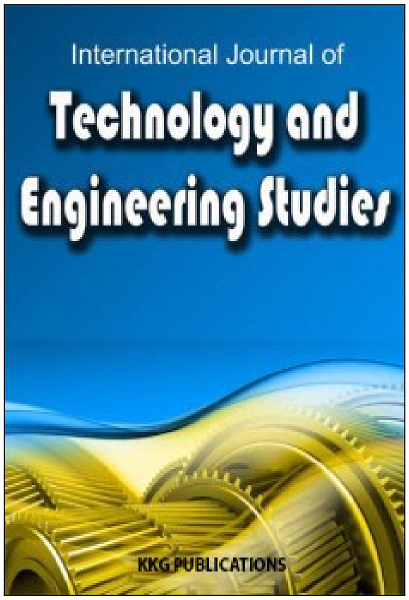

Key Knowledge Generation

Publication details, including instructions for author and subscription information:
http://kkgpublications.com/technology/

Markerless Motion Capture for Entrance Guard Systems

TAINCHI LU [1], KUANCHIEH SUNG [2], BOLIN JIANG [3]

[1, 2, 3] Department of Computer Science and Information Engineering, National Chiayi University, Chiayi, Taiwan

Published online: 13 December 2016

To cite this article: T. Lu, K. Sung and B. Jiang "Markerless motion capture for entrance guard systems," *International Journal of Technology and Engineering Studies,* vol. 2, no. 6, pp. 172-179, 2016.
DOI: https://dx.doi.org/10.20469/ijtes.2.40002-6

To link to this article: http://kkgpublications.com/wp-content/uploads/2016/2/Volume2/IJTES-40002-6.pdf

International Journal of Technology and Engineering Studies
vol. 2, no. 6, pp. 172-179, 2016
IJTES

MARKERLESS MOTION CAPTURE FOR ENTRANCE GUARD SYSTEMS

TAINCHI LU [1]*, KUANCHIEH SUNG [2], BOLIN JIANG [3]

[1, 2, 3] Department of Computer Science and Information Engineering, National Chiayi University, Chiayi, Taiwan

Keywords:
Character Animation
Markerless Motion Capture
Pose Estimation
Motion Tracking
Entrance Guard System

Received: 12 June 2016
Accepted: 25 September 2016
Published: 13 December 2016

Abstract. This research presents a practical markerless motion capture method and uses motion estimation for developing an entrance guard system. In recent years, markerless motion capture is used to track human motions in various applications, from entertainment to surveillance. The markerless-based technology is more flexible and less cumbersome in comparison with marker-based optical systems. In the first step, a binary image is calculated to carry out background subtraction and region partition. Afterward, we reconstruct an articulated skeleton model to measure the model-to-image similarity with the binary image to possess locations of the specific end joints. In addition, a decision tree is defined in advance to facilitate pose estimation. As a result, a valid or invalid command can be determined to control the door lock through the proposed entrance guard system. The current experiments demonstrate that this method can capture human motions and applies to the entrance guard system.

INTRODUCTION

Over a decade, motion capture technologies are successfully applied to a wide variety of applications, such as athletic sports, movies, animations, games, surveillances, and entertainment. It is more convenient and effective to obtain realistic continuous motions by making good use of the motion capture. In addition, we do not pay more attention to tediously adjust skeletons and joints for the purpose of acquiring accurate motions. It would be beneficial to decrease the system development time and to perform motion analysis for further versatile usages. For example, we can capture continuous motions of athletes to evaluate their professional physical performance from motion captured data. Generally, motion capture technologies are categorized into two parts, one is motion tracking, and the other one is motion recognition. Motion tracking is to keep track of the marked joints while an actor is acting in order to capture the motion trajectories. In contrast, motion recognition is to take advantage of off-line statistical or training data to recognize user's behaviors or intentions. In this paper, we concentrate on the motion tacking parts and make use of pose decision tree to facilitate the motion recognition in an entrance guard system.

In practice, motion data can be acquired by marker-based or markerless motion capture systems. With respect to marker-based motion capturing, discontinuous trajectories of motions may happen due to inaccurate markers, and it is cumbersome to use while setting up the system and wearing clothes. The aim of the paper is to provide low-cost motion capturing experience towards humanoid characters; therefore, the markerless method is adopted in this paper. Kinect technology released by Microsoft is the most popular markerless motion capture device in the past few years. Even though Kinect provides efficient motion capture results, it also suffers from problems of body occlusion, joint rotation, and variant illuminance. In this paper, we make use of webcams to capture human motions, and take advantage of articulated skeleton to reconstruct a human model. Region partition and pose decision tree are proposed to benefit human motion capture in order to reduce computational time. For verifying the proposed method, we apply it to an entrance guard system to evaluate its feasibility and flexibility.

The remainder of the paper is organized as follows: Section 2 addresses the related studies about markerless motion capture and articulated skeletons. In Section 3, we present the adopted technologies in detail. Implementation and experimental results are shown in Section 4 for demonstrating the effectiveness of the proposed method. Section 5 concludes the paper and addresses some future directions.

LITERATURE REVIEW

Markerless motion capture system is a motion trail recording technology that does not need marked points. General markerless motion capture system records motion trail by Microsoft Kinect. Although the commercial products that are based on Kinect can provide efficient markerless motion

*Corresponding author: Tainchi Lu
†Email: tclu@mail.ncyu.edu.tw

capturing in real time, but there are still some problems, such as unstable accuracy and unable to rotate human body substantially.

Tracking human motions from video had great improvement in the past few years. Method [1], [2], [3], [4] used human model which had activity limitation and two-dimensional image projected by three-dimensional model.

These methods needed to raise accuracy by machine learning. Due to machine learning strategies, they could not track motions which are not similar to the training data. Previous markerless motion tracking methods [5], [6] utilized image features for pose detection which the recorded target poses preliminarily during data training.

The human model that simplified by articulated skeleton can achieve easier and faster calculation. In [7], the articulated skeleton comprised of a rectangular torso and stick-limbs, but it led to more deviations. [8] proposed another articulated skeleton which consisted of a cuboid trunk. They used the model to estimate pose and analyze silhouette.

They also used the Gaussian Mixture Model (GMM) to get the color distribution of each pixel in the video. Furthermore, they strengthened the skeleton by the Gaussian mixture model. More precise model of a specific topic proposed by [9] suggested that tapered super-quadric synthesis and three-dimensional reconstruction could constitute the joint model. [10] employed the Gaussian blobs to define spherical volume, and attached spheres to the skeleton model.

In [11], the author defined a skeleton model manually, and used a three-dimensional mesh and skeleton association to calculate skinning weights. In this paper, we referred and adopted the articulated skeleton in [12], which defined the joints limitation and degree of freedom to avoid abnormal deforming of a mesh.

MARKERLESS HUMAN MOTION CAPTURE

Overview

The main concept of the proposed method is a markerless motion capture system, which can capture human motion without any specific costumes, wearable devices, or markers on users. Human actions will be recorded as serial images by common cameras at first. After employing several image processing procedures and combining with the articulated skeleton, precise human motions could be obtained for the entrance guard system. Figure 1 shows the framework of the proposed method in this paper. Due to the fact that the entrance guard system should be designed to respond in real time while users take some actions, the complexity of the proposed method needs to be reduced. First, we utilized the OTSU algorithm to convert input RGB images into binary images, and employed background subtraction techniques to separate foreground and background. Second, we segmented these binary images of foreground into numerous regions, and defined a sensor node for each region. Furthermore, we also segmented the articulated skeleton into parts, and matched the skeleton parts with the segmented foreground regions to obtain the joint position. Thus, the variation of the joint positions can be regarded as motion trajectories for the pose recognition of the entrance guard system. Because of the fact that human actions are complicated, especially the transitions between two actions, we predefined the motion decision tree to constrain poses and the transitions of poses for raising the accuracy of pose recognition. Eventually, we defined several poses and pose transition orders as constraints for the pose recognition. If users start to act with correct initial pose and transit poses with correct transition order, the proposed method will send an electrical signal to Arduino controlling system and make the door lock open or close.

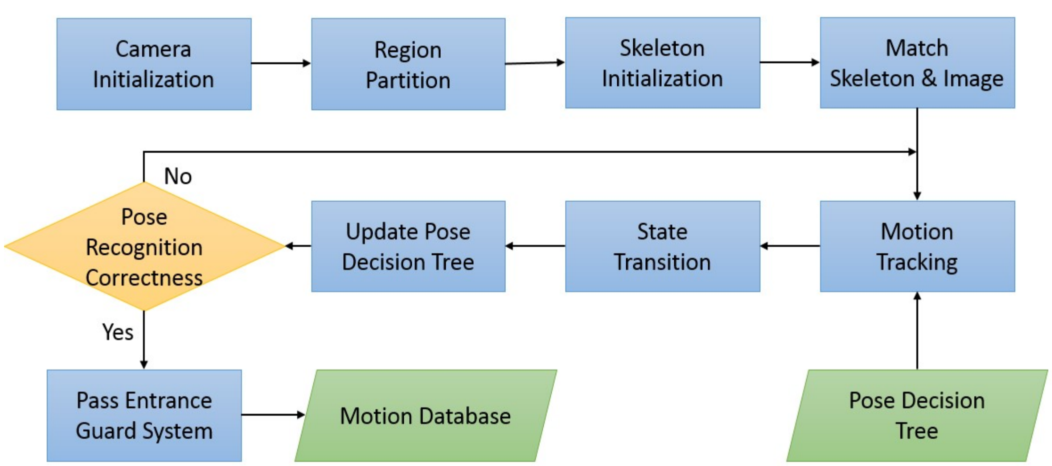

Fig. 1. The flowchart of proposed method

Camera Initialization

Our system adopted a webcam to record human motion video. We had to calibrate and modify webcam for suitable positions at the beginning. The color and luminance of video should be confirmed for avoiding deviations during recording. Then make an actor show T-pose on the center of an image and start to grab image stream by the webcam. In order to determine human position from images, we combined human feature with OTSU binary method [13] which consisted of two processes, erosion and dilation. As shown in Equation 1, we got the mean of first cluster (u_1). As shown in Equation 2, we got the mean of second cluster (u_2). As shown in Equation 3, we got the variance of first cluster (σ_1). As shown in Equation 4, we got the variance of second cluster (σ_2). As shown in Equation 5, we got the minimum sum of variance (σ). As shown in Equation 6, we got the threshold of binary (T).

$$u_1(T) = \frac{\sum_{i=0}^{T} h(i) * i}{\sum_{i=0}^{T} h(i)} \quad (1)$$

$$u_2(T) = \frac{\sum_{i=T+1}^{N} h(i) * i}{\sum_{i=T+1}^{N} h(i)} \quad (2)$$

$$\sigma_1(T) = \sum_{i=0}^{T} (u_1 - i)^2 h(i) \quad (3)$$

$$\sigma_2(T) = \sum_{i=T+1}^{N} (u_2 - i)^2 h(i) \quad (4)$$

$$\sigma(T) = \sigma_1(T) + \sigma_2(T) \quad (5)$$

$$T = argmin_T(\sigma(T)) \quad (6)$$

After applying several processes to the image stream, the features of a human, such as head, arms and legs can be obtained. Eventually we marked those features of human body parts to provide initial information of the articulated skeleton.

Region Partition

There is a lot of pixel information on an image [14]. Time complexity is high when calculating image information in real time [15]. In this paper, we divided an image into many uniform blocks that each block is a sensor node. First, we divided the recorded image into two parts and defined human endpoints. Second, we set left hand, right hand and head on the image. At last, we determined the image information of each sensor node by attaining the threshold or not for deciding human motion. As shown in Figure 2, the left red point represents right hand of the actor, the right red point represents left hand of the actor, the pink point represents head of the actor, and the yellow point on image center represents center of mass of human body.

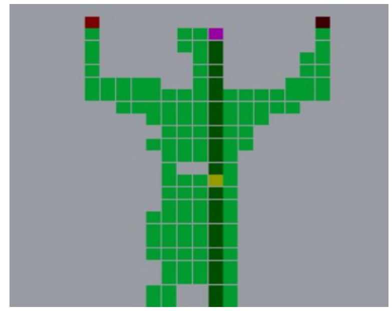

Fig. 2. Illustration of image partition

Human Skeleton Initialization

The proposed method referred to the articulated skeleton in [16]. We initialized a suitable ridged skeleton according to the silhouette and shape of the actor. The hierarchy of ridged skeleton refers to the definition in human structure. We defined the degree of freedom and limitation of rotation for each joint. The degree of freedom of each joint is shown in Figure 3. The limitation of rotation for each joint is shown in Table 1. We added Forward Kinematics (FK) as defined in Equation 7 and Equation 8. We also added Inverse Kinematics (IK) as defined in Equation 9 and Equation 10. We also defined the relationship of skeleton hierarchy to make the movement and rotation of each part of body conform the rules of human structure. As shown in Figure 4, we combined the human body in the image with the skeleton model after initializing for a follow-up controlling. In subsection 3.3, we will illustrate how we obtained the relative position of image and model.

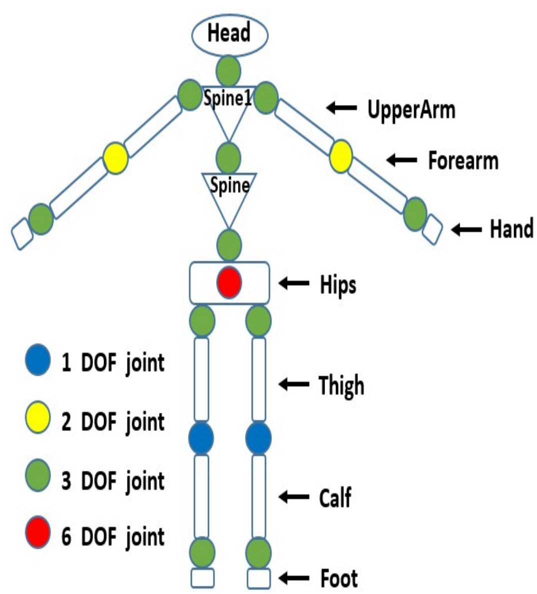

Fig. 3. The simplified skeleton model with its DOFs

TABLE 1
ROTATION LIMITATION OF HUMAN JOINTS

Bone / Axis	X Range	Y Range	Z Range
Head	-35 to 25	-45 to 45	-30 to 30
LUpperArm	-135 to 90	-105 to 100	-90 to 95
LForearm	-90 to 80	-145 to 0	locked
Lhand	-45 to 45	-25 to 35	-90 to 85
RUpperArm	-135 to 90	-100 to 105	-95 to 90
RForeArm	-90 to 80	0 to 145	Locked
RHand	-45 to 45	-35 to 25	-85 to 90
Spine 1	-45 to 20	-45 to 45	-30 to 30
Spine	-45 to 70	-45 to 45	-30 to 30
Hips	Locked	Locked	Locked
LThigh	-155 to 45	-85 to 105	-15 to 90
LCalf	0 to 150	Locked	Locked
LFoot	-30 to 60	-25 to 25	-75 to 15
RThigh	-155 to 45	-105 to 85	-90 to 15
RCalf	0 to 150	Locked	Locked
RFoot	-30 to 60	-25 to 25	-15 to 75

$$x = d_1 cos\theta_1 + d_2 cos(\theta_1 + \theta_2) \quad (7)$$

$$y = d_1 sin\theta_1 + d_2 sin(\theta_1 + \theta_2) \quad (8)$$

$$\theta_2 = cos^{-1}(\frac{x^2 + y^2 - {d_1}^2 - {d_2}^2}{2d_1 d_2}) \quad (9)$$

$$\theta_1 = tan^{-1} = (\frac{-x(d_2 sin\theta_2) + y(d_1 + d_2 cos\theta_2)}{y(d_2 sin\theta_2) + x(d_1 + d_2 cos\theta_2)}) \quad (10)$$

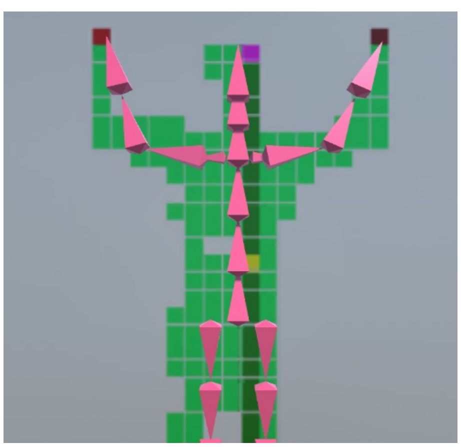

Fig. 4. Illustration of image correspondence

Human Pose Decision Tree

Human motions are linked by a lot of single postures. We determined different motions by recognizing a series of different continuous single-postures. We defined the pose decision tree as shown in Figure 5, and trained the probability model by the predetermined motion. Each motion starts from the T-pose (X1), and transfers to another posture according to the well-trained probability mode. Probability of transferring between two postures gets higher when the motion is more continuous and changing frequently.

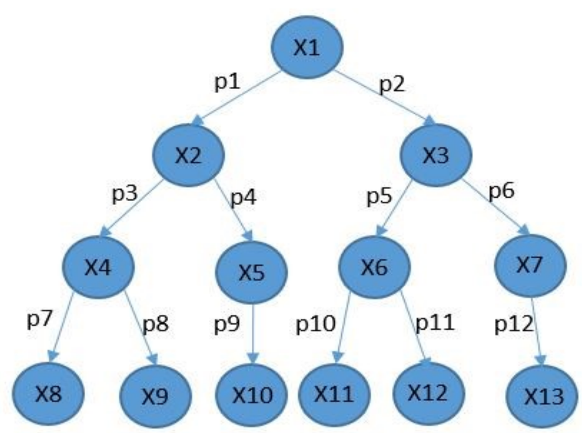

Fig. 5. Decision tree illustration

Joints Position Tracking

The proposed method starts tracking motion from key-pose. First we matched skeleton with head, endpoint of arms, and legs on the recorded image, then drove skeleton to track human motion by IK. As shown in Equation 11, we calculated Euclidean distance to detect the distance between key joint of continuous motions for removing unreasonable motion transition. We let the skeleton tracking in reasonable range that can reduce incidence of error determination. The difference between skeleton pose and pre-determination pose lower than threshold is regarded as a successful tracking, then pose state will be updated. When a motion trail passes determination of pose continuously with correct order, the entrance guard system will unlock the door. In contrast, the state is back to motion tracking step to restart tracking.

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + x(_2 - y_2)^2 + ...(x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (11)$$

Signal Sending

In this paper, users can define the entrance pose when performing initialization. Afterwards, motion tracking will be performed to determine that the acting pose conform to the entrance pose or not. If correct, the proposed method will change the signal in the signal library, and then connect with the Arduino Microcontroller System as shown in Figure 6. The microcontroller system will load information in signal library continuously. Once figuring out that the signal is changed and conformed the preconditions, microcontroller system will send the electronic signal to the entrance guard system to open the door lock. Signal library will be initialized after sending over to avoid sending electronic signal repeatedly.

Fig. 6. Entrance guard system and Arduino microcontroller system

Motion Exportation

The proposed method does not only apply to the entrance guard system, but also can save the captured human motion with the defined articulated skeleton in BVH format to build a human motion database for other purposes.

EXPERIMENT RESULTS

In this section, we demonstrate the proposed method for showing the implementation results with an Arduino-based door lock. The initialization flow is shown in Figure 7, we first took a snapshot of clean field as the background initialization (7a), and then an actor can enter the field to act the initial pose, T-Pose (7b).

Afterward we applied binary image processing (7c) and background subtraction (7d) to obtain the binary image of the foreground. With a 480 by 320 pixels binary image of the foreground, we divided it into 25 by 25 squares and determined the position for each part of human body (7e).

Eventually, we matched the defined skeleton to processed image and completed the initialization. So far, the implemented system is ready to capture or recognize motions of the actor.

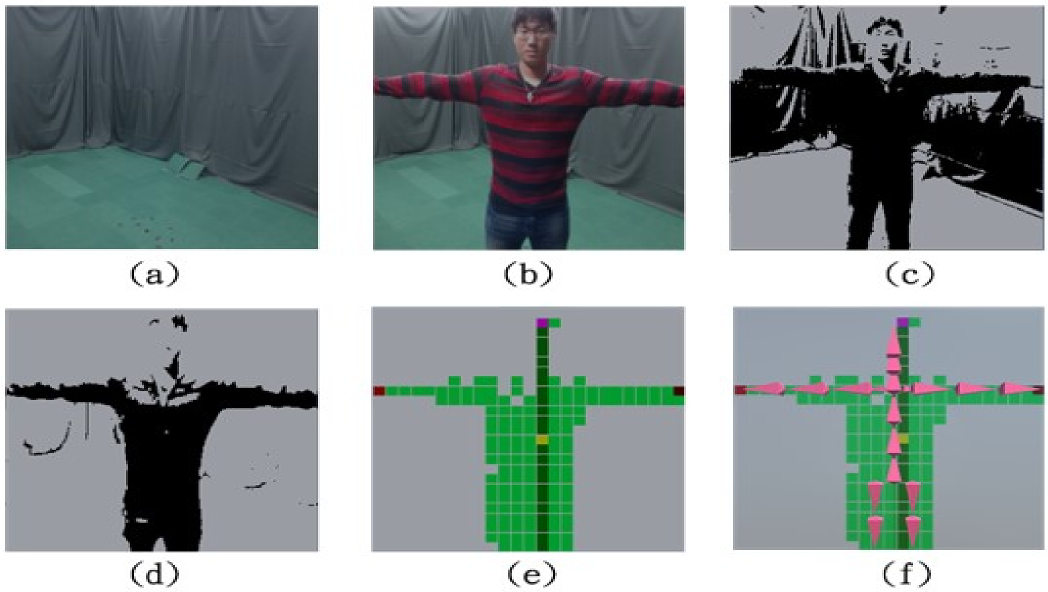

Fig. 7. Steps of image processing. (a) Original image in the testing environment, (b) Original image when the user enters the environment, (c) Binary image after OTSU, (d) Image after background subtraction, (e) Marking user key joints after image partition, (f) Image combines with the human skeleton

Figure 8(a) to 8(g) show seven types of defined motions for the entrance guard system correspondingly: T-pose, both hands up, both hands down, waving right hand, waving left hand, right hand up while left hand down, and left hand up while right hand down. To unlock the door with the proposed system, users need to act from T-pose to other assigned poses. Table 2 shows the accuracy of two users who try to unlock the entrance guard system. It can be clearly found that the accuracy of simple motion like 8(b) and 8(c) was higher than others, which tend to misjudge and decrease the accuracy.

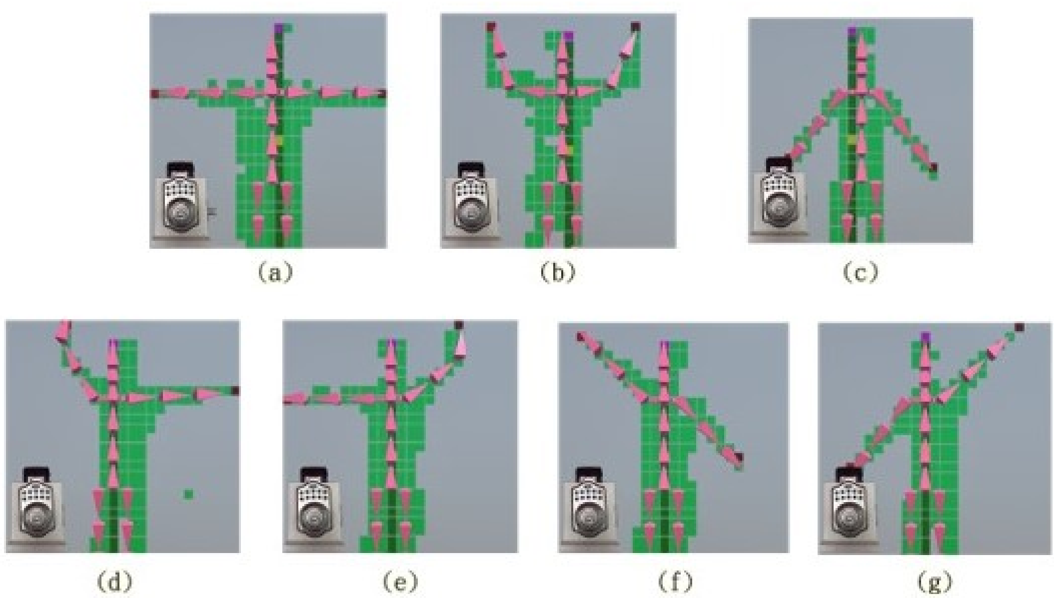

Fig. 8. Motion capture combines with the entrance guard system

TABLE 2
MOTION RECONIZE RESULTS

Motion	Tester 1			Tester 2			Failure Causes
	Total test number	Number of successes	Success rate (%)	Total test number	Number of successes	Success rate (%)	
Hands Up	100	88	88%	100	100	100%	No
Hands down	100	100	100%	100	80	80%	No
Waving left hand	100	72	72%	100	44	44%	Higher threadshold
Waving right hand	100	71	71%	100	58	58%	Higher threadshold
Tilting both hands (1)	100	72	72%	100	64	64%	Misunderstanding with hands down
Tilting both hands (2)	100	75	75%	100	74	74%	Misunderstanding with hands down

CONCLUSION AND RECOMMENDATIONS

In this paper, we have successfully taken advantage of optical webcams to capture human motions for applying to an unsophisticated entrance guard system. The proposed system can be worked well under the good illuminance and adequate motion speed; on the contrary, the poor illuminance has led to incomplete image background subtraction with incorrect motion recognition. Accordingly, we have adopted an articulated skeleton with joint constraints and pose decision tree in order to increase the recognition rate.

However, it would suffer from failed motion recognition in the case of the high-speed human motions and self-occlusion of actors. In the near future, we tend to increase more than one webcam to acquire depth information to tackle the above problem.

Declaration of Conflicting Interests

No conflicts of interest.

Acknowledgments

This work was supported by grants from the Ministry of Science and Technology at Taiwan, R.O.C., under the project grant numbers MOST 104-2221-E-415-025.

REFERENCES

[1] S. Choppin. (n.d.). *Engineering sport blog* [online]. Available: https://goo.gl/MKke4k

[2] R. Li, T. P. Tian, S. Sclaroff and M. H. Yang, "3d human motion tracking with a coordinated mixture of factor analyzers," *International Journal of Computer Vision,* vol. 87, no. 1, pp. 170-190, 2010.

[3] C. S. Lee and A. Elgammal, "Coupled visual and kinematic manifold models for tracking," *International Journal of Com- puter Vision,* vol. 87, no. 1, pp. 118-139, 2010.

[4] L. Bo and C. Sminchisescu, "Twin gaussian processes for structured prediction," *International Journal of Computer Vision,* vol. 87, no. 1, pp. 28-52, 2010.

[5] A. Bissacco, M. H. Yang and S. Soatto, "Fast human pose estimation using appearance and motion via multi-dimensional boosting regression," in *IEEE Conference on Computer Vision and Pattern Recognition,* 2007, pp. 1-8.

[6] M. Van Den Bergh, E. Koller-Meier and L. Van Gool, "Real-time body pose recognition using 2D or 3D haarlets," *Inter- national Journal of Computer Vision,* vol. 83, no. 1, pp. 72-84, 2009.

[7] D. Moschini and A. Fusiello, "Tracking human motion with multiple cameras using an articulated model," in *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications,* 2009, pp. 1-12.

[8] P. Kohli, J. Rihan, M. Bray and P. H. Torr, " Simultaneous segmentation and pose estimation of humans using dynamic graph cuts," *International Journal of Computer Vision,* vol. 79, no. 3, pp. 285-298, 2008.

[9] A. Sundaresan and R. Chellappa, "Model driven segmentation of articulating humans in Laplacian Eigenspace," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, no. 10, pp. 1771-1785, 2008.

[10] F. Caillette, A. Galata and T. Howard, "Real-time 3-D human body tracking using learnt models of behaviour," *Computer Vision and Image Understanding,* vol. 109, no. 2, pp. 112-125,2008.

[11] J. Gall, C. Stoll, E. De Aguiar, C. Theobalt, B. Rosenhahn and H. P. Seidel, "Motion capture using joint skeleton tracking and surface estimation," in *IEEE Conference on Computer Vision and Pattern Recognition,* 2009, pp. 1746-1753.

[12] K. Ogawara, X. Li and K. Ikeuchi, "Marker-less human motion estimation using articulated deformable model," in *Pr- oceedings IEEE International Conference on Robotics and Automation*, 2007, pp. 46-51.

[13] D. H. AlSaeed, A. Bouridane, A. ElZaart and R. Sammouda, “Two modified Otsu image segmentation methods based on Lognormal and Gamma distribution models,” in *International Conference on Information Technology and e-Services (ICITeS)*, 2012, pp. 1-5.

[14] B. K. Kumar, Shreyamsha. “Image fusion based on pixel significance using cross bilateral filter,” *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1193-1204, 2015.

[15] S. Sanchez, R. Rui, S. Leonel, and P. Antonio. “Real-time implementation of remotely sensed hyperspectral image unmixing on GPUs,” *Journal of Real-Time Image Processing* vol. 10, no. 3, pp. 469-483, 2015.

[16] L. Ballan and G. M. Cortelazzo, “Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes,” in *Proceedings of 3DPVT'08 - the Fourth International Symposium on 3D Data Processing, Visualization and Transmission*, Atlanta, GA, 2008.

— This article does not have any appendix. —