandable of 3-D motions than avatars in the display, though precise motions cannot be well acted yet.

The system configuration is given in **Fig. 3**. A motion database needs to be produced in advance, in which multiple cameras are employed. In the motion capture stage, a single camera and a PC system captures a video image of a human unknown motion and it is retrieved in the motion database to find the best match. Once the matched motion is found, it is represented by a robot.

IV. EXPERIMENT AND RESULTS

We performed an experiment for showing performance of the proposed motion capture system. Twenty-three human motions are registered in a motion database. They include; raising a right/left hand horizontally/vertically, stooping and stretching, deep/shallow bowing, etc. But, in this particular experiment, one frame that represents each motion is chosen and registered in the motion database for simplicity. Some of the motions are shown in **Fig. 4**. A fixed and uncalibrated camera captures the initial background image and then an unknown motion video of a subject. Difference between the background and each image frame of the video are computed to obtain the images containing only human areas. They are processed to recognize the unknown motion by using the motion database. The recognition rate is approximately 80% to 85% in the laboratory.

The developed motion capture system was exhibited in a workshop and it showed successful performance there. Some robot motions demonstrated in the workshop are shown in **Fig. 5** along with corresponding human motions.

V. CONCLUSION

The present paper proposed a new optical motion capture system which neither perform camera calibration nor employ markers. Some of the reasons why the existent motion capture systems do not spread wide among human analysis-related researchers are that they require camera calibration and the system is too expensive. Many researchers request simpler operation of the system and lower price. The proposed system may offer a simple and easy-to-use motion capture system. If we restrict captured motions in some particular fields, it is easier to produce a motion database. Then we may establish a practical motion capture system in a special field.

It is obvious that extraction of a human area from the background is indispensable after obtaining a human motion video. If a camera is fixed and the background is not very complex, the extraction is not very difficult. On the contrary, in the case that the background is complicated such as the image taken outdoors, the technique [12] can be effectively employed to extract a human in motion.

When the proposed motion database becomes large, the retrieval time may increase. Then we can introduce a tree structure in order to maintain fast retrieval [10]. This is particularly important in order to realize a low cost system. Otherwise we may have to introduce parallel processing to achieve fast data retrieval on the motion database.

This study was supported by KAKENHI of Japan Society for the Promotion of Science.

REFERENCES

[1] Gonzalez RC, Wintz P(1987), Digital Image Processing. Addison-Wesley

[2] Abdel-Aziz YI, Karara HM (1971), Direct linear transformation from comparator coordinates into object space coordinates in close range photogrammetry. Proc. American Society of Photogrammetry 1-19

[3] Sugimoto N, Xu G(1999), A linear algorithm for motion from three weak perspective images using Euler angles. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(1):54-57.

[4] Tan JK, Ishikawa S(2001), Deformable shape recovery by factorization based on a spatio-temporal

measurement matrix. Computer Vision and Image Understanding 82(2):101-109

[5] Yamaguchi I, Tan JK, Ishikawa S(2007), 3-D modeling of remote dynamic situations using mobile stereo cameras. Artificial Life and Robotics 11(1):96-100

[6] Rahman MM, Ishikawa S(2005), Human motion recognition using an eigenspace. Pattern Recognition Letters 26:687-697

[7] Bobick A, Davis J(2001), The representation and recognition of action using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(3):257-267

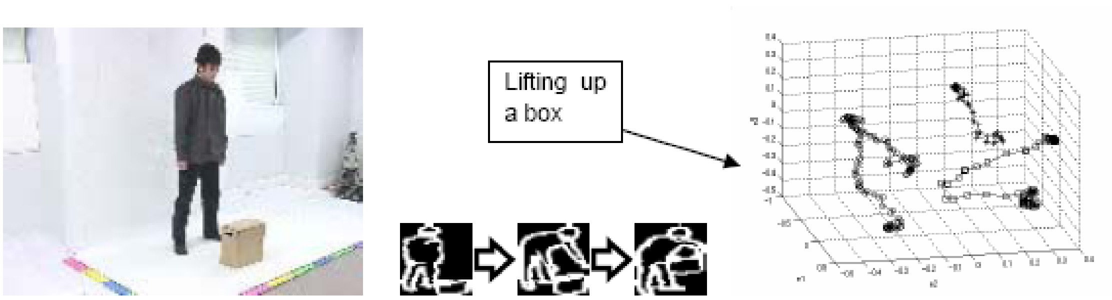
[8] Ahad MAR, Ogata T, Tan JK, Kim HS, Ishikawa S(2007), Performance of multi-directional MHI for human motion recognition in the presence of outliers. Proc. of IECON 2366-2370

[9] Tan JK, Ishikawa S(2007), "High accuracy and real time recognition of human activities. Proc. of IECON 2377-2382

[10]Tan JK, Kouno K, Ishikawa S, Kim HS, Shinomiya T(2007), High speed human motion recognition employing a motion database. Imaging & Visual Computing 36(5):738-746. (in Japanese)

[11] http://www.honda.co.jp/ASIMO/

[12] Ohba O, Tan JK, Kim HS, Ishikawa S(2007), Recognizing human postures in a complex background employing MRRF method. Proc. of Int. Conf. on Control, Automation and Systems 1159-1162



(a) (b) (c)

Fig. 1. Human motion representation in an eigenspace: (a) Original motion (Lifting up a box); (b) its image sequence given by the differentiated images; and (c) its eigenspace representation.

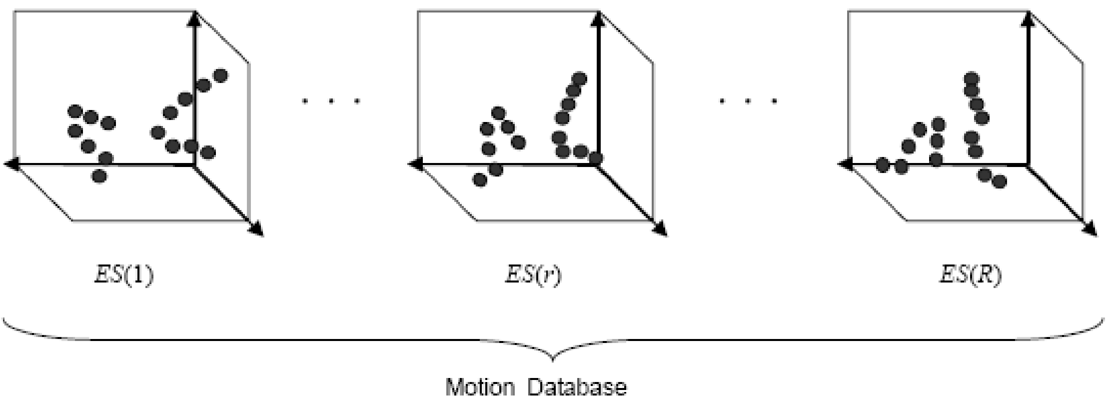


Fig. 2. Motion database containing R eigenspaces. Eigenspace $ES(r)$ corresponds to the rth observation orientation.

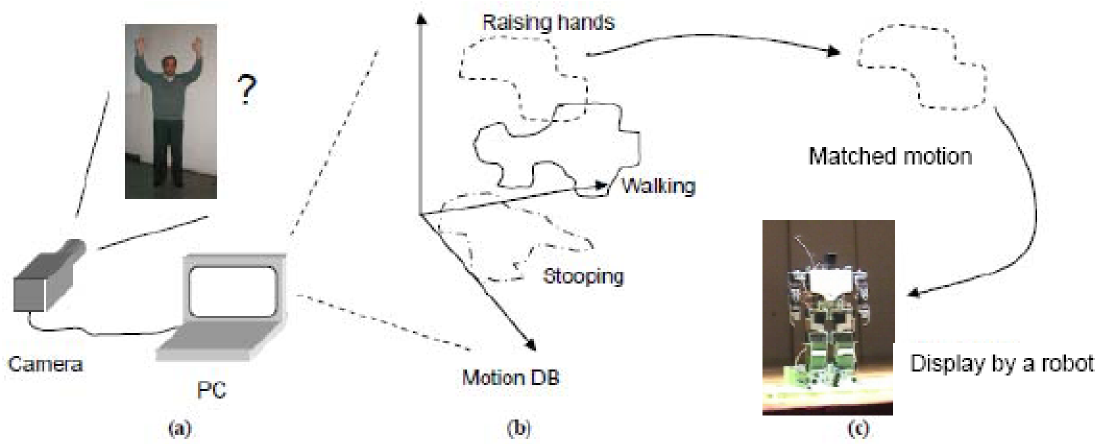


Fig. 3. System configuration: (a) A single camera and a PC provide a video capture and motion recognition system: (b) A motion database is prepared in advance: (c) The recognized motion is repeated by a robot having a human figure.

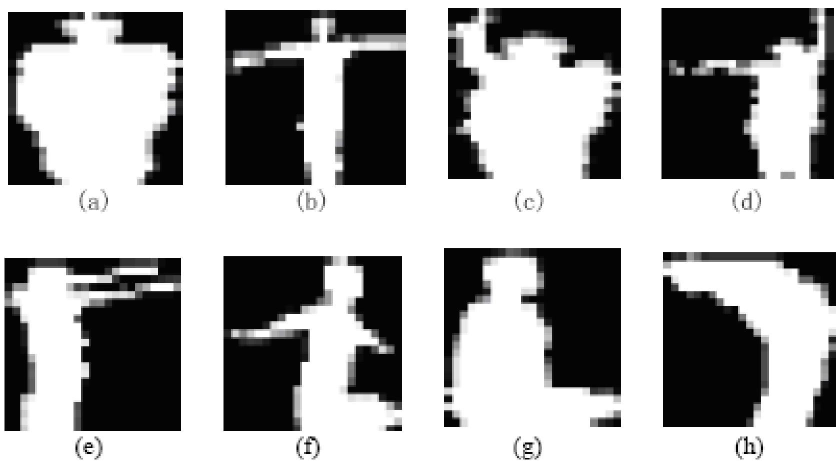


Fig. 4. Examples of employed motions.

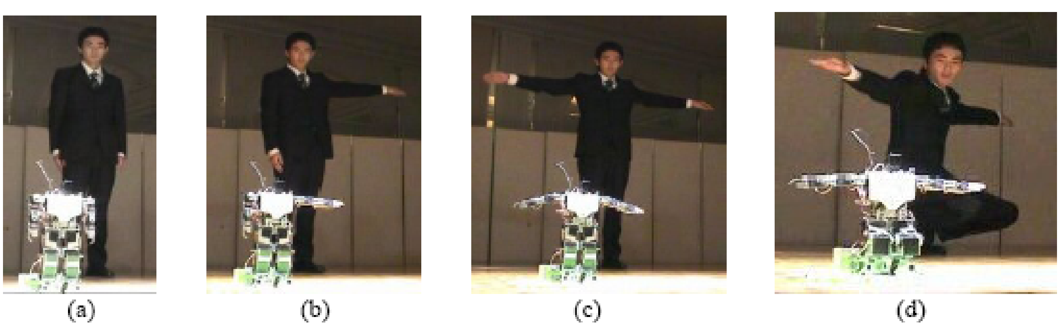


Fig. 5. Some human motions and a robot repeating the motion. The system recognizes human motion and it is repeated by the robot. (a) Standing: (b) Stretching the left arm: (c) Stretching the both arms: (d) Stooping and stretching the both arms.