

## An Interactive Aerobic Training System Using Vision and Multimedia Technologies

Thanarat H. Chalidabhongse, and Alongkot Noichaiboon

Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

(Tel : +66-2-737-2551; E-mail: thanarat@it.kmitl.ac.th)

**Abstract:** We describe the development of an interactive aerobic training system using vision-based motion capture and multimedia technology. Unlike the traditional one-way aerobic training on TV, the proposed system allows the virtual trainer to observe and interact with the user in real-time. The system is composed of a web camera connected to a PC watching the user moves. First, the animated character on the screen makes a move, and then instructs the user to follow its movement. The system applies a robust statistical background subtraction method to extract a silhouette of the moving user from the captured video. Subsequently, principal body parts of the extracted silhouette are located using model-based approach. The motion of these body parts is then analyzed and compared with the motion of the animated character. The system provides audio feedback to the user according to the result of the motion comparison. All the animation and video processing run in real-time on a PC-based system with consumer-type camera. This proposed system is a good example of applying vision algorithms and multimedia technology for intelligent interactive home entertainment systems.

**Keywords:** Interactive system, motion capture, multimedia, real-time vision

### 1. INTRODUCTION

The ability to detect, track, and recognize people in action is a fundamental and crucial task in many vision systems. The task is very challenging due to the dynamics of the articulated human body, partial occlusion and non-rigid clothing. The problem can be divided into several sub-problems ranging from low-level image processing to high-level modeling and reasoning problems. The analysis and recognition of human motion has a number of promising applications such as security and surveillance systems. A vision system that has ability to detect, track, and recognize people in action can provide a cost-effective method for automatically extracting and identifying information from the video stream. In perceptual user interfaces (PUI), the idea is to let the people interact with machines in the same natural fashion as they interact with one another. Another area of application might be sign-language translation and gesture driven control of appliance for assistance disabled people. In the entertainment industry, motion capture systems are used to detect human movements and transfer them to 3D graphical models used in animation for movies, games, advertises, etc. Also, they can be applied to medical gait analysis, engineering design, ergonomics, virtual reality, and sport simulations such as golf swing analyzers and figure skating interpreters.

This paper presents a development of an interactive aerobic training system using low-cost vision-based motion capture that can detect and track unconstrained whole-body human motion without the use of markers. The original inspiration of our work was the work done by Davis and Bobick [1]. Their system, called PAT, provided virtual Personal Aerobic Trainer to the user. The system allowed the user to create and personalize an aerobic session to meet the user's needs and desires. However, their system was running on high cost hardware setup which is composed of SGI R10000 O2 machines equipped with a professional CCD camera. Our goal is to build a similar system using a home PC with a consumer-level web camera. The algorithm used is also different. Rather than computing motion templates [2], we employ a robust and fast background subtraction (previously developed by one of the authors [3]) to segment the silhouette of the user. We then analyze the silhouette to locate the positions of principal body parts such as head, hands, and feet using the cardboard model [4]. The cardboard model is simple

but work well for this application since the user is mostly in the upright posture. By utilizing the temporal information of these body parts, we obtain the trajectory of each part. They are then compared with the motion pattern of the animated character and provide an appropriate feedback to the user.

This report is organized as follows: the next Section describes an overview of the interactive aerobic training system configuration and program components. The details of the vision-based motion capture are presented in Section 3. Section 4 presents the results. The conclusion of the paper is in Section 5.

### 2. SYSTEM OVERVIEW

The interactive aerobic training system using vision-based motion capture has the system configuration shown in Fig.1. The PC is equipped with a web camera and runs an interactive comic animation written in ActiveX Control and Shockwave Flash Objects. The user stays in front of the camera trying to imitate the motion of the comic aerobic trainer. The system watches the user and response interactively according to the user's motion. If the user correctly follows the trainer's movement, positive comments are provided verbally and the trainer moves to the next pattern. Otherwise, the trainer provides negative comments and repeats the pattern until the user gets it done correctly or the pattern is time-out. After finishing all the workouts, the system ends the session by saying "thank you". The system is not designed only for adult aerobic training; it could be also applied for children entertainment systems such an interactive "Simon Says" game.

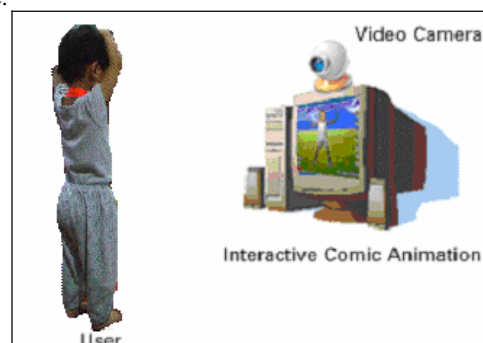


Fig. 1 The system configuration.

In analyzing the user motion, the system uses background subtraction technique to segment the user's silhouette from the environment. Thus, when starting the system, the user will be greeted and asked to initialize the background model by letting the web camera grab a video of empty scene for a few seconds. After background initialization is done, the user can now start the exercise. The system starts with the first exercise and music. Fig. 2 is the system's component diagram that illustrates the control and communication of the system while the aerobic training session is performed. First, the input video from the web camera is subtracted from the background model yielding the extracted silhouette of the user. The motion analysis is then performed on the silhouette and is compared with the exercise's motion ground truth. If the user imitates the comic's motion correctly, the positive audio feedback is provided and the system will move on to the next workout by changing the animation on the screen. Otherwise, the negative audio feedback is provided and the system repeats the pattern until the user gets it right or the pattern is time-out.

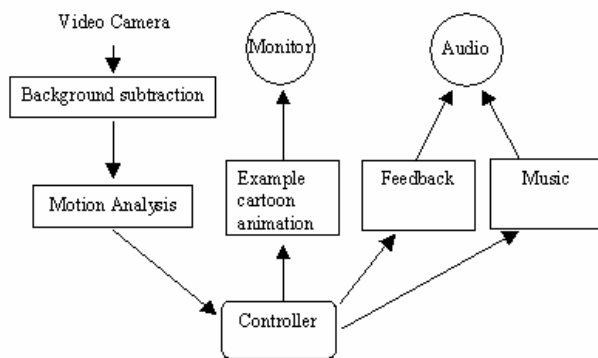


Fig. 2 Program components and media flows

The controller program is written in C/C++, MS Vision SDK 1.2 [9], and uses Video for Windows (VFW) library for video stream grabbing. The multimedia processing such as animation and audio synthesis used in interacting with the user, is written using ActiveX Control and Shockwave Flash Objects.

### 3. HUMAN MOTION ANALYSIS

In this Section, we describe a real-time vision-based system for detecting and tracking human motion. It employs a feature-based approach in recovery of articulated body models. Each input images is analyzed to identify the locations of principal body parts such as torso, head, hands, and feet.

First, the system performs a statistical-based background subtraction presented in [3] and [6]. The idea of background subtraction is to subtract the current image from a reference image, which is acquired from a static background during a period of time. The subtraction leaves only non-stationary or new objects, which include the objects' entire silhouette region (see Fig.3). The algorithm is a robust and efficiently computed method that is able to cope with local illumination changes such as shadows and highlights, as well as, some global illumination changes. The algorithm is based on a computational color model which separates the brightness from the chromaticity component.

After background subtraction is done, the system then performs silhouette analysis and template matching to locate the positions of principal body parts. To initially locate the body parts, we employ two approaches; cardboard model and body contour extremity detection. A geometric cardboard

human model of a person in a standard upright pose is used to model the shape of the human body and to locate the body parts (see Fig.4). The height of the bounding box of the object is taken as a height of the cardboard model. The lengths of the initial bounding boxes of the head, torso, and legs are calculated as  $1/5$ ,  $1/2$ , and  $1/2$  of the length of bounding box of the object, respectively. As opposed to using the cardboard model, using body contour extremities in locating body parts does not require the upright position. The person can be in any generic posture. Moreover, it works even when the entire silhouette is not in the scene or not detected. In addition, it is very likely that the principal body parts we are interested in usually lie on the body contour.

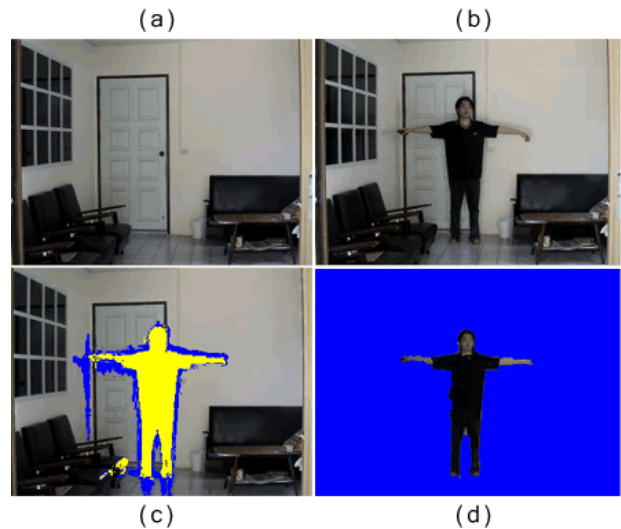


Fig.3 The user's entire silhouette region is segmented after background subtraction

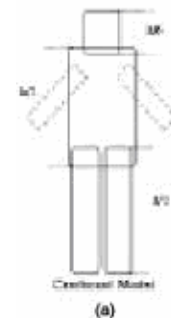


Fig.4 Cardboard human model

To identify extremities on the contour, we considered four methods:

1. Recursive convex hull: as employed in [5]
2. Star shape distance: defined as the shortest distance from the silhouette centroid to the contour pixel
3. Path length: defined as the shortest path length from the silhouette centroid to the contour pixel
4. Curvature: defined as the curvature or angle of the contour pixel.

We found that the first three methods have disadvantage in the case of multiple people occluding each other. The curvature computation is local and can be carried out in the presence of occluding multiple people. To compute the curvature, first the median angle of each contour pixel is computed and threshold. Then, nonmaxima suppression is applied. Fig.5

shows the result of the extremities detection. After predicting the locations of the head and hands using the cardboard model and extremities, their positions are verified and refined using dynamic template matching. Multiple cues, such as distance, color, and shape, which define feature appearance are used in matching.

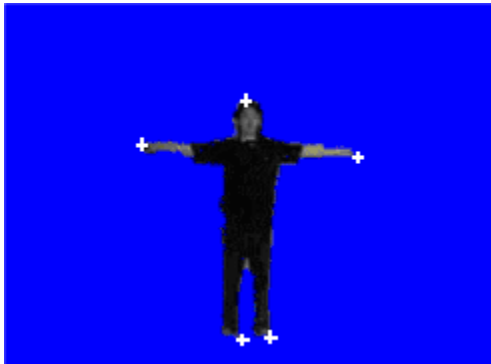


Fig.5 Extremities detection using curvature

#### 4. RESULTS

Fig.6~7 show the screen snapshots of the system while running. At first, the system requested the user to initialize the system by constructing the background model. The image shown in the lower left corner of the screen is the modeled background image (see Fig.6). After initialization, the user can now play with the system by copying the motion pattern of the comic aerobic trainer which is appeared on the upper right corner of the screen. The image shown in the upper left corner of the screen is the input video. The system will then detect the principal body parts such as head and hands and use them to analyze the motion of the user. The detected user motion is then compared with the comic motion pattern (see Fig.7). The system will then response to the user based on his performance.



Fig.6 Screen snapshot after background initialization

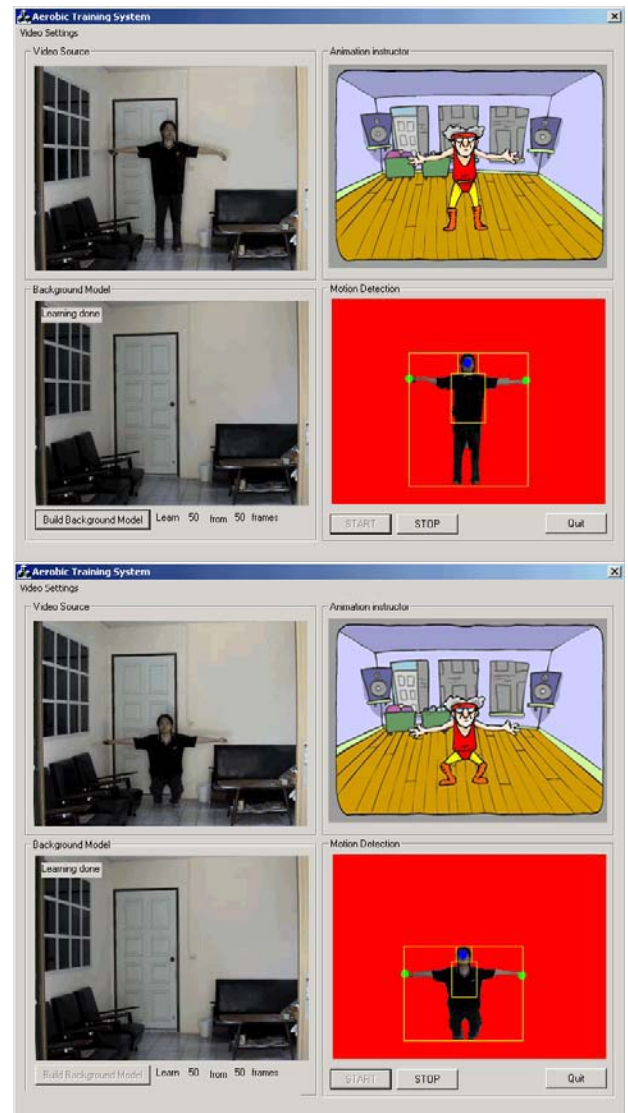


Fig.7 Screen snapshot while the system is running

#### 5. CONCLUSION AND FUTURE WORKS

We presented a development of an interactive aerobic training system using vision-based motion capture. The system applies the background subtraction technique to extract the user's silhouette from the input video. Subsequently, the user's gesture is analyzed by computing motions of principal body-parts. The simple cardboard model is used for body part localization. The recovered user motion is then compared with the motion ground truth of the cartoon animation. The system responds interactively to the user according to how well the user performs. Our system runs in real-time on home PC with a typical web cam. The system is not only for adult aerobic training; it also could be applied for children interactive games such as the classical "Simon Says".

In the current version, only a few principal body parts are used. Thus, the motion support is quite limited. We are working on more sophisticate approach in body part detection and tracking. Spatio-Temporal motion modeling is also under investigated. Another limitation of the current system is that it is developed under assumption of single user. The system does not work with multiple persons simultaneously interact with the system. Regarding this concern, we are investigating approaches proposed by [7] and [8].

## REFERENCES

- [1] J. Davis, and A. Bobick, "Virtual PAT: a virtual personal aerobic trainer," *Proc. of Workshop on Perceptual User Interfaces*, pp. 13-18, 1998.
- [2] J. Davis, and A. Bobick, "The representation and recognition of action using temporal templates," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.928-934, 1997.
- [3] T. Horprasert, D. Harwood, and L.S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," *Proc. of IEEE ICCV'99 Frame-rate Workshop*, 1999.
- [4] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: who? when? where? what? a real time system for detection and tracking people," *Proc. of Int'l Conf. on Face and Gesture Recognition*, 1998.
- [5] I. Haritaoglu, D. Harwood, and L.S. Davis, "Ghost: a human body part labeling system," *Proc. of 14<sup>th</sup> Int'l Conf. on Pattern Recognition*, 1998.
- [6] T. Horprasert, D. Harwood, and L.S. Davis, "A robust background subtraction and shadow detection," *Proc. of Asian Conf. on Computer Vision*, 2000.
- [7] A. Elgammal, and L.S. Davis, "Probabilistic framework for segmenting people under occlusion," *Proc. of the 8<sup>th</sup> IEEE Int'l Conf. on Computer Vision*, 2001.
- [8] E. Polat, M. Yeasin, and R. Sharma, "A tracking framework for collaborative human computer interaction," *Proc. of the 4<sup>th</sup> IEEE Int'l Conf. on Multimodal Interfaces*, 2002.
- [9] Vision Technology Group, "The Microsoft Vision SDK Version 1.2," *Microsoft Research*, 2000. [Online] Available: <http://research.microsoft.com/projects/VisSDK>