Abstract

Measurement of human motion is widely required for various applications, and a significant part of this task is to identify motion in the process of human motion recognition. There are several application purposes of doing this research such as in surveillance, entertainment, medical treatment and traffic applications as user interfaces that require the recognition of different parts of human body to identify an action or a motion. The most challenging task in human motion recognition is the ability and reliability of the motion capture system to track and recognize dynamic movements because human body structure has many degree of freedom. This paper introduces the 3D motion analysis of human upper body using an optical motions capture system for the purpose of gesture recognition.

1. Introduction

An optical motion capture system or MoCap, is a technique of digitally recording movements of a real object and mapping it onto a computer-generated objects for the purpose of analyzing various types of movements. Motion tracking or motion capture was started as a photographic analysis tool in biomechanics research, and was expanded into education, training, sports and recently computer animation for cinemas and video games as the technology matured. There are two main techniques used in optical motion capture system, reflective marker and Pulsed-LED (light emitting diodes). An optical motion capture systems tend to utilize proprietary cameras to track the motion of reflective markers (or pulsed LED's) attached to joints of the performer's body. The performance of motion capture system depends on the number and resolution of a camera, which are used in the system.

Recently, various types of motion capture systems are introduced for the purposes of measuring human motions such as gloves, magnetic motion capture and optical motion capture, which has a certain advantage in the comparison with each other, and are used depending on the user application. An optical motion capture system is unburden to a performer and has the ability of measuring unlimited movements, however it has problems due to the occlusion when bending the body joints and the ambiguities of the markers. Gloves and magnetic sensors are able to solve these problems, but the system gives burden to the performer and has the limited movements because it is equipped with cables connecting sensors, which are attached at the body parts. An optical motion capture system gives a best solution for the measurement of human motions if occlusion problem can be solved, and recently plural cameras are used in the system depending on the application such as tracking rigid parts of the body or a whole body.

Interest of researches using an optical motion capture system to track a moving object are divided into two; which are motion analyses and problem solving of occlusion of moving markers. In the case of motion analyses, the information captured from moving markers such as positions, angles, velocities and accelerations are mapped to computer-generated objects for the purpose of animation or interfacing with another computer-oriented machines. Others, small numbers of researches focus on developing a new algorithm for solving an occlusion problem, while tracking moving markers. Comparing with marker-less tracking system, an optical motion capture system gives precise marker data with less noise, and promises a good result for tracking system.

This paper introduces the analytical studies of human upper body movements using an optical motion capture system for the purpose of gesture recognition. An image processing technique is introduced to track optical markers attached at feature points of a human body for constructing a human upper body model and estimating its three dimensional motions.

This paper is structured as follows: Section 2 addresses the related researches to the application, and problem of motion capture system are presented. Section 3 describes the purposes of study which are related to the ability of commercialise optical motion capture system. Section 4 describes the configuration of the experimental equipment and the subject. Section 5 describes the construction of human upper body model and its kinematics constraints that include the calculation of joint angle. The algorithms for estimating joint angles together with a prediction process are reported. Section 6 presents experimental results, including an evaluation in terms of prediction a missing marker. We conclude the article with a short summary in section 7.

2. Related researches

Many studies for analyzing human gesture for the purpose of human machine interfacing has been reported
so far, and most of them are marker-based using tracking system because marker-less tracking system needs high computational cost especially for the system, which have more than two camera with a high resolution. Working with marker-less tracking system has to deal with noisy data, and the development of database system to match a present hand posture and features stored in database are reported in [1][2][3]. An optical motion capture system gives a precise data compared with marker-less tracking system, however it has to deal with occlusion because of high degree of freedom at the body joints while doing movements, especially fingers that have many joints and distances between them are very close. N. Miyata and her team used an optical motion system with seven cameras to track and reconstruct hand link structure while grasping a cylinder [4]. More than two cameras are usually required to track the moving markers at fingers position for recognizing hand shapes because of the complexity of the joints and the visibility of markers. The applications of an optical motion capture system for tracking a whole human body movements have been studied and reported [5], [6], and give a good reference for the application of tracking and recognition rigid parts of human body. There are still few researches working on tracking and recognizing rigid parts of human body using an optical motion capture system especially for the purpose of gesture recognition, and the ability of its functionalities remain unknown unless research results of the application are published. In this paper, we introduce the analytical studies of human upper body motion using an optical motion capture system for the purpose of gesture recognition to create a user friendly interface between human and machine.

3. Purpose of the study

As mentioned earlier, although an optical motion capture system provides high accuracy for capturing moving markers, pre-processing steps are required before capturing data are proceeded for post-processing to recognize gestures or poses. The main problem working with an optical motion capture system is the capability of cameras to track and recognize moving markers. A commercialized optical motion capture system provides automatic tools for calculating motion features, however it still remains to be done manually such as to correct wrong markers position and solve hidden markers. In this study, a skeleton model of a performer is created based on captured 3D data in the initial pose, and is used to detect the wrong marker positions in the motions. The calculation of joint angle is presented, and is used to predict the position of a body joint that is missing.

4. System configuration

An optical motion capture system which tracks 9 markers simultaneously in real time is used for the motions measurement. The system is equipped with 2 high-speed cameras with the ability of capturing 120 frames per second. The movements of arm motions are mainly dealt to analyze human gestures by excluding hand motions, due to the complexities of finger configurations and the occlusion of markers. Nine markers are attached at the feature points of the body of the performer, which are the chest, head, shoulder, elbow and wrist for both right and left upper body, as shown in Figure 1. The system estimates the motion of human upper body by tracking the reflective markers attached to the body of a performer.

Figure 1. A schematic diagram for marker positions

5. Methods

5.1 Upper body structure model

Based on the positions of 9 reflective markers, a human body structure model is constructed as illustrated in Figure 2. The body parts are relatively connected, and the model is determined by considering the human kinematics calculated on the global coordinate system. According to the motions of human upper body, only the length of the sticks are considered as the constants for tracking and predicting hidden markers. Angles of the body joints are calculated for the purpose of motion measurement.

Figure 2. Skeleton model of human upper body
Lengths of the sticks $T_j (j = 1, 2, 3, ..., 8)$ are given based on the earliest frames of a certain gesture. The moving range of each joint is set according to the ability of human movements [7]. Positions of the body parts are estimated and determined by 9 parameters $\Theta = \theta_i (i = \gamma_1, \theta_1, \eta_1, \alpha_1, \theta_2, \epsilon_2, \eta_2, \alpha_2)$, however the correspondence between markers and their positions on the body is hard to be established. The correspondence between 3D positions of marker and body parts are done by comparing with the amount of movement of the 3D points. The distances between markers are used to identify body parts in motions. For discriminating body parts using the location data of markers, equations (1) and (2) shown below are established by referring to figure 2.

$$H_{RT} = \begin{bmatrix} \alpha_1 & T_3 & \theta_1 \epsilon_1 \eta_1 & T_2 & \gamma_1 \Omega \end{bmatrix}$$

(1)

$$H_{LT} = \begin{bmatrix} \alpha_2 & T_6 & \theta_2 \epsilon_2 \eta_2 & T_7 & \gamma_2 \Omega \end{bmatrix}$$

(2)

The equations are used for determining the points that are connected by sticks $T_3, T_4, T_5$ and $T_6, T_7, T_8$ of the right and left hands represented by the transformations $H_{RT} = (x_R, y_R, z_R)^T$ and $H_{LT} = (x_L, y_L, z_L)^T$, respectively.

The average values of the sticks are determined in the earliest frames, and are used as a reference for the trackers while doing measurement. The threshold value is set for discriminating each stick, depending on the relation between different body joints, for example, the motion of lower arm is larger than the upper arm. Equation (3) shows the mathematical relation for determining the body length.

$$T_n = I_{frame} \pm Th_n$$

(3)

A coordinate transformation, which includes the translation and the rotational component are used to obtain the position of markers as time proceeds [3][4]. As mentioned earlier, the sticks $T_n$ are used as a model for tracking correct markers while a performer gives gesture. Stick data of the earliest frames $I_{frame}$ of particular gesture are determined to be a model for the rest of the frames in the case where trackers failed to track moving markers or tracked wrong markers.

5.2 Joint angle calculation

In this paper, the chest angle is assumed static while movement is done, and constraints are given to the motion of arm. A human arm and two local coordinate systems are shown in Figure 3. Figure 4 shows the position of elbow with respect to the rotational angle $\eta$.

Base on the structure of human arm in Figure 3 and Figure 4 as above, the rotational angles of shoulder and elbow can be determined by a kinematics constraint. Considering the coordinate system in Figure 3, the pronation and supination $\theta$ are the rotations around $y$-axis, when hand is hanging by side of the body, and this is defined in absolute coordinate as the pitch angle. Similarly, flexion/extension $\eta$, and abduction/adduction $\epsilon$, are rotations around the $x$ and $z$-axes, respectively, and thus are described as roll and pitch angles, respectively. The formulation that is devised by John Seocthing [10] is used to calculate the joint angle, and by following his approach, and a reference frame at shoulder remains stationary and another frame rotates with the moving arm. Given the shoulder position $S$, the elbow position $E$ and the wrist position $H$ in the shoulder and elbow proximal frames, $\theta$, $\eta$, and $\epsilon$ are calculated by the combination of Eulerian angles in the rotational order of Y-X-Y and Z-X-Z. Notice that $T_5$, $T_4$ are the length of the lower arm and upper arm respectively, and then three shoulder angles can be found as:
\[ \varepsilon = \arccos(E_y, T_4) \quad (4) \]
\[ \theta = \arctan(E_z, T_4 \cdot \sin \varepsilon) \quad (5) \]
\[ \eta = \arcsin(A(B \cdot E_y - A \cdot E_x), T_4(B^2 + A^2)) \quad (6) \]

Therefore:
\[ \eta = \begin{cases} \pi - \eta & \text{if } \sin(\eta) > 0 \\ \eta & \text{otherwise} \end{cases} \quad (7) \]

where \( A \) and \( B \) denote \( \sin(\theta) \) and \( \cos(\theta) \cdot \cos(\varepsilon) \), respectively.

Equations (4), (5), and (6) provide a rotational matrix that transforms any vector in \( E \) frame into \( S \) frame, and consequently computes the hand position with respect to shoulder. However, to describe a vector that always points to hand, the degree of freedom existing at elbow must be calculated and the 2DOF at elbow are flexion/extension and pronation/supination. Because the configuration of human arm is regarded as a stick that has a joint at the elbow and a joint at the shoulder, a pronation or supination of the elbow should not change the position of the hand.

Figure 5. Shoulder angle calculation. (a) Abduction/Adduction. (b) Pronation/Supination. (c) Flexion/Extension
This results that only a single degree of freedom exists at elbow, and by giving the shoulder position $S$ and the hand position $H$, the elbow angle $\alpha$ is uniquely determined by the following equation:

$$\alpha = \arccos \left( \frac{\left( T_4^2 + T_5^2 - d^2 \right)}{2 \cdot T_4 \cdot T_5} \right)$$

(8)

where $d$ is the distance from shoulder to wrist joint of the hand.

Figure 5 shows an example of measurement of human shoulder, which are abduction/adduction, pronation/supination and extension/flexion when hanging arm by body side, which changes the roll, pitch and yaw angle, respectively. According to the movement of the human upper body joint, the range of angle is restricted as shown in Table 1. However, the actual values of these limitations are not universal and differ between individuals.

Table 1. Parameters of upper body

<table>
<thead>
<tr>
<th>Body Joint</th>
<th>Angle</th>
<th>Max</th>
<th>Min</th>
<th>Stick</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest</td>
<td>$\gamma_1$</td>
<td>180</td>
<td>-180</td>
<td>$T_1T_2T_3T_6$</td>
</tr>
<tr>
<td>Right shoulder</td>
<td>$\varepsilon_1$</td>
<td>-90</td>
<td>90</td>
<td>$T_4$</td>
</tr>
<tr>
<td></td>
<td>$\eta_1$</td>
<td>-45</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_1$</td>
<td>-90</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Right elbow</td>
<td>$\alpha_1$</td>
<td>180</td>
<td>0</td>
<td>$T_5$</td>
</tr>
<tr>
<td>Left shoulder</td>
<td>$\varepsilon_2$</td>
<td>-90</td>
<td>90</td>
<td>$T_7$</td>
</tr>
<tr>
<td></td>
<td>$\eta_2$</td>
<td>-45</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>-90</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Left elbow</td>
<td>$\alpha_2$</td>
<td>180</td>
<td>0</td>
<td>$T_8$</td>
</tr>
</tbody>
</table>

5.3 Tracking and prediction of hidden marker

Position of a missing marker in the next frame is predicted using the information from the previous frames based on the skeleton model of the human upper body. The position of missing marker $N$ at time $t$ is predicted using position data of another marker $P$, which is physically connected to it, for example, the missing marker at wrist joint is predicted using elbow joint at time $t$ and the relation between them in the previous frame at time $t-1$. The relation between joints depending on the amount of DOF existed in the reference frame, for example the position of hand is respect to the 3DOF, which are 2DOF at shoulder and 1DOF at elbow. As a reference angle, another markers are attached at lower and upper arm when a marker is missing, and are used to estimate joint angle based on the threshold in the initial pose as shown in section 5.1. Equation (9) shows the mathematical relation for the prediction of missing marker.

$$N_t = R_n(n)P_t + (N_{t-1} - P_{t-1})$$

(9)

$R_n(n)$ is a rotational matrix of the joint and $n$ is the number of DOF exists in reference frame. $(N_{t-1} - P_{t-1})$ is a translation matrix between $N$ and $P$ in the previous frame.

6. Result

An experiment was conducted to predict the missing markers while a subject gave an arm motion. During the motion, the left hand was occluded by the right hand. The motion of the lower arm stick was predicted by using a threshold value obtained in the initial pose and estimated angle of elbow joint.

Figure 6 shows the prediction of the missing marker at left hand in two different motion (a) and (b).

Figure 6. Prediction of missing marker at left hand in two different motion (a) and (b).
the lower arm location from others depending on the relation between different body parts, which was calculated in the preliminary experiment. A reference angle as explained in section 5.3 was used to estimate the position of the left hand.

7. Conclusion

In this paper, an image processing technique to track optical markers attached at feature points of a human body for constructing a human upper body model and estimating its three dimensional motion was introduced. Prediction of missing marker in the present frame was introduced by using the angle of the connected joint in the reference frame. For the purpose of gesture recognition, an optical motion capture system provides high accuracy for capturing moving markers in 3D space, however it still has a weakness in pre-processing step, which is the ability for cameras to track moving markers. By using the proposed approach, the ability of an optical motion capture system to track moving markers is improved and can be used to analyze upper body motion. For tracking a complex motion with more markers, it still requires improvements to be done, but for the purpose of analyzing arm gestures, the proposed method was proved to be enough for practical use. For the further studies, particular features from the motions will be extracted for the purpose of gesture recognition and also for the identification of particular person among plural people by observing characteristic body motion.

Reference


