GENERATING A 3D HAND MODEL FROM FRONTAL COLOR AND RANGE SCANS

Muhammad Asad[1], *Enguerrand Gentet*[2], *Rilwan Remilekun Basaru*[1], *Greg Slabaugh*[1]

[1]City University London, UK
[2]Ecole Normale Suprieure de Cachan, France

ABSTRACT

Realistic 3D modeling of human hand anatomy has a number of important applications, including real-time tracking, pose estimation, and human-computer interaction. However the use of RGB-D sensors to accurately capture the full 3D shape of a hand is limited by self-occlusions, relatively smaller size of the hand and the requirement to capture multiple images. In this paper, we propose a method for generating a detailed, realistic hand model from a single frontal range scan and registered color image. In essence, our method converts this 2.5D data into a fully 3D model. The proposed approach extracts joint locations from the color image using a fingertip and interfinger region detector with a Naive Bayes probabilistic model. Direct correspondence between these joint locations in the range scan and a synthetic hand model are used to perform rigid registration, followed by a thin-plate-spline deformation that non-rigidly registers a synthetic model. This reconstructed model maintains similar geometric properties as the range scan, but also includes the back side of the hand. Experimental results demonstrate the promise of the method to produce detailed and realistic 3D hand geometry.

***Index Terms*—** Hand model, pose estimation, reconstruction

1. INTRODUCTION

The recent introduction of real-time depth sensors and powerful computing devices has resulted in an increased research interest for natural interaction methods. The advent of mobile and wearable computers that utilize an ego-centric perspective have further necessitated the need for a novel interaction method that can overcome the limitations of the existing tangible and voice activated interfaces [1].

To this end, using the hand as an input for human computer interaction has undergone extensive research. A number of hand pose estimation methods have been proposed which require the use of synthetic hand models to reason how movements of a real hand result in corresponding variations in the input data [2, 3, 4, 5]. Furthermore the existing 3D reconstruction methods are limited by the noise, relatively smaller size of the hand and the requirement to capture multiple images for producing a complete hand model [6]. A realistic hand modeling method is required which can reconstruct a watertight realistic hand model encoding user-specific variations in hand geometry.

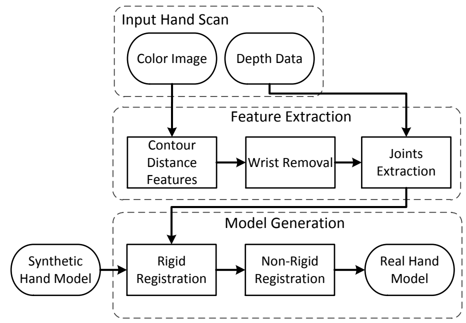


Fig. 1: Flow chart of the proposed model generation method.

We present a method for generating detailed, realistic hand model from a single pair of frontal range scan and color image (as shown in Fig. 1). The proposed method extracts the location of joints using fingertips and valley points through a Naive Bayes probabilistic model. Rigid registration of a synthetic hand model is performed using direct correspondence between joint locations on the range scan and the synthetic model. Non-rigid registration is then achieved using thin-plate-spline deformation, producing a realistic hand model that also includes the back side of the hand.

Previous work on 3D model construction using depth scanning has involved using a temporal sequence of depth and color images to reconstruct different surfaces [6]. In [7] a sequence of range images of the hand are used to build a user-specific hand model. Our method differs from these as we use a single front range and color scan of hand for modeling the geometry of the hand.

This paper is inspired by [8], which also produces 3D models from images. However, unlike [8] which only relies on color data, our method additionally considers a depth channel captured by a range scanner. The range data provides additional information regarding the shape of the hand, but only for the unoccluded frontal side. Our method detects creases using a probabilistic formulation and uses the range data as part of a 3D registration.

1.1. Contributions

To our knowledge, this is the first paper to produce a full watertight 3D hand model using only a frontal color and range image. In addition to the overall framework, we contribute a Naive Bayes approach to extract finger contours using gradient and spatial priors. The final result is a watertight synthetic hand model that captures the details in shape, size and pose of a real human hand. In addition, the proposed method implicitly infers the back side of the hand by applying the shape and size variations from the frontal scan.

2. METHOD

2.1. Problem Definition

Let $\mathcal{S} = \{I, D\}$ be a pair of frontal scans containing pairs of aligned color (I) and depth (D) images for different hands. We propose a method to estimate the 3D hand model $\mathcal{M}$ for each scan in $\mathcal{S}$. The estimated hand models vary in shape, size and pose, whereas the texture comes from an existing hand model and is common for all models in $\mathcal{M}$.

The overview of the proposed 3D hand model generation method is shown as a flowchart in Fig. 1. For each input image pair I and D, the proposed method extracts the contour distance features (*CDF*) using a segmented hand region [1]. Fingertips and valley points are then extracted from the *CDF* as local maxima and minima points. We then remove the excess wrist area from the hand by locating points on the radial and ulnar side of the palm and fitting them to an Euler spiral curve completion method [9]. A Naive Bayes probabilistic method is proposed to infer the finger joints' location which combines the probabilities of joint location from both the gradient image and an anatomical prior. A synthetic hand model with direct joint correspondence is then registered to a skeleton configuration acquired from finger joints in the real hand. Following this non-rigid registration is done using a thin-plate spline [10]. The final outcome is a complete 3D hand model with realistic shape, size and pose and a synthetic texture map on both the front and back. We discuss the proposed method in detail below.

2.2. Fingertips and Valley Extraction

CDF have been previously used for both pose and orientation estimation of the hand [1, 11]. We use these features as an initial step to localize the fingertips and valley points between the fingers. The fingertips $\mathbf{F}_i$ and valley points $\mathbf{V}_i$ are extracted from the *CDF* vector as the local maxima and minima, respectively, with each finger $i \in \{thumb, index, middle, ring, pinky\}$. The proposed method uses an open hand pose, hence the number of visible fingertips and valley points remain constant.

In order to localize a finger completely we require the fingertip and two corresponding valley points on each side of the finger base. As only one valley point lies on the finger base for the thumb, index and pinky, we propose a method for extracting the additional base points on the radial and ulnar side of the hand, respectively.

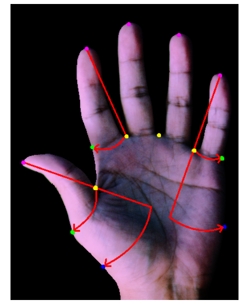


Fig. 2: Fingertip and valley points extraction shows fingertips F_i in magenta, finger valley points V_i in yellow, additional finger valley points V_j extracted using Eq. 2 in green, points extracted for wrist removal in blue and the arcs used for extracting the additional points in red. This figure is best viewed in color.

Given the fingertip point $\mathbf{F}_i$ and the nearest valley point $\mathbf{V}_i$, a circle can be defined centered at $\mathbf{F}_i$ (shown in Fig. 2):

$$(x - F_i^x)^2 + (y - F_i^y)^2 = r^2, \quad (1)$$

where radius r is given as:

$$r = \sqrt{(F_i^x - V_i^x)^2 + (F_i^y - V_i^y)^2}. \quad (2)$$

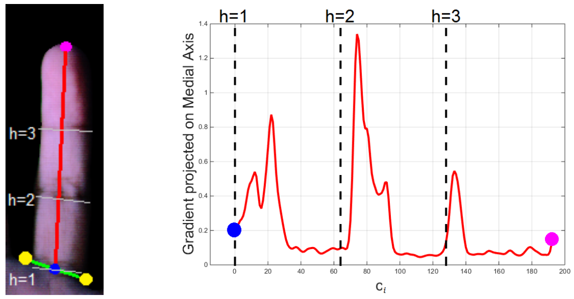
Starting from a given valley point $\mathbf{V}_i$, and assuming the hand is roughly vertical in the 3D scan, we can grow the arc towards radial or ulnar side by decreasing or increasing the value of x in Eq. 2, respectively. The additional valley points $\mathbf{V}_j$ are extracted when the arc intersects with the hand contour, where $j \in \{radial_{thumb}, radial_{index}, ulnar_{pinky}\}$.

2.3. Wrist Removal

In order for the proposed method to work across a range of different users, with varying wrist occlusions, we propose a method for wrist removal. This method uses an Euler spiral curve completion technique from [9, 11] and is similar to [8]. We extract two points on the ulnar and radial side of the hand using Eq. 2 for fingertip $\mathbf{F}_k$ where $k \in \{thumb, pinky\}$. The radius r is increased by a scaling factor α, so that the arc intersects with the contour at points near the wrist. These points are then fed to the curve completion algorithm to remove the excess wrist area from both the hand depth D and color I image.

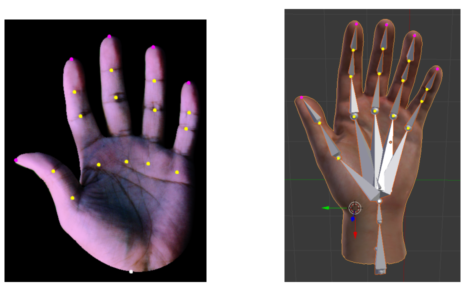
2.4. Joint Extraction

The analysis of the surface anatomy of the hand reveals its direct correspondence with the location of different skeletal



(a) (b)

Fig. 3: Crease detection using anatomical prior for the ring finger shows (a) fingertip F_i in magenta, the two corresponding finger valley points V_i in yellow, finger valley midpoint V_{mi} in blue, the finger's medial axis formed using F_i and V_{mi} and the location of the anatomical prior from Eq. 4 depicted by gray horizontal lines, and (b) the corresponding gradient projection on medial axis. This figure is best viewed in color.



(a) (b)

Fig. 4: Correspondence between joints extracted from the real and synthetic hand model shows the location of fingertips and joints from (a) the real hand image, and (b) the rigidly registered synthetic model taken from [12].

joints. We utilize this observation to formulate the problem of joint localization by first locating the finger creases from the gradient of the color image. This is achieved using a Naive Bayes probabilistic approach to combine the probability of the crease location from both the prior anatomical knowledge and the gradient of the frontal image of a given hand.

2.4.1. *Crease detection*

For a given finger i, we first estimate its length L_i by computing the Euclidean distance between the fingertip F_i and the finger base defined by the finger valley midpoint V_{mi}. Anatomical analysis of the hand reveals that the finger creases lie at a specified ratio β of finger length L_i. Given all the candidate crease locations $c_i \in \{0, \cdots, L_i\}$, we use the prior anatomical observation to formulate a conditional probability for detecting a crease $h \in \{1, 2, 3\}$ as:

$$p\left(h \mid c_i\right) = e^{-\frac{1}{2(\sigma_h)^2}(c_i - \mu_h)^2}, \tag{3}$$

where σ_h and μ_h encode the standard deviation and location, respectively, as depicted in Fig. 3(a), and are defined as:

$$\mu_h = \beta\left(h-1\right)L_i, \qquad \sigma_h = \frac{1}{2}\beta L_i, \tag{4}$$

where $\beta = 1/3$ relates the length L_i, of each finger i, anatomically to the most probable location of the creases.

The conditional probability $p\left(h \mid c_i\right)$ provides a region with high probability of finding a crease, however complete localization is achieved by formulating a gradient-based crease probability. The gradient of the color image is projected to each finger's medial axis defined along the fingertip F_i and the finger valley midpoint V_{mi}. This projection is achieved by accumulating the gradient output along the normal of medial axis for all points between F_i and V_{mi}. The extracted signal is noisy due to the highly textured hand surface. We suppress this noise using a Gaussian filter with a standard deviation $\sigma_f = 1.5$. This filtered signal represents the probability of crease location $p\left(c_i\right)$ in gradient of color image.

Assuming that both probabilities are independent, a Naive Bayes method can be applied to achieve a combined probability $p\left(c_i \mid h\right)$ given by:

$$p\left(c_i \mid h\right) \propto p\left(h \mid c_i\right) p\left(c_i\right). \tag{5}$$

for each crease $h \in \{1, 2, 3\}$ as shown in Fig. 3.

Finally the location of the crease is determined on the medial axis by the maximum-a-posteriori (*MAP*) estimation for each crease in $h \in \{1, 2, 3\}$.

$$c_i{}^* = \arg\max_{c_i} p\left(c_i \mid h\right). \tag{6}$$

The extracted crease locations are then used to acquire the joint location from depth image. The interphalangeal joints (*IP*) directly lie under finger creases near $h = 2$ and $h = 3$, whereas the metacarpophalangeal joint (*MP*) lies 20% of finger length L_i below $h = 1$ on medial axis (Fig. 3(a) and Fig. 4(a)).

2.5. Rigid Registration

The proposed method utilizes direct joint correspondences of a synthetic hand model with the extracted joints of a real hand for rigid registration. The joint configurations from real hands are extracted in the form of joint angles, which are directly applied to the hand model. Fig. 4 shows the corresponding joints in both the real and synthetic hand model.

2.6. Non-Rigid Registration

The size and shape of the rigidly registered synthetic hand model differs from the real hand (Fig. 4). This requires non-rigid registration which is performed by building correspondences based on the location of the fingertips and finger valley points on the contour of both the range data and the

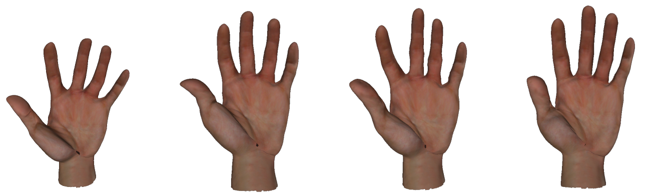


Fig. 5: Example 3D hand models generated by the proposed method.

rigidly registered synthetic model. A thin-plate-spline deformation model is first built using these corresponding contour points which model the non-rigid transformation [10]. This is then applied to the synthetic hand model resulting in a reconstructed model of the real hand.

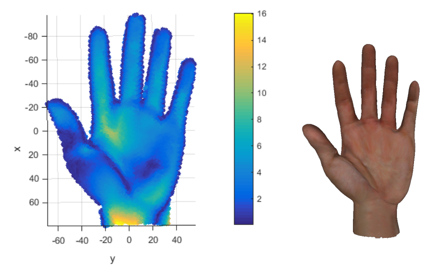
3. EXPERIMENTAL EVALUATION

We evaluate our method on the dataset from [13], which contains high resolution frontal 3D hand range and color scans acquired from different individuals. The error for each vertex on the generated 3D model is calculated by computing the nearest Euclidean distance, measured in mm, with the frontal range scan of the hand. Table 1 presents minimum, maximum and average error for 7 reconstructed models, whereas Fig. 6 localizes these errors for one of the model surfaces. The total average error of 3.65mm in Table 1 depicts the method's ability to reconstruct the majority of the vertices with high accuracy. It can be observed that the majority of the errors exist in the inner palm and wrist region. This is due to the inability of our method to capture the complex shape of the palm and the exclusion of the wrist region from the modeling data. Another source of error is the thumb region, which arises due to the inability of our synthetic hand model to register the complex joint configurations of the thumb in real hands. Despite these limitations, overall the method is effective in reconstructing an accurate 3D hand model that matches the range data (Fig. 5).

In a separate experiment, the ability of the proposed method to automatically infer the back side of the hand is evaluated using a full 3D model of a person's hand, providing a ground truth (*GT*) model. We use computer graphics to render this *GT* model and acquire frontal range and color images. These are then used in our proposed method to generate a 3D hand model. We then compare this reconstructed model to the original *GT* which is presented in Table 1 and depicted on model vertices in Fig. 7. Average error for both the front and the back is 2.51mm which is comparable to the average error for front only validation. Fig. 7 visualizes the error on the *GT* model, where larger errors follow our earlier observation for front side, particularly for the thumb, while the generated surface has no big error on the back side.

Table 1: Euclidean distance error between input data and Predicted model. All measurements are in millimeters (mm).

Input data	Min error	Max error	Average error
1	0.04	25.02	5.58
2	0.10	10.54	2.74
3	0.03	20.47	2.65
4	0.07	12.20	3.43
5	0.03	16.06	3.69
6	0.08	15.06	2.58
7	0.09	19.29	4.85
Total Average	**0.06**	**16.95**	**3.65**
GT Model (Fig. 7)	0.12	13.91	2.51



(a) (b)

Fig. 6: Error (a) in the hand model generation visualized using Euclidean distance between the input range and the generated model for model 5 in Table 1 and (b) the generated model.

4. CONCLUSION

We proposed a method for generating realistic 3D hand models using only single frontal depth and color images of hands. This method used the fingertip and valley points and a probabilistic Naive Bayes approach to locate joints from the hand range scan and color image. The direct correspondences between the extracted joint locations with a synthetic model were used to perform rigid registration. This was followed by a thin-plate-spline deformation to produce a reconstructed model. The experimental evaluation showed an average error of 3.65mm. A *GT* hand model was also used to validate the accuracy of the inferred back side of the hand where average error for the whole reconstructed model was 2.51mm.

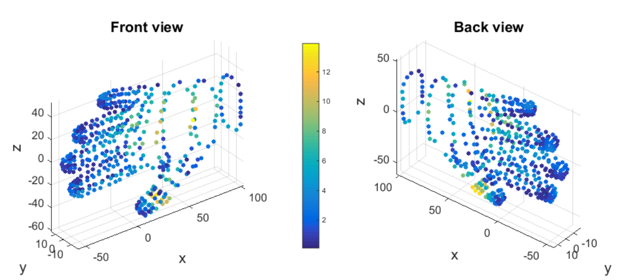


Fig. 7: Euclidean distance error using a *GT* synthetic model with both the front and back side.

5. REFERENCES

[1] Muhammad Asad and Greg Slabaugh, "Hand orientation regression using random forest for augmented reality," in *International Conference on Augmented and Virtual Reality*, 2014.

[2] Martin de La Gorce, David J Fleet, and Nikos Paragios, "Model-based 3d hand pose estimation from monocular video," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 9, pp. 1793–1805, 2011.

[3] Iason Oikonomidis, Nikolaos Kyriazis, and Antonis A Argyros, "Efficient model-based 3d tracking of hand articulations using kinect.," in *BMVC*, 2011, vol. 1, p. 3.

[4] Danhang Tang, Tsz-Ho Yu, and Tae-Kyun Kim, "Real-time articulated hand pose estimation using semi-supervised transductive regression forests," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3224–3231.

[5] Cem Keskin, Furkan Kıraç, Yunus Emre Kara, and Lale Akarun, "Real time hand pose estimation using depth sensors," in *Consumer Depth Cameras for Computer Vision*, pp. 119–137. Springer, 2013.

[6] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al., "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011, pp. 559–568.

[7] Jonathan Taylor, Richard Stebbing, Varun Ramakrishna, Cem Keskin, Jamie Shotton, Shahram Izadi, Aaron Hertzmann, and Andrew Fitzgibbon, "User-specific hand modeling from monocular depth sequences," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 644–651.

[8] Taehyun Rhee, Ulrich Neumann, and John P Lewis, "Human hand modeling from surface anatomy," in *Proceedings of the 2006 symposium on Interactive 3D graphics and games*. ACM, 2006, pp. 27–34.

[9] Benjamin B Kimia, Ilana Frankel, and Ana-Maria Popescu, "Euler spiral for shape completion," *International journal of computer vision*, vol. 54, no. 1-3, pp. 159–182, 2003.

[10] Fred L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[11] Erdem Yoruk, Ender Konukoglu, Bülent Sankur, and Jérôme Darbon, "Shape-based hand recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1803–1815, 2006.

[12] Marin Šarić, "Libhand: A library for hand articulation," 2011, Version 0.9.

[13] Vivek Kanhangad, Ajay Kumar, and David Zhang, "A unified framework for contactless hand verification," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1014–1027, 2011.