3D Gesture Recognition by Superquadrics

Ilya Afanasyev and Mariolino De Cecco
Mechatronics Lab., University of Trento, via Mesiano, 77, Trento, Italy

Keywords: Superquadrics, Gesture Recognition, Microsoft Kinect, RANSAC Fitting, 3D Object Localization.

Abstract: This paper presents 3D gesture recognition and localization method based on processing 3D data of hands in color gloves acquired by 3D depth sensor, like Microsoft Kinect. RGB information of every 3D datapoints is used to segment 3D point cloud into 12 parts (a forearm, a palm and 10 for fingers). The object (a hand with fingers) should be a-priori known and anthropometrically modeled by SuperQuadrics (SQ) with certain scaling and shape parameters. The gesture (pose) is estimated hierarchically by RANSAC-object search with a least square fitting the segments of 3D point cloud to corresponding SQ-models: at first – a pose of the hand (forearm & palm), and then positions of fingers. The solution is verified by evaluating the matching score, i.e. the number of inliers corresponding to the appropriate distances from SQ surfaces and 3D datapoints, which are satisfied to an assigned distance threshold.

1 INTRODUCTION

Gesture recognition, having the goal of interpreting human gestures via mathematical algorithms, is the important topic in computer vision with many potential applications such as human-computer interaction, sign language recognition, games, sport, medicine, video surveillance, etc. The model-based methods of hand gesture tracking have been studied by a high number of researchers (Rehg and Kanade, 1995); (Starner and Pentland, 1995); (Heap and Hogg, 1996); (Zhou and Huang, 2003); (La Gorce, et al., 2008). Some publications used hand tracking by color gloves with data acquired by fixed-position webcams (Geebelen et al., 2010) or a single camera (Wang and Popović, 2009). The hand tracking with quadrics was used by the authors (Stenger et al., 2001), but they had a model consisted of 39 quadrics, representing only palm and fingers.

The proposed method of 3D gesture recognition by SQ is close to the corresponding hierarchical method (Afanasyev et al., 2012) for 3D Human Body pose estimation by SQ applied for processing 3D data captured by a multi-camera system and segmented by a special preprocessing clothing algorithm. In this paper, the object of recognition is hand gesture; the sensor is MS Kinect; 3D point cloud segmentation is provided by analyzing Kinect RGB-depth data of color gloves. As far as a hand and fingers can be a priori modeled with anthropometric parameters in a metric coordinate system, we propose using the hierarchical RANSAC-based model-fitting technique with the composite SQ-models. As known SQs can be used for description of complex-geometry objects with few parameters and generation of a simple minimization function of an object pose (Jaklic et al., 2000) and (Leonardis et al., 1997). The logic of 3D Gesture Recognition algorithm is clarified by the block diagram (Fig. 1).

The gesture recognition starts with pre-processing 3D datapoints (captured by MS Kinect), segmenting them into 12 parts (forearm, palm and 10 for fingers) according to colors of gloves. Then the algorithm recovers 3D pose of the hand as the largest object ("Hand Pose Search") and after that restores fingers pose ("Fingers Pose Search"). To cope with measurement noise and outliers, the Pose Search is estimated by RANSAC-SQ-fitting technique. The fitting quality is controlled by inlier thresholds (for hand & fingers), which are a ratio of the optimal amount of inliers to whole data points. The tests showed that Hand Pose Search can give a wrong palm position satisfying a palm threshold, but troubling Fingers Pose Search. For this reason, when a finger inliers solution less than a finger threshold, the algorithm restarts the Hand Pose Search again until finding suitable results for every fingers.

Afanasyev I. and De Cecco M..
3D Gesture Recognition by Superquadrics.
DOI: 10.5220/0004348404290433
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 429-433
ISBN: 978-989-8565-48-8

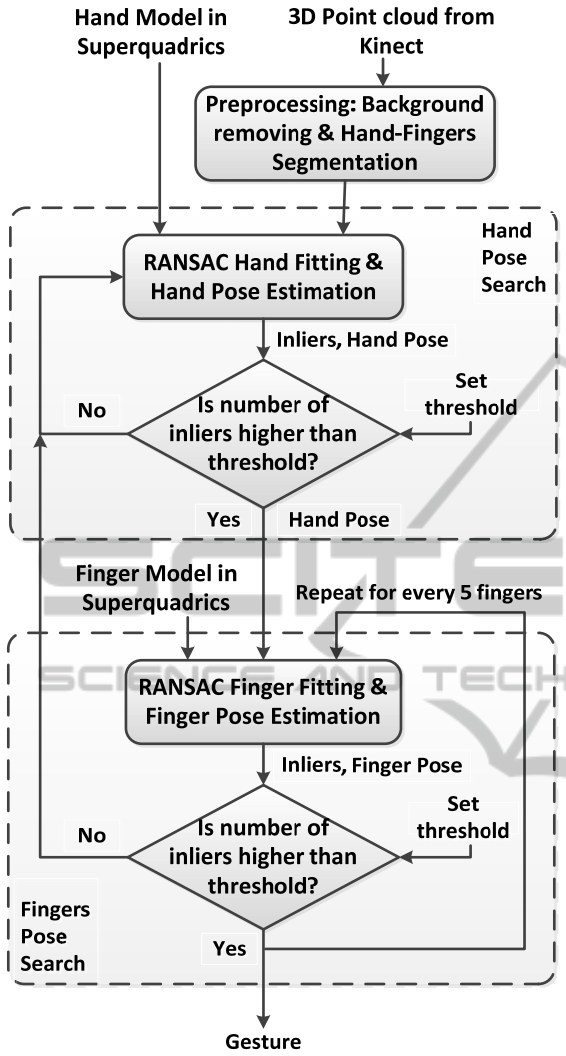

Figure 1: The block diagram of 3D Gesture Recognition algorithm.

2 3D GESTURE RECOGNITION ALGORITHM

2.1 About 3D Sensor and Data

The proposed method of 3D gesture recognition works with 3D sensors captured 3D coordinate and color information, like RGB-D (Red Green Blue - Depth) cameras, multicamera systems, etc. We used MS Kinect with 3D scanning software "Skanect" developed by Nicolas Burrus (Burrus, 2011). The software corrects the image distortions and captures 3D object rawdata in a metrical coordinate system with the origin at Kinect 3D depth sensor (Fig. 2, 3).

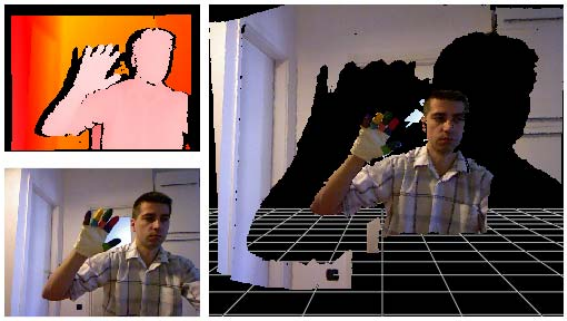

Figure 2: Images acquired by Kinect RGB camera (left, bottom) and Kinect 3D depth sensor (left, top); combined image from Skanect software (right).

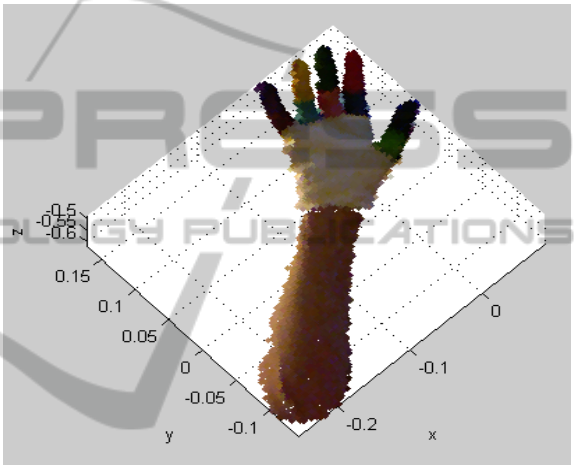

Figure 3: 3D cloud point of a hand with a color glove captured by Kinect with color segmentation.

2.2 Superquadric Parameters

The implicit SQ equation is well suited to mathematical modeling for fitting 3D data (Jaklic et al., 2000) and (Leonardis et al., 1997):

$$F(x,y,z)=\left(\left(\frac{x}{a_1}\right)^{\frac{2}{\varepsilon_2}}+\left(\frac{y}{a_2}\right)^{\frac{2}{\varepsilon_2}}\right)^{\frac{\varepsilon_2}{\varepsilon_1}}+\left(\frac{z}{a_3}\right)^{\frac{2}{\varepsilon_1}} \quad (1)$$

where x, y, z – superquadric coordinate system;
a_1, a_2, a_3 – scale parameters of the object;
$\varepsilon_1, \varepsilon_2$ – object shape parameters.

The explicit SQ equation is used for SQ visualization (where η, ω – spherical coordinates):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} a_1 \cdot signum(\cos\eta)\cdot|\cos\eta|^{\varepsilon_1}\cdot signum(\cos\omega)\cdot|\cos\omega|^{\varepsilon_2} \\ a_2 \cdot signum(\cos\eta)\cdot|\cos\eta|^{\varepsilon_1}\cdot signum(\sin\omega)\cdot|\sin\omega|^{\varepsilon_2} \\ a_3 \cdot signum(\sin\eta)\cdot|\sin\eta|^{\varepsilon_1} \end{bmatrix} \quad (2)$$

Figure 3 illustrates 3D point cloud of hand & finger

modeled in 12 superquadrics – superellipsoids with the shape parameters $\varepsilon_1 = \varepsilon_2 = 0.6$ and the following scaling parameters for:
- Forearm: $a_1 = a_3 = 0.025$, $a_2 = 0.115$ (m).
- Palm: $a_1 = a_3 = 0.04$, $a_2 = 0.018$ (m).
- Phalange: $a_1 = a_3 = 0.008$, $a_2 = 0.027$ (m).

2.3 Hand & Fingers in Superquadrics

2.3.1 Transformation for a Hand

SQ position of a Hand is defined by rotations α, β, γ among *x, y, z* (clockwise) correspondingly and the translation of SQ center (x_c, y_c, z_c) along *x, y, z*. The transformation matrix T_H for the HAND is:

$$T_H = R_H \cdot \begin{bmatrix} 1 & 0 & 0 & x_c \\ 0 & 1 & 0 & y_c \\ 0 & 0 & 1 & z_c \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad where$$

$$R_H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) & 0 \\ 0 & \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

2.3.2 Transformations Hand – Wrist, and Bottom – Upper Phalange Joint

The transformations Hand - Wrist (*H-W*) and Bottom – Upper Phalange Joint (B*P-UPJ*) are the similar and correspond to the matrix:

$$T_W^H = T_{UPJ}^{BP} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & P_W^H \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

$$where \quad P_W^H = P_{UPJ}^{BP} = \begin{bmatrix} 0 \\ a_2 \\ 0 \end{bmatrix}$$

2.3.3 Transformation Wrist - Palm

The transformation Wrist - Palm (*W-P*) is calculated by rotations ξ, ρ, σ among *x, z, y* (clockwise) correspondingly and the translation of SQ center on distance a_2 along y.

$$T_P^W = T_P^W(\xi, \rho, \sigma) = R_P^W \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (5)$$

where R_P^W is the rotation matrix of Palm:

$$R_P^W = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\xi) & -\sin(\xi) & 0 \\ 0 & \sin(\xi) & \cos(\xi) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\rho) & -\sin(\rho) & 0 & 0 \\ \sin(\rho) & \cos(\rho) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\sigma) & 0 & \sin(\sigma) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\sigma) & 0 & \cos(\sigma) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

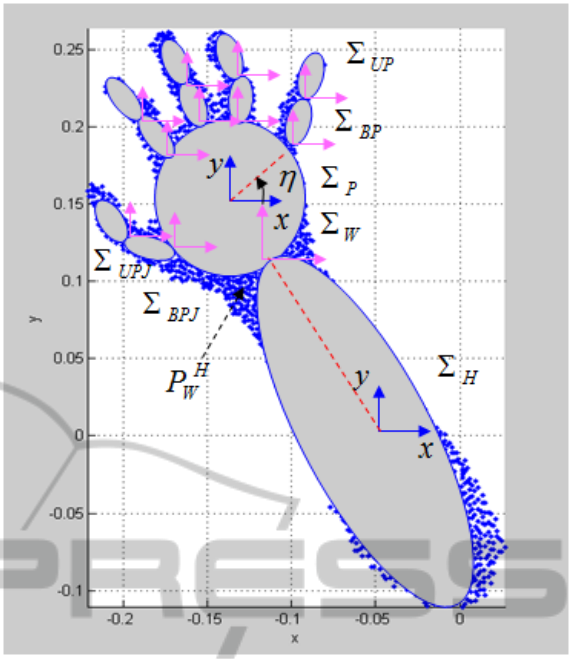

Figure 4: Presentation of Hand in 12 parts: *H* – hand, *P* – palm, *W* – wrist, *BPJ/UPJ* – Bottom/Upper Phalange Joints, *BP/UP* – Bottom/Upper Phalanges, etc.

2.3.4 Transformation: Palm – Bottom Phalange Joint

The transformation Palm – Bottom Phalange Joint (*P-BPJ*) corresponds to the matrix:

$$T_{BPJ}^P = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & P_{BPJ}^P \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad where \quad P_{BPJ}^P = \begin{bmatrix} a_1 \cos(\eta) \\ a_2 \sin(\eta) \\ 0 \end{bmatrix}. \quad (7)$$

2.3.5 Transformation: Bottom Phalange Joint - Bottom Phalange

The transformation Bottom Phalange Joint – Bottom Phalange (*BPJ-BP*) is created by rotations *δ and ε* among *x and z* (clockwise) correspondingly and the translation of SQ center on $-a_2$ along *y*.

$$T_{BP}^{BPJ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\delta) & -\sin(\delta) & 0 \\ 0 & \sin(\delta) & \cos(\delta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\varepsilon) & -\sin(\varepsilon) & 0 & 0 \\ \sin(\varepsilon) & \cos(\varepsilon) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (8)$$

2.3.6 Transformation: Upper Phalange Joint - Upper Phalange

The transformation Upper Phalange Joint - Upper Phalange (*UPJ-UP*) is created by the rotation *θ*

among x (clockwise) and the translation of SQ center on $-a_2$ along y.

$$T_{UP}^{UPJ}(\theta)=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \cos(\theta) & \sin(\theta) & -a_2\\ 0 & -\sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & 0 & 1\end{bmatrix}^{-1} \quad (9)$$

2.3.7 Full Transformation: Hand – Upper Phalange

Finally, taking into account equations (4)-(9), the full transformation for every point of system "Hand – Upper Phalange" (*H-UP*) is:

$$P^{UP}=\left(T_W^H\cdot T_P^W\cdot T_{BPJ}^P\cdot T_{BP}^{BPJ}\cdot T_{UPJ}^{BP}\cdot T_{UP}^{UPJ}\right)^{-1}\cdot P^H \quad (10)$$

where P^H, P^{UP} – coordinates of Hand and Upper Phalange points correspondingly (Figure 3).

2.4 RANSAC-SQ-Fitting Algorithm

The Hand and Fingers Pose Searches are very similar and have the common logic (Figure 1). RANSAC Hand Fitting algorithm is used to find the hand pose hypothesis, i.e. 6 variables: 3 rotation (α, β, γ) and 3 translation coordinates (x_C, y_C, z_C). These variables are needed to calculate the transformation matrix T_H (3). The model described by the superquadric implicit equation (1) is fitted to 3D datapoints sorted by segmentation. Each RANSAC sample calculation is started by picking a set of random points (s = 6) in the world coordinate system (x_{Wi}, y_{Wi}, z_{Wi}). The following equation is used to transform these points to the SQ centered coordinate system (x_{Si}, y_{Si}, z_{Si}):

$$F_{s_i}(x_{s_i},y_{s_i},z_{s_i})=T_{HAND}^{-1}\cdot\begin{bmatrix}x_{w_i}\\ y_{w_i}\\ z_{w_i}\\ 1\end{bmatrix}, \quad (11)$$

where T_{HAND}^{-1} is the matrix of inverting homogeneous transformation for the hand (3).

Then the inside-outside function is calculated according to the superquadric implicit equation (1) in world coordinate system:

$$F_{w_i}=\left(\left(\frac{F_s(x_{s_i})}{a_1}\right)^{\frac{2}{\varepsilon_2}}+\left(\frac{F_s(y_{s_i})}{a_2}\right)^{\frac{2}{\varepsilon_2}}\right)^{\frac{\varepsilon_2}{\varepsilon_1}}+\left(\frac{F_s(z_{s_i})}{a_3}\right)^{\frac{2}{\varepsilon_1}}. \quad (12)$$

The inside-outside function for superquadrics has 11 parameters (Jaklic et al., 2000) and (Solina and Bajcsy, 1990):

$$F_{w_i}=F(x_{W_i},y_{W_i},z_{W_i},a_1,a_2,a_3,\varepsilon_1,\varepsilon_2,\alpha,\beta,\chi,x_c,y_c,z_c), \quad (13)$$

where 5 parameters of the SQ size and shape are known (a_1, a_2, a_3, ε_1, ε_2) and other 6 parameters (α, β, γ, x_C, y_C, z_C) should be found by minimizing the cost-function:

$$\min_{F_W}\sum_{i=1}^{s}F_i(x_i)^2=\left(F_{w_i}^{\ \varepsilon_1}-1\right)^2, \quad (14)$$

where additional exponent ε_1 ensures that the points of the same distance from SQ surface have the same values of F_W (Solina and Bajcsy, 1990).

SQ fitting to the random dataset by minimizing the inside-outside function of distance to SQ surface is realized by nonlinear least-square minimization method with "Trust-Region algorithm" (or "Levenberg-Marquardt algorithm"). The amount of inliers is estimated by comparing the distances between every point of 3D point cloud and SQ model with distance threshold t (t = 1 cm):

$$d_i=\sqrt{a_1\cdot a_2\cdot a_3\cdot\left(F_{w_i}^{\ \varepsilon_1}-1\right)^2}. \quad (15)$$

3 RESULTS

The Figure 5 shows the workability of the RANSAC-based model-fitting the composite SQ-model to 3D point cloud for Gesture Recognition. The presented example concludes 4848 points of 3D data, achieving about 65% of inliers for distant threshold 1 cm. The algorithm has been developed in MATLAB. The RGB-D information was obtained with Microsoft Kinect and then processed offline (taking about several minutes for a gesture). The quality of gesture recognition depends on quality of segmentation that requires good illumination condition and using gloves with bright colors. For some gestures (when fingers are hidden) the method cannot correctly recognize the finger poses.

4 CONCLUSIONS

The paper describes a method of 3D Gesture Recognition by SuperQuadrics (SQ) from 3D point cloud data captured by Microsoft Kinect and clustered according to the colors of the color gloves. The hand was modeled by a composite SQ-model consisted of forearm, palm and fingers with a-priori known anthropometric dimensions. The proposed method based on hierarchical RANSAC-pose search with a robust least square fitting SQs to 3D data: at

first for the hand, then for the fingers. The solution is verified by evaluating the matching score and comparing this score with admissible inlier threshold for the hand and fingers. The gesture estimation technique described has been tested by processing 3D data offline, giving encouraging results.

ACKNOWLEDGEMENTS

The work of Ilya Afanasyev on creating the algorithms of 3D Gesture recognition has been supported by the grant of EU\FP7-Marie Curie-COFUND-Trentino postdoc program, 2010-2013. The authors are very grateful to colleagues from Mechatronics Lab., UniTN, for help and support.

REFERENCES

Afanasyev I., Lunardelli M., De Cecco M., et al. 2012. 3D Human Body Pose Estimation by Superquadrics. *In Conf. Proc. VISAPP'2012 (Rome, Italy), V.2, 294-302.*

Burrus N. Kinect software "Skanect-0.1". 2011. http://manctl.com/products.html.

Heap A.J. and Hogg D.C., 1996. Towards 3-D hand tracking using a deformable model. In *Conf. Proc.* on Face and Gesture Recognition, P.140–145.

Geebelen G., Cuypers T., Maesen S., and Bekaert P., 2010. Real-time hand tracking with a colored glove. In *Conf. Proc.* 3D Stereo Media.

Jaklic A., Leonardis A., Solina F., 2000. *Segmentation and Recovery of Superquadrics.* Computational imaging and vision 20, Kluwer, Dordrecht.

La Gorce M., Paragios N., Fleet D., 2008. Model-Based Hand Tracking with Texture, Shading and Self-occlusions. In *IEEE Conf. Proc. CVPR*. P.1-8.

Leonardis A., Jaklic A., Solina F., 1997. Superquadrics for Segmenting and Modeling Range Data. In *IEEE Conf. Proc. PAMI-19 (11).* P. 1289-1295.

Rehg J.M. and Kanade T., 1995. Model-based tracking of self-occluding articulated objects. In *IEEE Conf. Proc. on Computer Vision*, P. 612–617.

Solina F. and Bajcsy R., 1990. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions PAMI-12(2)*:131-147.

Stenger B., Mendonca P.R.S., and Cipolla R., 2001. Model-based 3D tracking of an articulated hand. *In IEEE Conf. Proc. CVPR 2001 (2)*: 310-315.

Starner T. and Pentland A., 1995. Real-time american sign language recognition from video using hidden Markov models. *In IEEE Proc. Computer Vision*, P. 265-270.

Wang R.Y. and Popović J., 2009. Real-time hand-tracking with a color glove. *ACM Transactions on Graphics (TOG)*, 28 (3), 63.

Zhou H. and Huang T. S., 2003. Tracking articulated hand motion with eigen dynamics analysis. In *IEEE Conf. Proc.* on Computer Vision, V. 2, P. 1102–1109.

APPENDIX

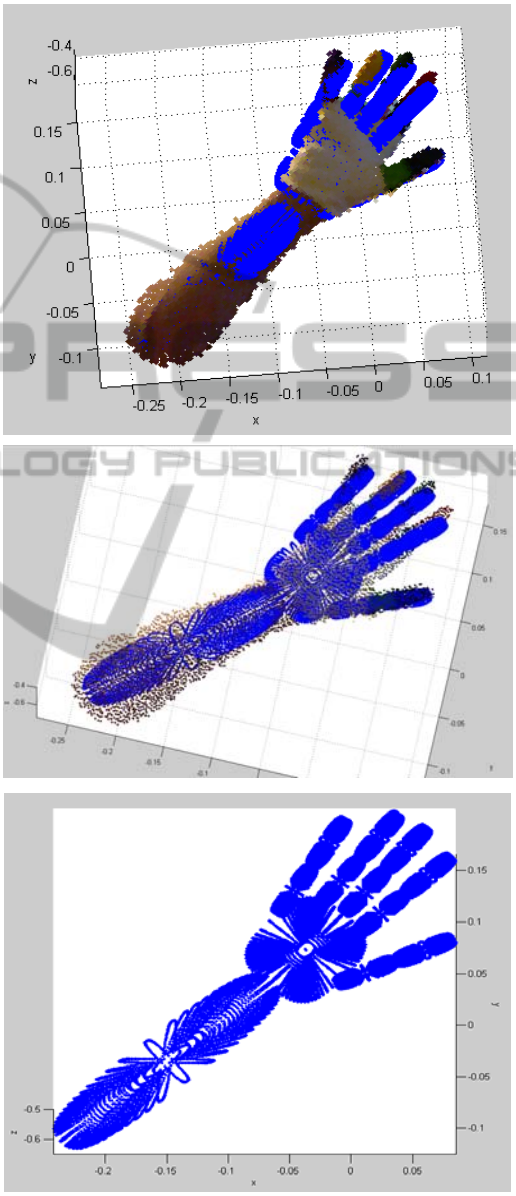

Figure 5: 3D gesture recognition by Superquadrics.