Palm area detection for reliable hand gesture recognition

Giulio Marin, Marco Fraccaro, Mauro Donadeo, Fabio Dominio, Pietro Zanuttigh

Department of Information Engineering, University of Padova
Via Gradenigo 6/B, 35131 Padova, Italy

***Abstract*—Hand gesture recognition applications require as a first step a reliable identification of the hand region and its subdivision into fingers and palm areas. In this work the hand is firstly extracted from the depth data acquired by a depth sensor (e.g., Microsoft Kinect). Then the center of the palm and the hand orientation are identified. Finally circular and elliptical shapes are fitted on the extracted samples in order to reliably identify the palm and fingers area. The proposed approach has been tested inside a simple gesture recognition scheme and preliminary results show its reliability.**

I. Introduction

Hand gesture recognition is a quite challenging problem for which several different approaches have been proposed, based both on color information extracted from video sequences framing the hand and, more recently, also on depth data acquired by novel acquisition devices like Microsoft's Kinect [1]. A key aspect in gesture recognition is the identification of the finger positions. Among the various approaches proposed for this task, an effective solution is to use descriptors representing the distance of the various fingertips from the center of the hand. Following this rationale a set of approaches exploits the distance of the hand contour from the hand centroid or some other reference point. For example, Ren et al. [2] build an histogram of the distances of the hand contour points from the centroid of the palm. This approach is affected by the fact that the palm contour is also considered in the histogram construction. Better performance can be obtained if the palm and finger areas are recognized before building the histogram or other descriptors based on the contents of the two regions.

The work presented in this paper is part of a larger project that aims at the construction of a hand gesture recognition system exploiting the combination of several different features in order to recognize hand gestures from the data acquired from consumer depth cameras like the Kinect. Here in particular we focus on the extraction of the palm and finger samples, that is a one of the key steps in producing the data that can be exploited for the feature extraction.

MMSP'13, Sept. 30 - Oct. 2, 2013, Pula (Sardinia), Italy.
978-1-4799-0123-4/13/$31.00

II. Extraction of the hand region

The first step is the extraction of the hand samples from depth data. We assume that the hand is the closest object to the acquisition camera, a quite common situation in many gesture acquisition settings. Let us denote with $X_i, i = 1, .., N$ a 3D point acquired by the depth camera and with $D(X_i)$ the corresponding depth value, in particular $D(X_{min})$ will denote the depth of the closest sample. The set of all the points with depth within a threshold T_h from X_{min} and with an euclidean distance from X_{min} smaller than T_{h2} is computed:

$$\mathcal{H} = \{\mathbf{X_i} | D(X_i) < D(X_{min}) + T_h \wedge \|X_i - X_{min}\| < T_{h2}\} \quad (1)$$

The values of T_h and T_{h2} depend on the hand size (typical values are $T_h = 10$cm and $T_{h2} = 30$cm). This approach allows to reliably divide the hand from the scene objects and from the other body parts (as shown in the example of Fig. 1c). However, it does not completely avoid the inclusion also of the wrist and the first part of the forearm into set $\mathcal{H}$. The wrist and forearm will be removed as described next.

a) b) c) d)

Fig. 1. Extraction of the hand samples: a) Acquired color image; b) Acquired depth map; c) Extracted hand samples (the closest sample is depicted in green); d) Output of the Gaussian filter applied on the mask corresponding to $\mathcal{H}$ with the maximum (i.e., $\mathbf{C_g}$) in red; (*Best Viewed in colors*).

A 2D mask H_{dm} in the depth image space corresponding to the samples in H is built and then filtered by a Gaussian filter with a standard deviation σ, that depends on the distance of the hand from the Kinect. The point $\mathbf{C_g}$ corresponding to the maximum of the filtered image is then detected. As exemplified in Fig. 1d, since the Gaussian filter support is larger than the hand, and the palm is larger than the forearm and denser than the finger region, point $\mathbf{C_g}$ lies somewhere close to the center of the palm region. In case of multiple points with the same maximum value, the closest to X_{min} is selected. Principal Component Analysis (PCA) is then applied to $\mathcal{H}$ in order to extract the main axis $\mathbf{x}$ that roughly

corresponds to the direction of the vector going from the wrist to the fingertips.

III. Circle fitting for palm area detection

The next step of the proposed method is the detection of the largest circle, with center $\mathbf{C}$ and radius r, that can be fitted on the palm region. The circle fitting uses the following procedure: a circle with initial center position $\mathbf{C} = \mathbf{C_g}$ is first expanded by increasing its radius r until 95% of the points inside it belong to $\mathcal{H}$ (we left a tolerance of 5% to account for errors due to noise or artifacts of the depth sensor). After the maximum radius value satisfying the threshold is found, the coordinates of point $\mathbf{C}$ are shifted towards the direction that maximizes the number of samples in $\mathcal{H}$ contained in the circle of center C and radius r. Then, r is increased again and we continue to iterate the two phases until the largest possible circle has been fitted on the palm area (see Fig. 3). The final position of $\mathbf{C}$, with coordinates $\mathbf{C_f}$ represents the palm center and will be the starting point for the next step.

IV. Improved palm detection with elliptical structure

The fitting of the circle on the palm allows to obtain a reasonable but not always accurate estimate of the palm area. This happens for two main reasons: the first is that the palm can have the length in one direction that is longer than the other, e.g., for people with thin hands. The second is that in many acquired gestures the hand is not parallel to the imaging plane and the circular shape gets distorted by the affine warping between the palm plane and the imaging one and by the perspective projection. In order to deal with these issues a more accurate model is needed and we decided to fit an ellipse to the palm region. We start from the center of the circle $\mathbf{C_f}$ ($\mathbf{C_g}$ can also be used to speed up the computation avoiding the circle fitting, but its position is less accurate) and build 12 regions corresponding to different partially superimposed angular directions (we used an overlap of 50% between each sector and the next one as shown in Fig. 2). For each region we select the point of the hand contour inside the region that is closest to the center. In this way we get a polygon contained inside the hand contour that approximates the hand palm. Notice how the choice of using partially superimposed sectors and to take the minimum distance inside each sector ensures that the polygon corners are chosen at the basis of the fingers and the finger samples are not included in the polygon. We then employ the method of [3] to find the ellipse that better approximates the polygon in the least-square sense.

On the basis of the computed data, $\mathcal{H}$ is segmented in three sets corresponding to the palm (the points inside the ellipse), the fingers and the wrist or forearm (the distinction between the last two sets is made on the basis of the direction of main axis identified by the PCA). The computed data is then used to extract histograms showing the distances of the finger points from the hand's centroid in a way similar to [2]. There are however two key differences with respect to [2]: firstly only finger samples are considered for the construction of the histograms and secondly the distances are not the ones from the hand centroid but the ones from the palm edge (i.e. the border of the ellipse).

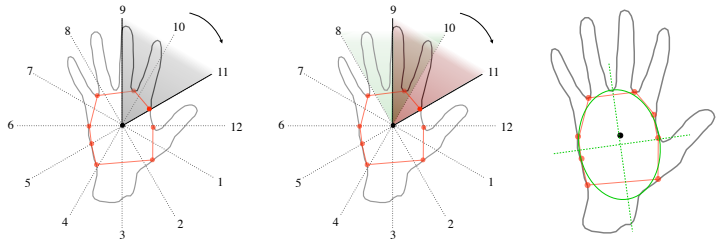

Fig. 2. Angular sectors and extraction of the minima for the computation of the ellipse

V. Preliminary results

Figure 3 shows some examples of the detection of the palm area by the proposed method. Notice how the proposed approach is able to reliably extract the palm region in most situations. Furthermore the ellipse allows a more accurate fitting of the palm area if compared with the simpler circular structure. We also performed some preliminary gesture recognition experiments using distance histograms representing the distances of the samples in the finger set from the side of the ellipse for each angular direction obtaining an accuracy of 92.5% on the database of [2].

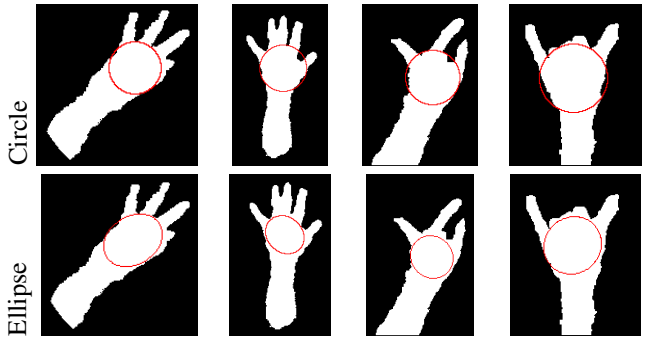

Fig. 3. Examples of fitting of circles and ellipses over the hand palm

VI. Conclusion

In this paper a reliable approach for the extraction of the hand and of the finger and palm regions is presented. It constitutes a first step towards the construction of a gesture recognition scheme that will exploit this data to extract a set of relevant features that will be used into a machine learning framework.

References

[1] C. Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, *Time-of-Flight Cameras and Microsoft Kinect*, ser. SpringerBriefs. Springer, 2012.

[2] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer-interaction," in *Proc. of ICICS*, 2011, pp. 1 –5.

[3] A. Fitzgibbon and R. B. Fisher, "A buyer's guide to conic fitting," in *In British Machine Vision Conference*, 1995, pp. 513–522.