Finger phalanx detection and tracking by contour analysis on RGB-D images

Thalisson Santos
Intelligent Vision Research Lab
Federal University of Bahia
Salvador, Brazil
thalisson.nobre@ufba.br

Luciano Oliveira
Intelligent Vision Research Lab
Federal University of Bahia
Salvador, Brazil
http://www.ivisionlab.eng.ufba.br/

***Abstract*—In this paper we propose a method for identification of the finger phalanges based on the analysis of hand contour in RGB-D sensors. The proposed method is able to partially identify and track the kinematic structure of the fingers. The tracking was performed using the ORB algorithm to match points between a template with some hand images (in different poses) and the image captured. The principal component analysis was performed to compute the hand orientation relative to the image plane. The system will be used as a starting point for a full tracking of the fingers articulated movement.**

***Keywords*-phalanges; hand kinematic; RGB-D cameras;**

I. INTRODUCTION

The analysis of the hand kinematics can be useful in many applications that accurate information about the joints of the fingers is necessary, since human machine interaction up to medical diagnosis or treatment. The use of RGB-D cameras has been widely exploited, eliminating the mandatory use of invasive sensors (e.g. data glove) for accurate results. Visual analysis of kinematics in clinical practice often seeks the use of reflective markers on the joints of the hand to facilitate the mapping of points of interest. However, there may be occlusion and displacement during the execution of movements, hindering the efficient diagnosis of patients, especially those with prostheses [1]. The analysis the movement of hand requires high accuracy during the acquisition of the kinematic variables (joint angles), and complex kinematic systems are often used, these being expensive [2]. This project proposes the creation of a method for detection the phalanges of the finger in depth images captured from an RGB-D camera aimed at creating an inexpensive kinematics system for the analysis of hands. Our method is based on the hand contour analysis by to extract the fingers location (phalanges) without the markers. The kinematic model is built from the phalanges and its position is recalculated every image captured resulting in a constant hand detection process. Currently the method performing the procedure when the hand is fully open and parallel to the sensor image plane.

A. Related work

Hand tracking is a challenging task and many works use markers to achieve efficient results [3] and [4]. However this type of approach is intrusive and usually interferes the movement of the hand preventing its application in clinical research [2], for instance. Different techniques have been proposed to track the hand and estimate its movement. The approach proposed by [5] uses a 3D virtual model of the hand using a point cloud to obtain the hand pose by inverse kinematics while [6] uses the convolutional networks to recover poses of the hand. Using the medial axis technique for extracting fingers and phalanges, [7] presents an approach which demonstrates a solution with inverse kinematics to infer parts of occluded fingers. Hand tracking is invariant to rotation only for the camera view axis. Using the concept of inverse kinematics [8] proposed a multi-constrained approach to accurately reproduce the movement of hand joints from motion data previously captured.

II. OUTLINE OF THE PROPOSED METHOD

The acquisition of images from the RGB-D camera in the scene was strategically set to capture only the upper limb user when the hands are extended on the table and in parallel to lens of sensor.Therefore the lens of the kinect sensor was positioned vertically and slightly set above the computer for a top vision and targeted to hand action. Similar setup can be found in [9].In order to achieve the hand kinematics estimation was necessary to apply a pre-processing on the image to reduce the inherent noise and minimize the variability of pixels presented at the image edges. At the end, the detection of the phalanges is performed by using a template matching approach.

For fingers phalanges identification was utilized techniques of Euclidean Distance Transform (EDT) and thinning algorithm. For hand tracking the ORB algorithm was applied along with the Principal Component Analysis.

The following sections describe in detail all the approach.

A. Preprocessing

This step is necessary to prepare the original image for the following stages.

Be D the depth image acquired by an RGB-D camera, P_i the values of the pixels in the image, and L_{bfr} and L_{aft} the thresholds used to the start and the end of the range, respectively. The resulting image (D_{seg}) can be obtained by

$$D(x, y) = P_i \tag{1}$$

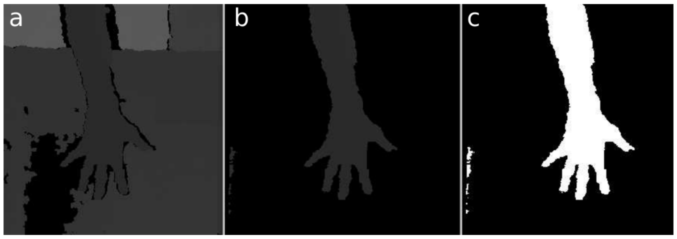

Fig. 1. Segmentation pipeline; (a) Depth map; (b) Segmented depth map; (c) Resulting binary image.

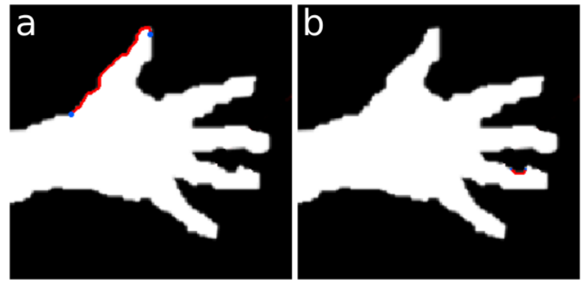

Fig. 2. Setting the k constant; (a) Setting a high value for k generates a hand contour with more point than necessary; (b) Setting a low value for k, generates a hand contour with inadequate amount of points.

$$D_{seg}(x,y) = \begin{cases} P_i & \text{para } L_{bfr} \leq P_i \leq L_{aft} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$D_{bin}(x,y) = \begin{cases} 255 & \text{for } P_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The values closer to 0.5 meters are susceptible to higher incidence of noise and considering the desktop table as background the values determined for L_{bfr} and L_{aft} were 0.8 and 0.95 meters, respectively. The binary image (D_{bin}) and the final segmentation result was extracted from the depth map using the equation 3. This binary image is passed to next step of the proposed method (figure 1c).

The direct use of the contour of segmented hand is not appropriated due the high instability of pixels presented in these areas. In certain cases the pixel intensity causes loss or significant deformation over the hand contour. The following algorithms perform a linear operation on contour points in order to replace them and resulting in a softer effect of the hand contour: the first aims to repair discontinuities while the second applies smoothing.

To apply the following methods is necessary to convert the hand contour to a point vector by using the method proposed in [10] and to reduce the computational cost we simplified the vector calculation with the method proposed in [11].

1) Filling discontinuities: Be V an array of points with n corresponding positions of contour size; $Pt_i(x_a, y_a)$ and $Pt_{i+k}(X_b, y_b)$ arbitrary points at distance of K elements being i the index of the first point and $(i + k)$ the second. A midpoint $M(x_m, y_m)$ can be calculated from the arithmetic mean of their respective coordinates.

$$M(x_m, y_m) = \begin{cases} x_m = & (x_a + x_b)/2 \\ y_m = & (y_a + y_b)/2 \end{cases} \quad (4)$$

Every point in the V vector present among the points Pt_i and Pt_{i+k} will have their coordinates recalculated by do a simple arithmetic mean with the coordinate of midpoint obtained. Finally two new points will be chosen and all points that are between them will replaced. This procedure is performed on all the hand contour resulting in a smoothing contour at end of process.

The purpose of this method is to repair isolated discontinuities belonging the same contour region. Considering high value for k implies long intervals encompassing points from different regions of the object (figure 2a). The inverse can not produce satisfactory results due the low number of points per range (figure 2b). Thus based on test analysis performed on hand contour k was set to value equal to eight.

2) Smoothing: This procedure is similar to algorithm introduced previously but the constant k is set to two. The goal is replace only the intermediate point of a set of three points. Be *P1*, *P2* and *P3* three consecutive points in hand contour the midpoint *M(xm, ym)* resultant is calculated by using the arithmetic mean and assigned to the intermediate point (*P2*). After apply this algorithm the regions with surpluses tend to be more uniform.

After applying these two last steps was necessary to apply a Gaussian filter to reduce the effect of expansion generated by smoothing process and consequently was necessary convert it back to binary image using the equation 5.

$$D_{BinPr}(x,y) = \begin{cases} 255 & \text{for } P_i > 245 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

For performance purposes the vector was again simplified. The figure 3 illustrates an overview of the steps of the preprocessing stage.

B. *Detection of the phalanges*

The detection stage is based on template matching technique and checks the presence of the hand in certain regions of image.

Two fixed regions (R_{lef} and R_{rgh}) of dimension 120x120 pixels were defined on the resulting binary image (D_{BinPr}) to verify the presence of the hand (left and right respectively).

The method consists of applying a logical operation between the pixels of two images: the hand model I_{obj} and the target hand I_{targ}. The hand model is composed by two images extracted from the detection region (R_{lef} and R_{rgh}) on the D_{BinPr} in each frame captured.The target is a fixed model and was previously extracted of same detection region but having their contour slightly simplified to improve efficiency during the comparison process.

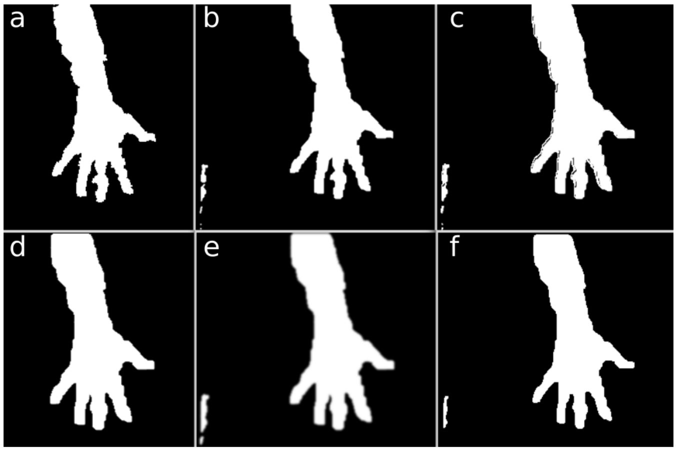

Fig. 3. Overview of the stages of preprocessing: (a) Image with noise; (b) Image of simplified contour; (c) Image after apply the filling algorithm; (d) Image after apply the smoothing algorithm; (e) Image after the Gaussian blur filter using a 7x7 window; (f) Final results after preprocessing.

The comparison calculation between the two images follows the negated exclusive disjunction operation where identical pixels generates true results and different pixels false ones. This operation was applied used the sum of the pixels of the two images producing a new image I_{dj} given by:

$$I_{dj}(x, y) = (I_{obj}(x, y) + I_{targ}(x, y)) \tag{6}$$

Values equal to zero indicate the true negative (tn) equal to 1 represent the false positives and negatives (fpn) and 2 is the number of true positives (tp). Accuracy (ACC) of the model is calculated by:

$$ACC = \frac{tp + tn}{tp + tn + fpn} \tag{7}$$

Due to inconsistency of the hand contour caused by noise is difficult to obtain exact match between the two images. Therefore an image is considered similar to another when ACC exceeds a threshold *P*. Noting the various comparisons *P* was defined as *0.9*.

C. Contourn analysis

The approaches convex hull and convex defects were used to determine the region of fingers and palm in the image. From contour analysis of hand the convexity can be found as the points located at more outer regions which was defined by seven points (figure 4b). By using each finger tip with the two wrist points the procedure for identifying the center of the palm and fingers is less complex.

For to complete this stage successfully is necessary the hand is always open and the fingers fully stretched for found all convexity points and consequently the fingers tips.

The convex hull and convex defect algorithm was enough to detect the location of finger tip and its base (figure 4c) but there are cases which the contour generate convex points not appropriate to fingertips due the instabilities on the edges therefore was necessary to apply the simplification algorithm used in section II-A. The final process is demonstrated in the figure 4.

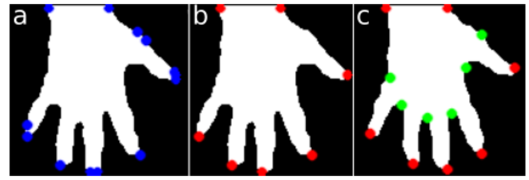

Fig. 4. Convexities of the hand. (a) Results after convex hull; (b) Simplified convex hull points by the method found in [11]; (c) Final results after finding the binary image of the hand with convex hull and convex defects.

D. Kinematics

The kinematic structure was found by considerate all centers of the joints of fingers in conjunction with the hand center (figure 5g). The joints of fingers was extracted from medial axis of each finger while the center of hand was identify using the highest pixel value after applying the EDT.

1) Medial axis of hand: To estimate the center of the phalanges was necessary to extract the medial axis of the hand. To improve the performance was necessary eliminate regions near edge reducing the amount of data to be processed by thinning morphological operation. Sharpen and blur median filters were applied on the picture aftert to apply EDT aim to maximize the difference between the center of hand and its contour and therefore eliminate unnecessary regions by applying a segmentation by threshold. The final result is achieved by applying the method in [12].

2) Finger medial axis extraction: The extraction of the branch for each finger was possible by applying an algorithm to select points of interest over all points resultants from the thinning algorithm and selecting the five first branches (left to right) and discarding the others which not correspond to the fingers (figure 6a).

To refine each branch to match the respective size of each finger was necessary identify the exact limits between dorsal side of the hand and the finger body. This was reached by calculate the midpoint among convex defects points and posteriorly using its location to find the smaller EDT relative to its respective branch. After defined all sizes of fingers the localization of phalanges was identified using statistical values proposed by [13].

E. Tracking

The proposed hand tracking was achieved by matching points between a group of four image of the hand in distinct poses and the hand image after the detection process both processed by EDT and median blur filter respectively. The algorithm ORB (a fusion of FAST keypoint detector and BRIEF descriptor) was used to match the points of interest among the images and the region of image containing the largest number of corresponding points was defined as initial location. This algorithm was chosen due the low computation cost in relative to others algorithms like [14] and [15]. The final location was identified as the position of the pixel with most value within a box (*120x120*) applied at center of initial localization. This size box corresponds to hand size acquired during the detection process described in the II-B section. The principal component analysis was used to calculate the tilt angle of the hand relative to axis of sensor image plane. A part from this approaches was possible to track the hand and its kinematics.

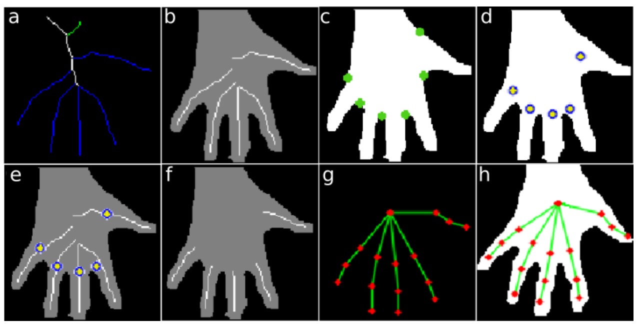

Fig. 5. (a) Extraction process of the medial axis of each finger; (b) Medial axis of the fingers; (c) Green dots indicate the convex defects points; (d) Yellow dots with blue edges indicate the midpoints of the defects of convexity; (e) Midpoints and the branchs to define the exactly location to eliminate the dorsal side of the branchs; (f) Approximate size of each finger; (g) The final kinematics structure; (h) The kinematics structure on the hand where the red dots indicate the phalanges and the hand center.

III. Preliminary results

In this work two solutions were presented for preprocessing stage in order to reduce the noise presented on the edges in images captured by an RGB-D camera. These approaches were based on the relocation of points in the contour of hand (binary image) smoothing only the critical regions avoiding filter algorithm on the entire image. We also presented a method to estimate the approximate location of each phalanx using statistic values representing the approximate position of phalanx on the body of each finger. Finally it was possible to build a kinematic model of the hand that could be used to track the fingers. The tracking of fingers presented is invariant to rotation in relative to axis plan of the sensor and there are few flaws even with presence of other objects in the scene. Critical points were observed when the hands were situated in regions close to the border of the image.

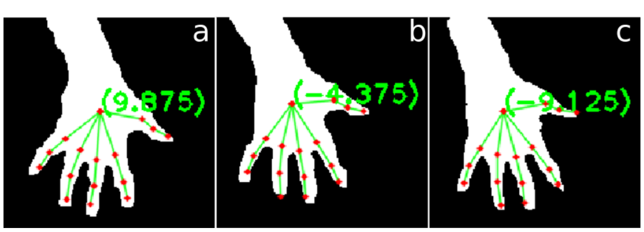

Fig. 6. Results from applying of the PCA algorithm on the binary image to detect the rotation in relative to axis plan of the sensor.

IV. Conclusion

In this paper a method was presented to detection of finger phalanges in RGB-D images. A kinematic model presented was restricted to make a rough estimate of the open hand position providing an initial framework for a more robust and accurate further processing to predict the full independent movement of each finger. The rotation of all the kinematic structure of a hand is shown in figure 6. To future work we intend to improve the hand tracking by using a descriptor of features with correspondence between similar points beyond consider the kinematic of various postures of the hand.

References

[1] F. Ricci, C. Perez, M. Fonseca, E. Guirro, and P. Santiago, "Protocolo experimental para análise cinemática da mão durante a utilização de órteses para membro superior," in *XXIV Congresso Brasileiro de Engenharia Biomédica*, June 2014, pp. 1993–1996.

[2] J. de Abreu, A. P. Fontana, and L. Menegaldo, "Reconstrução da cinemática da mão em pacientes com hanseníase," in *XXIV Congresso Brasileiro de Engenharia Biomédica*, Oct 2014, pp. 357–360.

[3] R. Y. Wang and J. Popović, "Real-time hand-tracking with a color glove," in *ACM SIGGRAPH 2009 Papers*, ser. SIGGRAPH '09, New York, NY, USA, 2009, pp. 63:1–63:8.

[4] A. P. Vicente and A. Faisal, "Calibration of kinematic body sensor networks: Kinect-based gauging of data gloves in the wild," in *Body Sensor Networks (BSN), 2013 IEEE International Conference on*, May 2013, pp. 1–6.

[5] M. Schroder, J. Maycock, H. Ritter, and M. Botsch, "Real-time hand tracking using synergistic inverse kinematics," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 5447–5454.

[6] J. Tompson, M. Stein, Y. Lecun, and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 169:1–169:10, Sep. 2014.

[7] Y. Ouedraogo and Y. Aoki, "Finger posture estimation using 3d medial axes," in *Human System Interactions (HSI), 2014 7th International Conference on*, June 2014, pp. 71–75.

[8] A. Samadani, D. Kulic, and R. Gorbet, "Multi-constrained inverse kinematics for the human hand," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, Aug 2012, pp. 6780–6784.

[9] C. Metcalf, R. Robinson, A. Malpass, T. Bogle, T. Dell, C. Harris, and S. Demain, "Markerless motion capture and measurement of hand kinematics: Validation and application to home-based upper limb rehabilitation," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 8, pp. 2184–2192, Aug 2013.

[10] S. Suzuki and K. be, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32 – 46, 1985.

[11] U. Ramer, "An iterative procedure for the polygonal approximation of plane curves," *Computer Graphics and Image Processing*, vol. 1, no. 3, pp. 244 – 256, 1972.

[12] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.

[13] A. Buryanov and V. Kotiuk, "Proportions of Hand Segments," *International Journal of Morphology*, vol. 28, pp. 755 – 758, 09 2010.

[14] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision 2*, 1999, p. 11501157.

[15] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proceedings of the ninth European Conference on Computer Vision*, May 2006.