

# HAND GESTURE TRACKING FOR WEARABLE COMPUTING SYSTEMS

Xiujuan Chai, Kongqiao Wang, Luosi Wei and Hao Wang  
*System Research Center, Nokia Research Center, Hepingli Dongjie, Beijing, China*

*School of Automation Science and Electrical Engineering, Beihang University, Beijing, China*

**Keywords:** Wearable computing, temporal differencing, motion region, skin detection, colour histogram.

**Abstract:** Wearable computing is a hot research field in recent years. For the important role in wearable computing systems, hand gesture tracking attracts many researchers' interests. This paper proposes a simple but efficient temporal differencing based hand motion tracking scheme which is used to build an augmented drumming system. In our method, the accurate motion information is gotten by a fine-coarse-fine strategy. Once getting the motion region candidates, a skin detector based on skin colour histogram is used to determine which region is our concerned hand. In the tracking procedure, motion direction constraint is also adopted in order to get a robust result. Different with the traditional skin detection for the whole image frame, combining with the motion region detection, the hand detection is no longer effected by the skin-like background. Experimental results show that our presented hand gesture tracking is robust and fast. We also adopt it into an augmented drumming system to show the good performance and powerful potential of our method in wearable computing systems.

## 1 INTRODUCTION

Wearable computing facilitates a new form of human-computer interaction(HCI). Hand detection and tracking is widely exploited for the important role in wearable computing systems for the potential applications, including the command control, games, text input and many other aspects (Manresa, 2005)(Buades, 2004)(MacCormick, 2000).

So far, the state-of-the-art tracking strategy can achieve high accuracy under restrict environment. However, when confronted with complicated background and irregular motion, the tracking performance will be decreased dramatically. Therefore, researchers' should pay more attention to the robust hand tracking under unrestricted conditions. Over the study of these years, there are many literatures focus on hand tracking and analysis. Roughly, the tracking methods can be divided into two categories: appearance-based method and the model-based method.

In general, the model-based method aims to find the accurate mapping from the 2D image to the 3D configuration model of hand (Wu, 2001) (Lu, 2003) (Chang, 2005). Although such tracking can achieve good performance even for the detailed finger

motion, the computations are always time-consuming for the iterative fitting to the elaborate 3D hand model.

While the appearance-based method aims to get the correspondence between sequential video frames based on the image features. Here, the image features include not only the color, edge, position, but also the transformed features, such as the histogram feature, wave-let feature, high level semantic feature etc. The time cost changes with the selected features and in general, it will be less than the model-based method. (Shamaie, 2003) proposes a Kalman filtering-based dynamic model to deal with the bimanual movements. (Shan, 2007) proposes mean shift embedded particle filtering to improve the sampling efficiency. (Bowden, 2002) adopts eigenspace approaches to model contour and appearance feature spaces. And there are also some papers focusing on much simpler features (Martin, 1998) (Huang, 2002). It is naturally that if the selected feature is simpler, then the time cost is lower.

Considering the efficiency, we also adopt appearance-based method to tackle hand tracking problem. First, a fine-coarse-fine strategy is performed to realize robust motion region finding.

Then skin color is used as a constraint for the determination of the hand region. Simultaneously, the moving direction of each motion rectangle is also computed, which is used to eliminate some unmeaningful motions and erect the correspondence of moving targets. To show the potential to wearable computing, the tracking is conducted in an augmented drumming system as an instance and shows good performance.

## 2 HAND MOTION TRACKING STRATEGY

Our scheme for hand motion tracking mainly includes three modules: motion region detection, skin detection and the final motion vector computing.

### 2.1 Motion Region Detection

In this part, a fine-coarse-fine strategy is adopted in the temporal differencing to achieve a good de-noising. For two consecutive video frames, i.e. . the differencing is operated between the current frame  $I_n$  and the previous frame  $I_{n-1}$ . With an experimental threshold, we get the binary difference image  $D_n$  according to Eq.(1), which usually contains many noise points for the illumination effect, as shown in Figure.1 (b).

$$D_n(x, y) = \begin{cases} 0, & \text{motion pixel} \\ 1, & \text{else(background)} \end{cases} \quad (1)$$

To further eliminate these noises, a fine-coarse-fine strategy is using here. By performing the de-noising operation in these transformable levels, the obtained differencing image is very clear.

Fine level: Performing the de-noising operation (erosion and dilation) to  $D_n$ , and we can get a binary image  $D'_n$  with less noise as shown in Figure. 1(c).

Coarse level: Performing down-sampling to  $D'_n$ , then doing the de-noising operation to get  $(D'_n)_s$ .

Fine level: Refining  $D'_n$  from  $(D'_n)_s$ :

$$D'_n(x, y) = \begin{cases} 0, & D_n(x, y) = 0, \\ & \text{and } (D'_n)_s(x', y') = 0 \\ D_n(x, y), & \text{else} \end{cases} \quad (2)$$

here,  $x' = x/w, y' = y/h$ , w and h are the down-sampling steps along x and y directions respectively.

### 2.2 Skin Detection

In the motion region detection stage, we can detect all the dominating motions, which are caused by hand movement, body movement, or the movement of anything else in the scene. It is obvious that some motions are meaningless for us. Therefore, how to move such motions from all detected candidates is important and here we exploit a skin detector based on color histogram.

In the model training procedure, first collecte many hand moving rectangles. Then by using color clustering, the skin pixels in these rectangles are determined to erect the statistical color histogram  $H$ , which could be defined as follows:

$$H_i = \sum_{(x,y)} N\{f(x, y) = i\} / M, (i = 0, 1, \dots, K). \quad (3)$$

In Eq.(3),  $f(x, y) = I^R(x, y) * bin + I^G(x, y) + I^B(x, y) / bin$ , and  $K = 255 * bin + 255 + 255 / bin$ , with  $bin = 16$ . Here  $M$  is the total number of the skin pixels and  $N$  is defined as:

$$N\{f\} = \begin{cases} 1, & f \text{ is true} \\ 0, & f \text{ is false} \end{cases} \quad (4)$$

Similarly, we can get a statistical non-skin color histogram  $\bar{H}$ . Therefore, one pixel is determined to be skin point if it satisfies the following two terms:

- (1)  $H_{f(x,y)} > 0$ ,
- (2)  $H_{f(x,y)} > \alpha \cdot \bar{H}_{f(x,y)}$ ,  $0 < \alpha < 1$ .

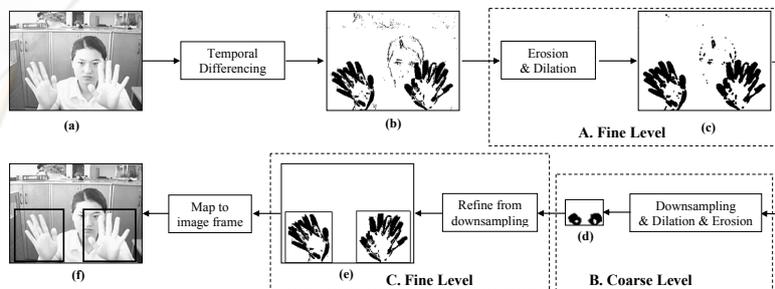


Figure 1: Flowchart of the motion region finding.

Figure 2 gives two examples for skin color detection. The results clearly show that the skin detection strategy can extract the hand region from so difficult background.



Figure 2: Examples for skin color detection.

With this skin detection procedure, the percentage of the skin pixels to the whole pixels in the moving rectangle can be computed. Through a comparison with a predefined threshold, we can eliminate the moving regions caused by non-skin color objects, as given in Figure 3.

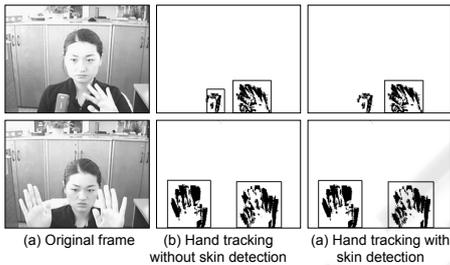


Figure 3: The comparison of the hand motion tracking with and without skin detection constraint.

### 2.3 Motion Vector Computing

To complete the tracking task, we try to erect the correspondence between these motion rectangles by using the moving direction information. Considering the simplicity and effectiveness, block matching algorithm is adopted in this paper. In our implementation, we aim to get the mean moving direction of each motion rectangle based on the moving vector of the whole frame. Figure.4 gives an example for the motion vector field and moving directions for hand regions.

## 3 AN INSTANCE - AUGMENTED DRUMMING SYSTEM

Augmented reality (AR) is a hot field of wearable computing research which deals with the real world

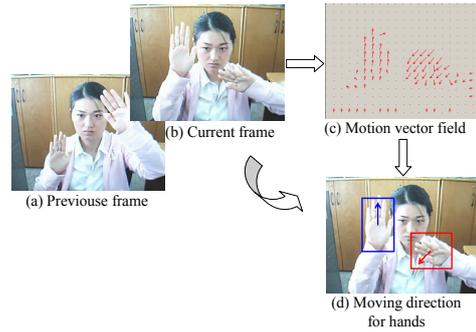


Figure 4: An example for the motion vector field and hand motion region direction.

and computer generated data. In this paper, we present a simple augmented drumming system as an instance. Through the hand motion tracking, the virtual drumming sound is generated and the virtual drum is displayed with the real person and background in the screen. Some examples are given in Figure. 5 of Section 4.

Simply speaking, if the hand motion region arrived at the virtual drum surface location with a downwards moving direction, then it is determined as a valid drum activity and the system is triggered to sound. Another important parameter is the volume of the sound and here it is determined through the location information according to Eq.(5):

$$v = V \cdot (\|C_n - C_{n-2}\|_y / h), \tag{5}$$

where,  $V$  is the predefined max value of volume.  $\|C_n - C_{n-2}\|_y$  is the vertical distance between the two centers  $C_n$  and  $C_{n-2}$  of  $n$ -th and  $(n-2)$ -th frames,  $h$  is the height of the image frame.

## 4 EXPERIMENTS

### 4.1 Experiment on Hand Motion Tracking

The qualitative results for hand motion tracking have been given in the Figure 2 and Figure 3 shown.

To evaluate the performance of the hand motion tracking, we adopt such a measurement principle:

$$r = n_h / n_d, \tag{6}$$

here,  $n_h$  and  $n_d$  are the numbers of hand motion region and totally detected motion region respectively.

In our experiment, 10 short videos are recorded, totally 2055 frames. We compare the hand motion

detection rates between with and without skin detection constraint as listed in the table 1. In some of our test videos, we add some motions caused by other objects, and the experimental results show that this kind of motion can be effectively eliminated and thus the hand motion detection rate with skin detection constraint improves remarkably. Importing the skin detection modular also causes the increasing of the time cost, as shown in Table 1. Fortunately, the increasing can be accepted for general HCI tasks and it can be compensated by high performance computers.

Table 1: The comparison results of the hand motion detections with skin model and without skin model conditions.

Evaluation \ Hand motion detection	skin detection constraint	
	Without	With
Detection rate	76.4%	93.04%
Time cost (ms/frame)	2.7	6.2

## 4.2 Experiment on Augmented Drumming

In this wearable computing instance, the aim of our hand motion tracking is to monitor an augmented drumming system. Assuming a virtual drum location first, through the hand tracking results, the rataplan activity can be determined and the drumbeat is played. By the integration of the hand motion tracking and motion vector computation, the augmented drumming system works well. Here, some examples are given on Figure. 5, which show the good performance of the interactive system.

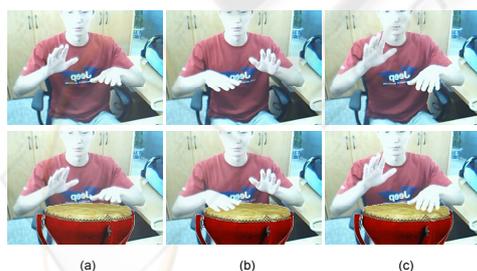


Figure 5: The examples of the augmented drumming system.

## 5 CONCLUSIONS

This paper proposes a robust hand gesture tracking

strategy. As an important visual analysis task for wearable computing system, it is also used for an augmented drumming system. In our motion detection method, a fine-coarse-fine strategy is adopted to eliminate lots of noise and get clear results. Based on the extracted motion rectangles, the skin detection using color histogram feature is performed on them to determine the hand region. The simple training procedure makes the distinction between hand pixels and the skin-like background become very easy and effective. Integrating the motion vector computing, our proposed hand gesture tracking strategy shows good performance in the augmented drumming system.

## REFERENCES

- Martin, J., Devin, V, Crowley, J. L. 1998. Active Hand Tracking. *Proc. of the 3rd. Int. Conf. on Face and Gesture Recognition*. pp. 573-578.
- Buades, J. M., Perales, F. J. , Varona, J., 2006. Real Time Segmentation and Tracking of Face and Hands in VR Application. *Third Int. Workshop on Articulated Motion and Deformable Objects*. pp. 259-268.
- Manresa, C., Varona, J., Mas, R., Perales, F. J., 2005. Hand Tracking and Gesture Recognition for Human-Computer Interaction. *Electronic Letters on Computer Vision and Image Analysis*. 5(3): 96-104.
- Lu, S., Metaxas, D., Samaras, D., Oliensis, J.2003. Using Multiple Cues for Hand Tracking and Model Refinement. *CVPR*, 2(443-450).
- Chang, W., Chen, C., Hung, Y., 2005. Appearance-Guided Particle Filtering for Articulated Hand Tracking. *CVPR*. pp.235-242.
- Wu, Y., Lin, J., Huang, T., 2001. Capturing Natural Hand Articulation. *ICCV*. pp. 426-432.
- Shamaie, A., Sutherland, A., 2003. A Dynamic Model for Real-Time Tracking of Hands in Bimanual Movements. *Gesture Workshop*. pp.172-179.
- Rosales, R., Sclaroff, S., 2006. Combining Generative and Discriminative Models in a Framework for Articulated Pose Estimation. *IJCV*. 67(3): 251-276.
- Stenger, B. 2005. Model-Based hand Tracking Using a Hierarchical Bayesian Filter. *Ph.D. Thesis*. University of Cambridge, St. John's College.
- MacCormick, J., Isard, M., 2000. Partitioned Sampling, Articulated Objects, and Interface-Quality Hand Tracking. *ECCV*. pp.3-19.
- Shan, C., Tan, T. Wei, Y. 2007. Real-Time Hand Tracking using a Mean Shift Embedded Particle Filter. *Pattern Recognition*. 40(7):1958-1970.
- Huang, Y., Huang, T., Niemann, H., 2002. Two-Handed Gesture Tracking Incorporating Template Warping With Static Segmentation. *FGR*. pp.275-280.
- Bowden, R., Sarhadi, M.,2002. A Non-linear Model of Shape and Motion for Tracking Finger Spelt American Sign Language, *Image Vision Comput*. 20: 597-607.