Hand Tracking by Extending Distance Transform and Hand Model in Real-Time[1,2]

C. Kerdvibulvech

National Institute of Development Administration (NIDA), 118 SeriThai Rd., Klong-chan, Bangkapi, Bangkok 10240, Thailand

e-mail: chutisant.ker@nida.ac.th

Abstract—Tracking a human's hand is not a trivial task. This paper contributes a new approach for hand tracking based on distance transform (DT) and edge points in real-time. In the beginning, we create a hand model geometrically in three dimensions. It is done by utilizing shortened quadrics. After that, the degrees of freedom, shortly called as DOF, for every joint angle correspond to each DOF to use in the later process. The edge likelihood is used for the feature extraction. A Bayesian classifier is utilized adaptively and accurately for the silhouette likelihood. For this reason, it is to cope greatly with any environmental changes visibly. By using these techniques, this method can be performed in real-time. Experimental results are provided.

DOI: 10.1134/S1054661815030098

INTRODUCTION

Although object tracking algorithms have been presented [1–5], this research problem is very challenging in computer vision-based work area, especially for hand tracking. Here, we contribute a model-based method for hand tracking. It is done by recovering of hand motion in real-time. A first challenge for tracking the hand is about the background. The background is non-uniform (e.g. natural scene). Thus, it is more challenging for background segmentation. Moreover, human hand is composed of many degrees of freedom (DOF), so that it is very difficult for tracing and tracking the movement of all DOF of the hand. In addition, the execution and computational time is very important for the system. If the system cannot run in real-time, it is very difficult to apply it in real-life applications. Furthermore, the issue of preprocess is critical. It must require to create many templates during preprocess, which is not easy to be done.

Cameron et al. [6] proposed a system to track a user's hand using in virtual reality application. However, this system used a white background without any object. Raheja et al. presented a system for fingertips detection and centers of palms detection distinctly using Microsoft KINECT [7]. Their aim was only to detect the fingertips using the depth vector. The related-work is presented by Pan et al. [8]. They built a real-time tracking system using averages of the velocity weighted models, integrated with the segmentation of skin color. Park et al. [9] presented an algorithm for tracking and tracing hand utilizing a three-dimensional depth sensor. A conventional filter was also employed in this system. A similar work [10] for hand gesture interaction was presented by Hendrik and Beyerer. [11]'s work by Asaari and Suandi also discussed a vision-based hand tracking system by combining Adaptive Kalman Filter (AKF) and appearance model. However, all of them do not aim the same as ours. In other words, their goals are not to utilize model-based approach to recognize the human hand which does not include the full details of hand, such as edges and features. Thus, these systems do not focus on full hand tracking.

The research that is quite similar to our work is [12] by Gorce et al. Their work aimed for 3D hand tracking from monocular video. They used minimization of an objective function. A similar detection algorithm which focuses on detecting the image regions that represent the scene using context-aware saliency [13] was presented by Goferman et al. A similar 3D model-based hand tracking was proposed in [14] by extending particle filter. However, most of these mentioned systems cannot run in real-time. For example, in [12] their system takes around 40 seconds per frame which is not real-time. Also in [14], by using the particle filter, the number of particles used climbs dramatically with the dimension and also the processing time. Hence, it cannot process in real-time. In this paper, we aim to compute it in real-time.

Other similar works can be found in [15] and [16] by using specialized mappings and image database indexing, respectively. The range of allowed hand motion, in [15] and [16], is fixed by the number of templates. In this way, they require building numerous templates sufficiently during preprocess. At the same time our goal is to try to avoid generating many tem-

[1] The article is published in the original.

[2] This paper uses the materials of the report submitted at the 11th International Conference "Pattern Recognition and Image Analysis: New Information Technologies," Samara, Russia, September 23–28, 2013.

Received March 12, 2014