Review Article
pISSN 2383-5389 / eISSN 2383-8116
Journal of International Society for Simulation Surgery 2016;3(1):9-11
http://dx.doi.org/10.18204/JISSiS.2016.3.1.009

ORCID
Jaesung Choi: orcid.org/0000-0001-5540-6081
Minkyu Lee: orcid.org/0000-0001-8670-0144
Sangyoun Lee: orcid.org/0000-0003-0394-6777

Multi-Finger 3D Landmark Detection using Bi-Directional Hierarchical Regression

Jaesung Choi, Minkyu Lee, Sangyoun Lee

Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea

Purpose In this paper we proposed bi-directional hierarchical regression for accurate human finger landmark detection with only using depth information.

Materials and Methods Our algorithm consisted of two different step, initialization and landmark estimation. To detect initial landmark, we used difference of random pixel pair as the feature descriptor. After initialization, 16 landmarks were estimated using cascaded regression methods. To improve accuracy and stability, we proposed bi-directional hierarchical structure.

Results In our experiments, the ICVL database were used for evaluation. According to our experimental results, accuracy and stability increased when applying bi-directional hierarchical regression more than typical method on the test set. Especially, errors of each finger tips of hierarchical case significantly decreased more than other methods.

Conclusion Our results proved that our proposed method improved accuracy and stability and also could be applied to a large range of applications such as augmented reality and simulation surgery.

Key Words Landmark Detection · Multi-finger · Hierarchical Structure.

Received: June 7, 2016 / **Revised:** June 8, 2016 / **Accepted:** June 11, 2016
Address for correspondence: Sangyoun Lee
Department of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: 82-2-2123-5768, **Fax:** 82-2-362-5563, **E-mail:** syleee@yonsei.ac.kr

Introduction

Over the past years, many application technologies related to Human Computer Interaction (HCI) have been developed for intelligent devices and wearable sensors. Moreover, they also have been received attention from both the research community and the industry. Almost of them make use of hand information such as hand gesture and posture. Consequently, the technical performance is highly depends on the information of humans' hand and also, the number of products equipped with these technologies rapidly increases. Therefore, the accurate estimation of hand pose becomes a crucial issue to improve the performance of HCI applications.

To detect landmarks, many algorithms have been developing in this research field. Most of them are based on the pixel-based classification (1) and deformable part models (DPM) (2). However, they does not fit well in finger pose estimation since there are large pose variations due to camera viewpoints and various finger gestures. It is challenging to precisely detect the principal landmarks of fingers, as they flexibly deform with a large degree of freedom.

In this paper, the regression based approach is used for estimating 3D multi-finger landmarks. Cascaded regression approaches (3, 4) already used for hand pose estimation. Our approach, however, concentrates on hierarchical structure of the detection strategy and initialization for better accuracy and stability. In our method, the bi-directional search structure is used for estimate landmarks with random forest regression. To

evaluate our approaches, we use ICVL database (5) with two different metrics relations with accuracy and stability.

Materials and Methods

Our method composed of two parts, which are initialization and landmark estimation.

In initialization step, the initial landmarks,θ^0 are estimated from the input depth image. The root joint of middle finger is set as a starting point of the regression of 16 finger landmark. To detect landmarks, the difference of random pixel pair as the feature (1, 2). Then, the tree-based classifier is used for training. Since a scale of feature selection is changed with depth value, it has translation and scale invariant property.

$$I(u_1)-I(u_2) \quad (1)$$

$$u_i=u+\frac{(\delta u_i)}{z(u)},\ i=1,2 \quad (2)$$

Where I is the intensity of image, and u is the reference pixel.

When a test query input, the trained weak learner detects candidates of initial landmarks. Consequently, an average point of candidate pixels becomes initial landmarks. After initialization, 16 three-dimensional landmarks are estimated from initial points, θ^0.

In landmark estimation step, 16 multi-finger landmarks are detected with regression methods. Since only one regression model has insufficient capacity to model the complex pose variation, we use cascaded regression approach.

Cascaded regression has multiple stage t (=1, ⋯ , T) which is consist of weak regressors, Fig. 1. As the stage progresses each estimated landmarks are more accurate with stage regressor R^t. In the end of stage T, 16 multi-finger landmarks are located in proper location (3).

$$\theta^t=\theta^{t-1}+{}^R t \quad (3)$$

To learn each R^t, we use random forest regression method. Fundamentally, the more trees make better performance. The holistic regression estimates the entire 16 landmarks at once. However, the pose variations of different fingers are significantly different and it may cause slow convergence problem. Furthermore, each tips and middle joints of finger have more movement variations than palm and root finger joints. It means that tips have a tendency of large errors than palm and root joints.

To improve the accuracy and the stability, we proposed bi-directional hierarchical approach for regression. As shown in Fig. 2, we divided two hierarchical steps. Palm, root joints and tips are estimated at first. Secondly, tips and middle joints of each finger are estimated from the tip locations at previous hierarchical stage. Consequently, the accuracy and stability of finger landmarks increase than holistic approach.

Results

We evaluate our approaches on the ICVL dataset with 3.4 GHz CPU and 16.00 GB RAM. ICVL dataset consists of 22 K training depth images and 700 test images. Both holistic and bi-directional hierarchical approaches have 8 regression stages and each stage has 10 trees which is trained randomly subsampled features.

Fig. 3 shows the results of our initialization process. From the initial landmark (red point) 16 landmarks are iteratively update with stage regressors. For a fair evaluation, we use two accuracy and stability metrics. The first one is the per-joint error (mm) and another is the success rate. Fig. 4 represent the average joint errors and success rate on training set.

To investigate the performance, we compare the Latent Regression Forest (LRF) (5) with two metrics on test set, Fig. 5.

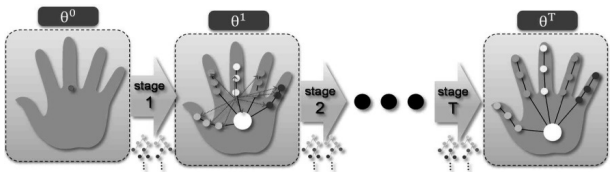

Fig. 1. Cascaded Regression for Landmark Detection.

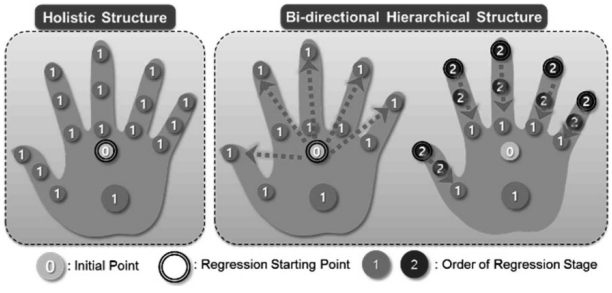

Fig. 2. Holistic and Bi-directional Hierarchical Structure.

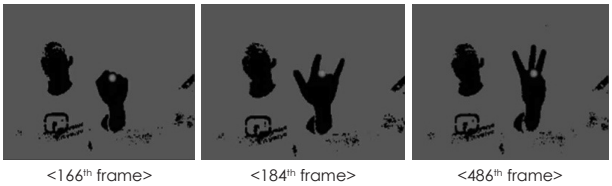

Fig. 3. The Results of Initialization Step (Red Point: θ^0).

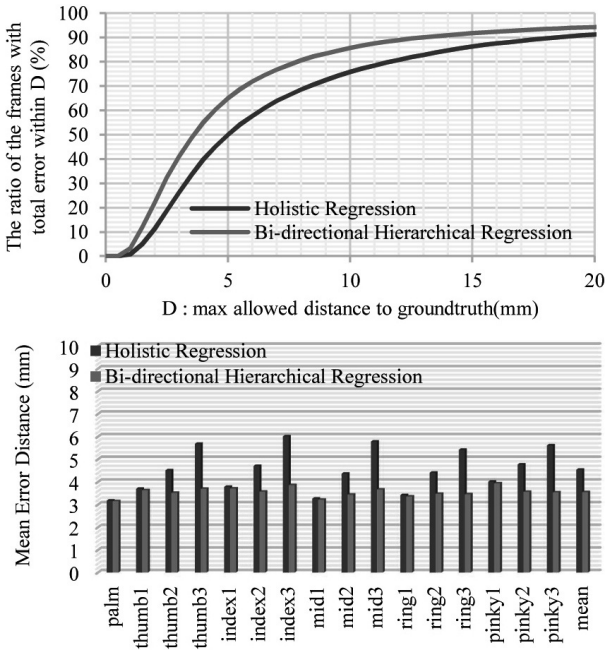

Fig. 4. Success Rate (Up), Average Joint Errors (Down).

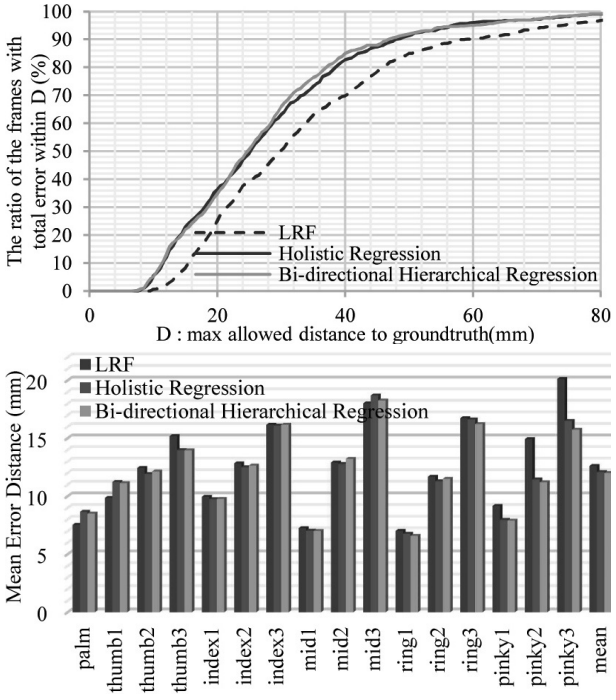

Fig. 5. A Comparison with Other Method on Test Set, Success Rate (Up), Average Joint Errors (Down).

Discussion

According to our experimental results, accuracy and stability increased when the bi-directional hierarchical regression is applied. Especially, a mean error distance of bi-directional hierarchical case has about 22% improvement than typical holistic case.

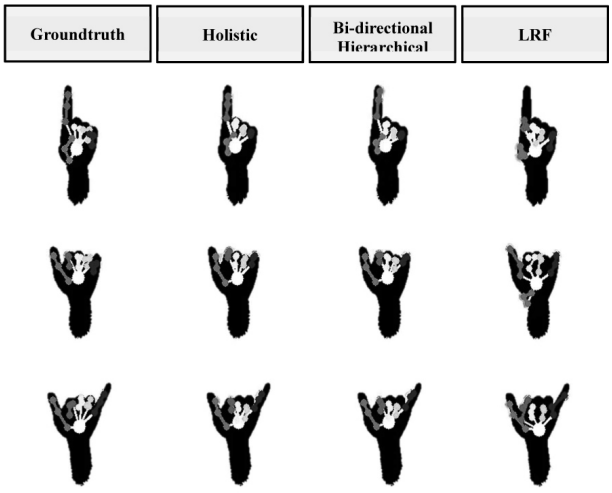

Fig. 6. The Result of 3D Multi-finger Landmark Detection.

On the test set, our approaches show more accurate and stable than typical method, Fig. 6. The accuracy of our results improve 5% than LRF. Generally, an error of each finger tips of hierarchical case has more accurate than other cases. Interestingly, there are a noticeable improvement on pinky case. The success rate also represent that our results are more stable than typical method.

Conclusion

In this paper, we propose a 3D multi-finger landmark detection method using bi-directional hierarchical regression. The experimental results show that the proposed method has improvement in terms of accuracy and stability compared with general methods. Our proposed method can be applied to a large range of HCI applications such as augmented reality and simulation surgery.

References

1. Keskin, Cem, et al. "Real time hand pose estimation using depth sensors." Consumer Depth Cameras for Computer Vision. Springer London;2013. p.119-137.
2. Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE;2008.
3. Dollár, Piotr, Peter Welinder, and Pietro Perona. "Cascaded pose regression" Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE;2010.
4. Sun, Xiao, et al. "Cascaded hand pose regression." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition;2015.
5. Tang, Danhang, et al. "Latent regression forest: Structured estimation of 3d articulated hand posture." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition;2014.