Rapid Finger Motion Tracking on Low-Power Mobile Environments for Large Screen Interaction

Yeongnam Chae(✉) and Daniel Crane

Rakuten Institute of Technology, Rakuten, Inc., Rakuten Crimson House, 1-14-1 Tamagawa, Setagaya-ku, Tokyo, Japan
{yeongnam.chae,daniel.crane}@rakuten.com

Abstract. Motion and gesture are garnering significant interest as the sizes of screens are getting larger. To provide lightweight finger motion tracking on low-power mobile environments, we propose an approach that breaks down the stereotypes of camera view points. By directing the camera view angle towards the ceiling, the proposed approach can reduce the problem complexity incurred by complicated background environments. Though this change incurs poor lighting conditions for image processing, by clustering and tracking the fragmented motion blobs from the motion image of the saturation channel, rapid finger motion can be tracked efficiently with low computational load. We successfully implemented and tested the proposed approach on a low-power mobile device with a 1.5 GHz mobile processor and a low specification camera with a capture rate of under 15 fps.

Keywords: Motion tracking · Mobile environment · Remote interface

1 Introduction

As screens are getting larger, multi-modal interaction including gesture and voice is becoming a significant research topic due to their intuitiveness and convenience in order to handle more information. Since the introduction of the RGB-D sensor, hand gesture and motion tracking have become rapidly studied fields [3,6,7] among multi-modal interfaces.

However, the approaches adopting RGB-D sensors belong to an area that is struggling to secure popularity due to the requirement of specialized depth sensors for the motion interface. On the other hand, with the growth of the computer vision technology, recognizing hand posture using RGB sensors is also seeing enhanced accuracy [2,4,5]; although this approach still has problems dealing with complex backgrounds, and working in low-power environments. In addition, both RGB and RGB-D based approaches have restrictions on the distance between the human and the sensor if the sensor is installed on a screen.

N. Munekata et al. (Eds.): ICEC 2017, LNCS 10507, pp. 465–468, 2017.
DOI: 10.1007/978-3-319-66715-7_64

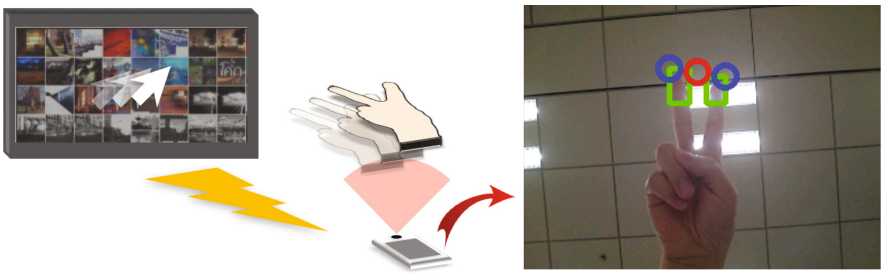

Fig. 1. Overview: remote finger motion interface for large screens

In order to overcome such limitations, in this paper we propose an approach that tracks rapid finger motion in low-power mobile environments. By breaking down the stereotypes of camera view points, we successfully show the potential of using low-power mobile devices as remote motion capturing devices. In addition, by focusing on finger motion rather than gesture we can successfully implement the tracking system on low-power mobile devices, and configure it as a remote interface for large screen interaction, even in the case of blurry images.

2 Proposed Approach

Compared to traditional motion capturing systems on large screens, we separate the capturing sensor from the screen and hand its functionality to remote mobile devices that are more broadly distributed than specialized sensors like RGB-D. By directing the camera view angle towards the ceiling from around 20 cm under the hand we can expect a relatively simple and static background compared to camera perspectives used in traditional motion capture systems. Challenging problems in the proposed system include minimizing lighting effects and detecting fast hand movement within a short distance between the hand and the mobile camera. In this paper we focus on the motion of two fingers to extract more robust finger position than using a single finger, and to open the possibility of extending the method to include recognition of gestures such as finger picking in the future. The overview of the proposed approach is illustrated in Fig. 1. We will describe each of the issues and how we resolved them in the following sub-sections.

2.1 Acquiring Image and Noise Removal

The reasons for most traditional motion capture systems used for large screens being directed towards the human body (along with potentially complex background information) include the camera's mount position, and the spatial relationship between the user and the camera. In the proposed system, by separating the capture function from the screen body to a remote mobile device, we can configure the spatial relationship between the user and the camera more flexibly. However, by directing the camera to the ceiling, the proposed system can not

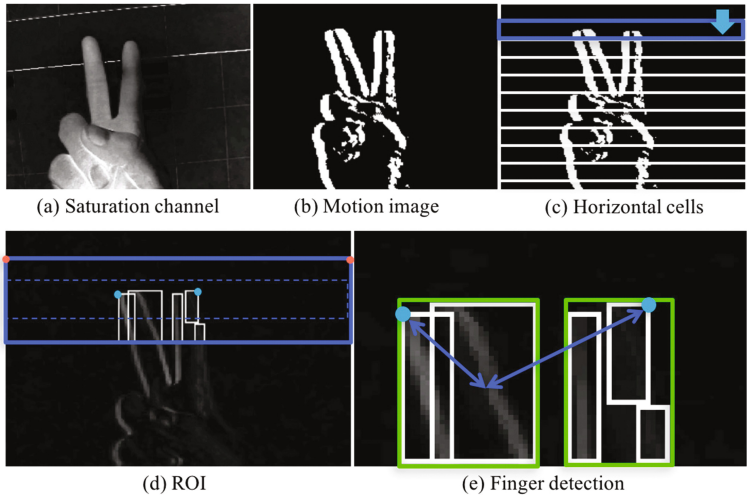

Fig. 2. Rapid finger motion detection (Color figure online)

avoid the effect of lighting which is one of the most challenging problems in the field of computer vision. In order to minimize the effects of lighting, we capture an HSV (Hue, Saturation, Value) image rather than RGB image, and use only the saturation channel, which is known to be robust to changes in lighting. Figure 2(a) shows the captured saturation channel. From the saturation channel, we calculate the motion difference after the noise removal using a morphological operation.

2.2 Rapid Finger Motion Detection and Tracking

Though saturation is relatively robust to changes in lighting compared to RGB, it is not invariant; as such, there can occur fragments of motion blobs in the differential image. Many motion detection approaches adopted Motion History Image (MHI)[1] which accumulates image differentials over time to acquire robust motion blobs. However, in the case of low-power environments it is hard to use MHI because the frame rate of the device is too low to capture rapid finger motion. In the proposed approach, we emphasize differential image by horizontal morphological operation accepting fragmented blobs, as shown in Fig. 2(b). In order to find the fingers in the fragmented motion image, we divide the entire image horizontally and compress each horizontal cell with the column-wise OR operation; Fig. 2(c) shows these horizontal cells. By checking the number of black and white transitions from the top compressed cell to the bottom, we can find the first cell that includes the fingers. Based on the detected horizontal cell, we set up wider ROI (Region Of Interest) and extract the image blobs of each fragmented finger motion, as seen in Fig. 2(d). In order to classify the motion blobs as left finger or right finger, we first find the upper-left and upper-right corners of the region spanned by the motion blobs (indicated by the light-blue dots in

Fig. 2(d)) by comparing the distance of the corners of each bounding box with upper-left and upper-right points of the ROI (indicated by red dots). Using the upper-left and upper-right corners of the region of motion, we can then classify each blob by the corner to which its center is nearest; illustrated in Fig. 2(e). In order to track the motion robustly, we track each finger's motion over time and verify it by comparing size, aspect ratio and overlapping region with the previous frame.

3 Performance

We implemented the proposed approach in a low-power mobile environment with a 1.5 GHz mobile processor and a low specification camera with a capture rate of under 15 fps. Under these conditions, our method successfully detected and tracked rapid finger motion at 21 ms/frame in 320×240 resolution.

4 Conclusion

In this paper, we proposed a rapid finger motion tracking methodology on low-power mobile environments for large screen interaction. In order to reduce the complexity of the problem caused by complicated and dynamic background conditions, we separated the capture function from the screen to a remote mobile device and changed the camera view, accepting fragmented motion caused by rapid finger motion and poor lighting conditions. By detecting finger motion efficiently from the fragmented motion blobs, we have successfully tracked rapid finger motion in a low-power environment. We implemented and verified the proposed approach on a low-power mobile device, which demonstrated real-time performance. In our future work, we will extend this approach to recognize finger motion gestures to provide more flexible interaction.

References

1. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. In: TPAMI (2001)
2. de La Gorce, M., Fleet, D.J., Paragios, N.: Model-based 3D hand pose estimation from monocular video. In: TPAMI (2011)
3. Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J.: Accurate, robust, and flexible real-time hand tracking. In: CHI (2015)
4. Simon, T., Joo, H., Matthews, I., Sheikh, Y.: Hand keypoint detection in single images using multiview bootstrapping. In: CVPR (2017)
5. Stenger, B., Thayananthan, A., Torr, P.H., Cipolla, R.: Model-based hand tracking using a hierarchical Bayesian filter. In: TPAMI (2006)
6. Wan, C., Yao, A., Gool, L.: Hand pose estimation from local surface normals. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 554–569. Springer, Cham (2016). doi:10.1007/978-3-319-46487-9_34
7. Ye, Q., Yuan, S., Kim, T.-K.: Spatial attention deep net with partial PSO for hierarchical hybrid hand pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 346–361. Springer, Cham (2016). doi:10.1007/978-3-319-46484-8_21