Hand Orientation Regression using Random Forest for Augmented Reality

Muhammad Asad and Greg Slabaugh

City University London, EC1V 0HB, London, United Kingdom
{Muhammad.Asad.2, Gregory.Slabaugh.1}@city.ac.uk

Abstract. We present a regression method for the estimation of hand orientation using an uncalibrated camera. For training the system, we use a depth camera to capture a large dataset of hand color images and orientation angles. Each color image is segmented producing a silhouette image from which contour distance features are extracted. The orientation angles are captured by robustly fitting a plane to the depth image of the hand, providing a surface normal encoding the hand orientation in 3D space. We then train multiple Random Forest regressors to learn the non-linear mapping from the space of silhouette images to orientation angles. For online testing of the system, we only require a standard 2D image to infer the 3D hand orientation. Experimental results show the approach is computationally efficient, does not require any camera calibration, and is robust to inter-person shape variation.

Keywords: orientation estimation, Random Forest regression, silhouette image, hand

1   Introduction

Technological advancements over the recent years have made computing devices powerful, portable and inexpensive. This has given the possibility to rethink how devices are designed to work with us. One of the major hurdles faced by existing technology is the requirement for humans to adapt and learn to use it. This is particularly true for the existing tangible human-computer interaction interfaces, such as keyboard and mouse, which have not seen significant change since their first introduction. Researchers and manufacturers are trying to think of ways technology can adapt to our lifestyle so it does not get in the way but works with us [1].

Recent years have seen increased interest in wearable devices that utilize an egocentric approach for acquiring interaction input from multi-modal sensors and performing everyday computing tasks. These include devices like augmented reality glasses, where a combination of an augmented display and a number of tangible and voice activated interfaces are used for interaction [2]. Such devices lack a novel interaction method which is both computationally efficient and robust. There is a need for an interaction method which can intuitively utilize the egocentric perspective to realize a natural interaction experience.

The human hand is already an effective interaction tool, as a number of hand gestures and postures are used for both communication and manipulation tasks [3]. Most of the previous research on hand-based interaction focussed on recognition of hand gestures and 3D hand pose [3] [4] [5] [6]. Some methods extract the orientation of hand for augmented reality [7] [8] [9] [10]. Our work is closely related to [7], which uses a calibrated camera and recovers a single person's hand geometry in addition to the camera pose. However our approach differs in that it does not require camera calibration, rendering it suitable for a much wider array of applications. In addition, by training the system on data from multiple people, it naturally handles person-to-person hand variations. Our method does not recover the camera pose, instead we focus on acquiring the orientation of the hand itself, which can be used to render an augmented object.

Our method utilizes silhouette images to extract relevant features for regression. Silhouette images have been previously used to extract features for regressing 3D human pose [11] [12]. Multi-view silhouette images have also been used for 3D modelling [13]. Albert *et al.* [14] used multi-view silhouette images to estimate hand pose. Our approach differs from these methods as in our case, we only regress the 3D orientation using hand silhouette images.

We propose a method to regress hand orientation from a dataset of hand images. This method first extracts contour distance features from the hand silhouette images and then uses these features along with the ground truth (GT) orientation angles to train a set of Random Forest regressors. The hand orientation dataset is captured using a commodity depth sensor, where for each hand orientation we have a pair of color image and GT orientation angles. These angles are generated by fitting a plane on the depth image. The dimensionality of contour distance features is then reduced using Principal Component Analysis (PCA). Offline training of a set of Random Forest regressors is performed using the dimensionally reduced features and GT orientation angles. Online testing involves using only dimensionally reduced contour distance features from silhouette images (as shown in Fig. 1) to predict orientation angles. The proposed method is evaluated using single fold and leave-one-out cross-validation.

The rest of the paper is organised in subsequent sections. Section 2 provides the details of the proposed method, while section 3 details the experimental evaluation. Discussion of the evaluation results is presented in section 4. An augmented reality based application of the proposed method is illustrated in section 5. The paper concludes with section 6.

1.1   Our Contribution

A considerable amount of existing research has focused on gesture and pose estimation of human body and hand [3]. However there has been significantly less work done to extract the 3D orientation of a hand from a 2D image [7]. To the best of our knowledge our proposed method is the first to recover hand orientation using silhouette images only. Moreover our proposed method does not require camera calibration and is capable of generalizing variations in hand shape, size and orientation.
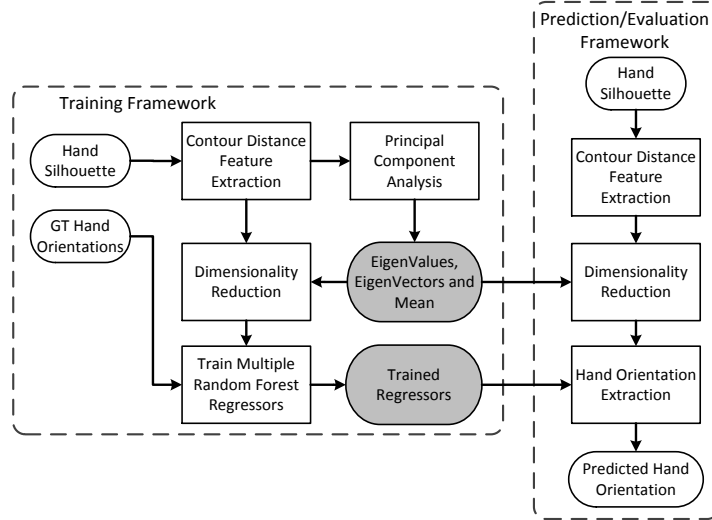
Fig. 1. Flowchart for training and evaluation of the proposed hand orientation method.

We also contribute a method for extracting GT hand orientation angles from a depth image. We note this GT orientation is only used for training of the regression model.

2 Method

Given a set of color images and GT orientation angles of hands, we are interested in finding the mapping between segmented silhouette images and the corresponding orientation angles. The framework is designed to work with uncalibrated cameras and across a range of different shapes, size and style variations of hand.

The flowchart in Fig. 1 presents the different steps in our proposed method. The framework consists of two stages, namely, training and prediction stage. Training is done offline while prediction is done online. Both training and predication require contour distance features to be extracted from hand silhouettes. For the training stage, Principal Component Analysis (PCA) of the training dataset is computed and the corresponding mean, eigenvalues and eigenvectors are used to reduce the dimensionality of the contour distance features in both training and prediction stages. Next, a set of Random Forest regressors are trained using the dimensionally reduced features and GT orientation angles [15]. For the prediction stage these regressors are used to infer the orientation using silhouette images only. The proposed approach is presented in further detail in the subsequent sections below.
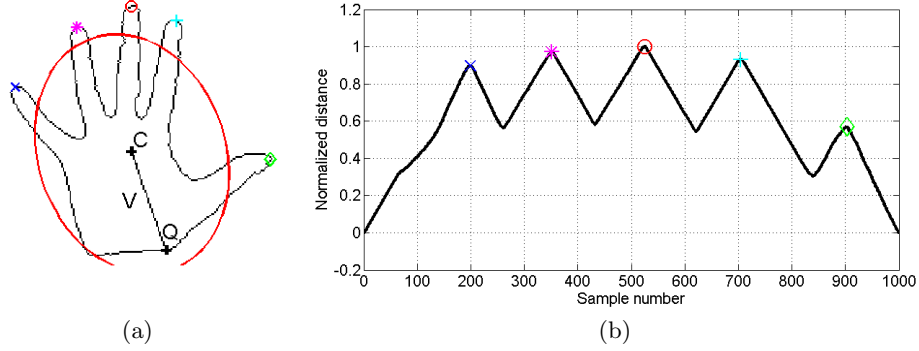
Fig. 2. Contour distance feature extraction from hand contour showing (a) the method for extraction of prevalent point Q on the wrist using a fitted ellipse, centroid C and a ray V and (b) the corresponding contour distance features.

2.1 Feature Extraction

Our method utilizes the contour distance features which are extracted from hand silhouette images. Contour distance features have been previously used for hand shape based gesture recognition [16]. While the main aim of our proposed method is not gesture recognition, contour distance features provide sufficient hand shape variations that can directly correspond to changes in orientation of the hand. Additionally we also employ a method for aligning and normalizing these features. Details on these feature extraction techniques are depicted in the following subsections.

Contour Distance Features: Let $\mathbf{S}_n = \{S_k\}_{k=1}^K$ be a set of hand silhouette images for the nth person. We propose a method to compute a corresponding distance feature set $\mathbf{D}_n = \{\bar{D}_k\}_{k=1}^K$.

The contour extracted from each silhouette image consists of points $P_k = \{P_{ki}\}_{i=1}^I$. The Euclidean distance of each of these contour points $P_{ki} = \{P_{ki}^x, P_{ki}^y\}$ to a prevalent point on the wrist $Q = \{Q^x, Q^y\}$ is determined as:

$$D_{ki} = \sqrt{(Q^x - P_{ki}^x)^2 + (Q^y - P_{ki}^y)^2}, \tag{1}$$

where $D_k = \{D_{ki}\}_{i=1}^I$ is the contour distance feature vector for a set of contour points P_i. This method is illustrated in Fig. 2. The extracted features have different number of samples and magnitude depending on the scale changes and inter-person hand shape variations. To deal with this we normalize a given feature vector as:
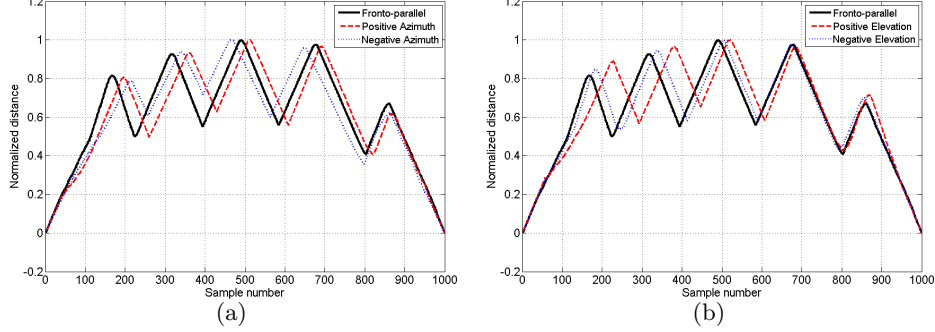
$$\bar{D}_k = \frac{D_k}{\sum_{i=1}^I D_{ki}}. \tag{2}$$

Fig. 3. Variation in the contour distance features with varying orientation in (a) Azimuth (ϕ) axis and (b) Elevation (ψ) axis only.

All extracted feature vectors are resampled to a specified number of samples ρ, in order to use PCA and train Random Forest regressors. For our experimental evaluation we empirically choose $\rho = 1000$.

Extraction of a prevalent point on the wrist: To align the values in the distance feature vectors, we propose a method to extract a prevalent point on the wrist. Given an orientation θ between x-axis and the major axis of an ellipse that fits the hand contour and centroid C, an equation of a ray emanating from C can be defined by:

$$V = \xi \lambda \hat{v} + C, \tag{3}$$

where $\hat{v}$ is the unit vector encoding the direction,

$$\hat{v} = \frac{\begin{bmatrix} 1 \\ \tan\theta \end{bmatrix}}{\sqrt{1^2 + \tan^2\theta}}, \tag{4}$$

ξ is a scalar for correcting the direction of $\hat{v}$,

$$\xi = \begin{cases} +1 & \text{if } \theta < 90° \\ -1 & \text{if } \theta \geq 90°, \end{cases} \tag{5}$$

and λ is a parameter that changes the length of the ray. The direction scalar ξ is calculated using Eq. 5 based on the assumption that the in-plane orientation θ of hand will always be within a predefined range of an upright hand pose with $\theta = 90°$. We define this range to be $0° < \theta < 180°$. This corrects the direction of the ray V so that it is always propagating towards the wrist.

The proposed method increases λ until the ray intersects with the contour at a point $Q \in P_{ki}$ on the wrist. This point is then used as a starting point
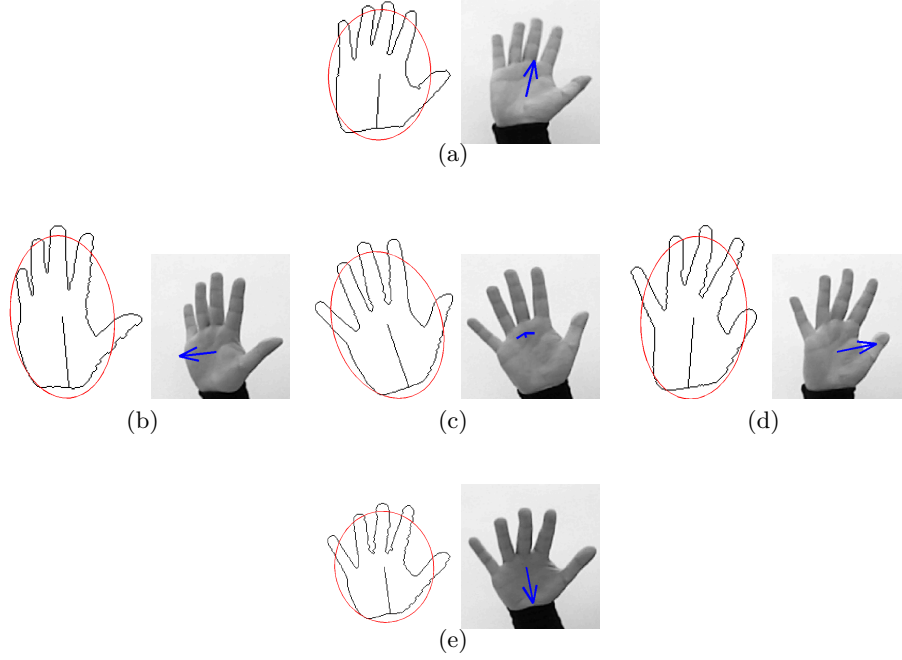
Fig. 4. Variation in the contour corresponding to contour distance features in Fig. 3 along with the normal vector encoding the GT Azimuth (ϕ_g) and Elevation (ψ_g) orientation angles. From Fig. 3 each plot corresponds to the contours in the following manner: (a) Negative Elevation: $\phi_g = +12.60°$ and $\psi_g = -38.96°$, (b) Negative Azimuth: $\phi_g = -36.67°$ and $\psi_g = +8.59°$, (c) Fronto-parallel: $\phi_g = +2.29°$ and $\psi_g = -0.57°$, (d) Positive Azimuth: $\phi_g = +47.56°$ and $\psi_g = +2.29°$ and (e) Positive Elevation: $\phi_g = +10.31°$ and $\psi_g = +41.83°$.

for distance feature calculation. θ represents the in-plane rotation of the hand, and is used along with the other predicted angles to define a complete hand orientation.

We note that changes in the hand orientation can directly induce variation in the contour distance feature. Fig. 3 shows these variations in contour distance feature corresponding to different orientations of hand. To visualize these variations effectively, we only show the contour distance feature for orientations near the ends of our defined orientation space. These orientations are called positive elevation, negative elevation, positive azimuth and negative azimuth. The corresponding hand contour and images, depicting the direct hand shape changes for each angle combination are shown in Fig. 4.
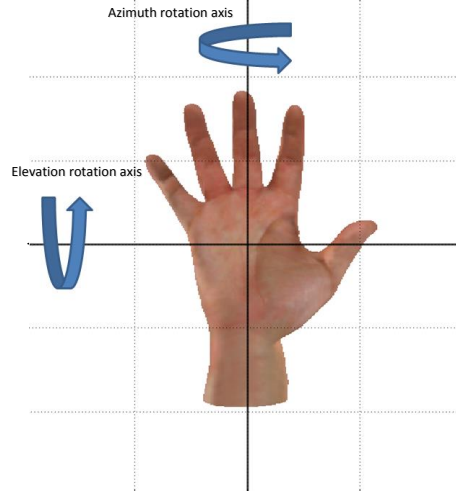
Fig. 5. Rotations axis about which Azimuth (ϕ) and Elevation (ψ) angles vary in the dataset. Image rendered using libhand [18].

2.2 Dimensionality Reduction

The contour distance features extracted from the hand silhouettes have a large number of dimensions. To extract the prominent variations in the dataset, we use PCA for projecting the feature vectors onto a reduced feature space.

We first extract the eigenvectors and eigenvalues of the corresponding feature vectors in the training data. The dimensions of these feature vectors are then reduced by selecting a set of eigenvectors E, that result in 90% energy for corresponding eigenvalues, and projecting the feature vectors onto a reduced space defined by:

$$\check{\mathbf{D}}_n = E^T \left(\mathbf{D}_n - \mu \right), \tag{6}$$

where μ is the mean of all the samples and $\check{\mathbf{D}}_n$ is a set of dimensionally reduced feature vectors [17].

2.3 Ground Truth (GT) Data Generation using Depth Maps

The dataset contains color images and GT orientation angles. These GT orientation angles are only used during training phase and are extracted from aligned depth images by fitting an equation of a plane. For our dataset collection we use an outstretched hand pose which is roughly planar.

We use RANSAC to fit an equation of a plane defined by:

$$n_0 = x n_x + y n_y + z n_z, \tag{7}$$

where the individual coefficients form a normal vector N such that:

$$N = [n_x, n_y, n_z]^T .$$

(8)

This N is used to calculate the corresponding orientation angles:

$$\phi_g = \cos^{-1} n_x, \ \ \psi_g = \cos^{-1} n_y,$$

(9)

where ϕ_g and ψ_g are GT azimuth and elevation angles respectively, as shown in Fig. 4 and Fig. 5.

2.4 Training using Random Forest

Random Forest has been previously used for fast and robust pose estimation for both the full-body and hand [19] [20] from depth images. The motivation to use Random Forest came from its performance with large datasets and that it can be easily implemented on a GPU [21] [15]. It has been proven to handle large variations in body shape, size and pose [19]. However, in this paper we are interested in regressing from the space of hand silhouette images to that of orientation angles.

The dimensionally-reduced features and GT orientations extracted in the previous steps are used to train two Random Forest regressors, one for each orientation angle. In our experimental evaluation we use Random Forest with 1000 trees and 2 features are sampled for splitting at each node.

We generate a dataset which contains 1624 color images and GT orientation from a total of 13 participants. The choice of hand orientation variations used to record the dataset holds significance in depicting the contribution of the proposed method. To generate this dataset we asked our participants to use an outstretched open hand pose throughout the data capture process. They were asked to rotate the hand back and forth, first along the azimuthal axis and then along the elevational axis only (as shown in Fig. 5). Color images and GT hand orientations were recorded while the participants performed these manipulations. As a result of different participants, the dataset contains significant variations in hand size, shape and style of rotations. This current dataset only contains data from participants' right hand, however taking the advantage of mirror symmetry the same dataset can be reflected to generate images for left hand. The GT orientation angles are only used for the training step and are not part of the final prediction method, where only hand silhouettes from color images are used. In the dataset both ϕ_g and ψ_g are limited from $-45°$ to $+45°$.

3 Experimental Evaluation

Evaluation of the proposed approach is done using two different methods. A single fold evaluation is done using 70% of the data for training while holding out 30% data for independent testing. Next, we perform a leave-one-out cross-validation, where in each trial we left one participant's data out for training
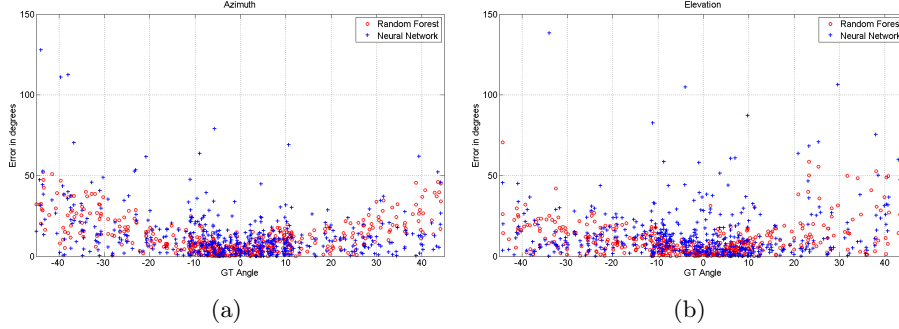
Fig. 6. Absolute prediction error (in degrees) illustrating error (a) ϕ_e and (b) ψ_e in single fold validation using Random Forest and Neural Network regression techniques.

Table 1. Average error in degrees for experimental evaluation in section 3.

Evaluation method	Regressor Used	Azimuth (ϕ_a)	Elevation (ψ_a)
Single Fold	Random Forest	$11.44°$	$9.57°$
	Neural Network	$15.31°$	$14.19°$
Leave-one-out	Random Forest	$12.93°$	$12.61°$
	Neural Network	$20.14°$	$18.85°$

the system, and tested the resulting system on the left out participant. This latter technique demonstrates how the system performs on unseen individuals. For comparison all experiments are also repeated using Neural Network regressor with 1 hidden state containing 1000 neurons. In our experiments, we varied the number of trees in Random Forest and neurons in Neural Network regressors. However changing these parameters did not significantly affect the output of our method. Therefore we empirically fixed the number of trees and neurons to be 1000 for all the experiments. The results are presented below which are then compared and discussed in Section 4.

3.1 Single fold validation

To evaluate the overall performance of the proposed method, we randomly divide the dataset into training and testing sets. The system is then trained and evaluated using the corresponding sets of data.

The absolute predicted errors for this validation are presented against GT orientation angles in Fig. 6. We also present plots of GT orientation angles against corresponding predicted angles in Fig. 7. For comparison, both these figures include results from Random Forest and Neural Network. Average error for orientation angles ϕ_a and ψ_a are presented in Table 1.
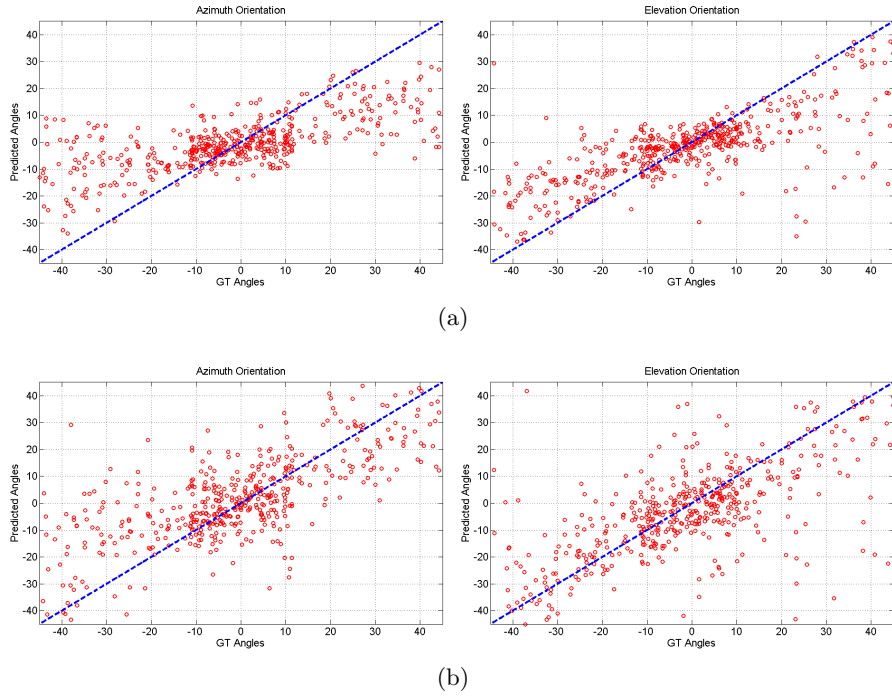
Fig. 7. *GT* vs Predicted Angle plots showing the accuracy of different regressors for predicted angles ϕ_p and ψ_p in single fold validation. The predicted angles are illustrated using (a) Random Forest regressors with number of trees = 1000 and (b) Neural Network regressor with 1 hidden state containing 1000 neurons.

3.2 Leave-one-out cross-validation

We further evaluate our method against a scenario where in each trial, we leave one participant's data out from the training dataset. This left out data is then used for testing. This is a scenario where an unseen hand is used with our method. It is also able to evaluate the ability of the method to handle variations in hand shape, size and orientation without the need for an additional calibration step. The average prediction error for each participant using Random Forest and Neural Network are presented in Fig. 8 (a) and (b) respectively, while Table 1 shows the results for average prediction error for all participants' cross-validation.

4   Discussion

Experimental results show that the proposed method is able to learn the mapping of 2D silhouettes to orientation angles. The method performs well when using Random Forest in both single fold and leave-one-out cross-validation. The average prediction error for single fold evaluation using Random Forest is close
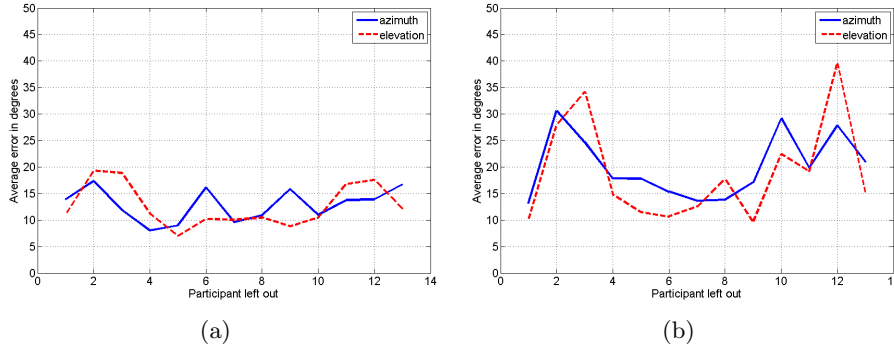
Fig. 8. Absolute prediction error in degrees for leave-one-out cross-validation of each participants' data using (a) Random Forest and (b) Neural Network regression techniques.

to $10°$ for both ϕ and ψ angles (as shown in Table 1). The average execution time of the proposed method for the given set of input silhouette images is found to be 16.93 ms per frame in Matlab implementation on 3.2 GHz Core-i5 CPU.

Fig. 6 shows the absolute prediction errors against GT orientation angles for both Random Forest and Neural Network regressors. It can be seen from this figure that Random Forest is able to model the underlying data well, with significantly less number of outliers as compared to Neural Network. Looking at the range $-10°$ to $+10°$ Random Forest is able to predict with exceptional accuracy, while Neural Network regressor has significant number of outliers falling within the same range.

In Fig. 7 we establish the relationship between GT and predicted orientation angles to illustrate the performance of different regressors in single fold validation. The diagonal line represents the region with optimum results, where we have correct predictions. The closer the predicted data is packed around this diagonal, the better the performance of the regressor is. It can be seen from this figure that for both ϕ and ψ, Random Forest is able to perform better with fewer outliers.

Leave-one-out cross-validation results show that the method is able to produce compelling results for the prediction of orientation for unseen hands. This evaluation method illustrates how well the system can perform with a training data containing different variations in hand shape, size and style. Comparing the average prediction errors for leave-one-out cross-validation with single fold validation in Table 1, there is a significant decrease in the performance of the Neural Network regressors. This highlights the inability of the Neural Network to model the variations in the dataset. Fig. 8 further validates these cross-validation results for each individual participant. In this validation Random Forest produces relatively lesser errors, which indicates its ability to generalize the inter-person variations.
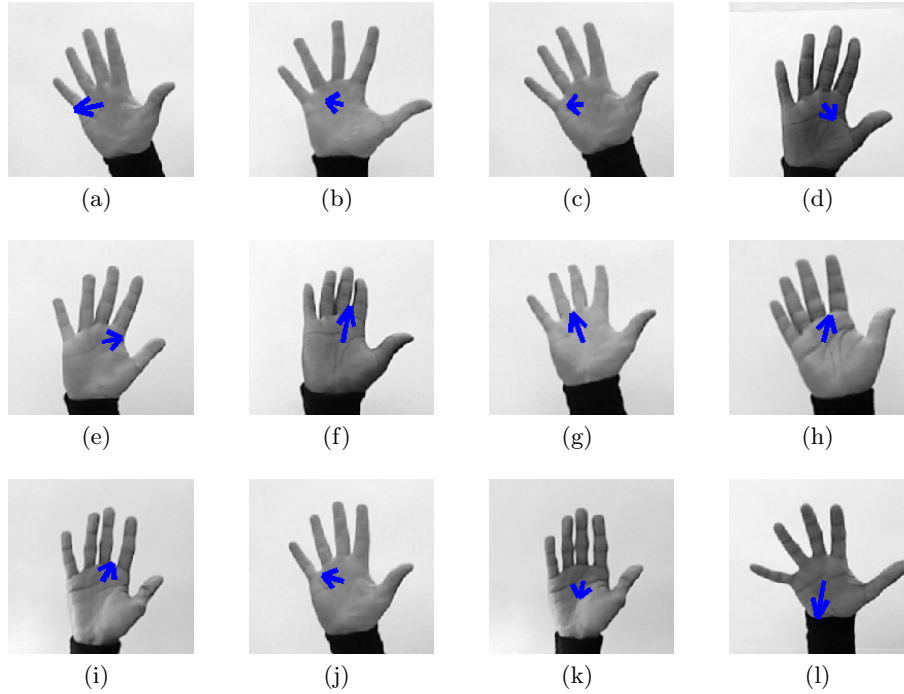
Fig. 9. Success cases for our proposed method. The GT normal vectors are super-imposed on each image to depict the orientation. Error for each case is presented separately in Table 2.

Table 2. Absolute prediction error in degrees for success cases shown in Fig. 9.

Fig. 9	Abolute Error		Fig. 9	Abolute Error	
	ϕ_e	ψ_e		ϕ_e	ψ_e
(a)	$0.89°$	$3.61°$	(g)	$5.17°$	$1.35°$
(b)	$4.33°$	$0.39°$	(h)	$9.08°$	$0.04°$
(c)	$1.18°$	$0.52°$	(i)	$7.83°$	$1.31°$
(d)	$0.81°$	$8.69°$	(j)	$7.12°$	$0.20°$
(e)	$1.17°$	$0.34°$	(k)	$1.07°$	$0.86°$
(f)	$3.18°$	$0.88°$	(l)	$7.18°$	$1.22°$

As shown above our method performs well to recover 3D hand orientation despite of a number of underlying variations in hand shape, size and style. In Fig. 9 we present different success cases using Random Forest in single fold validation. Table 2 shows the absolute error for each of these success cases. The variations in the dataset and the capability of our method is clear from these results.
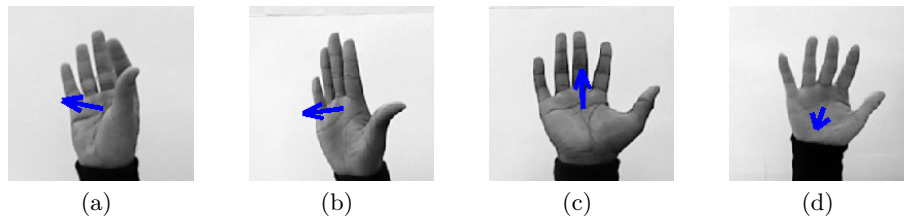
(a) (b) (c) (d)

Fig. 10. Failure cases for our proposed method. The GT normal vectors are super-imposed on each image to depict the orientation. Error for each case is presented separately in Table 3.

Table 3. Absolute prediction error in degrees for failure cases shown in Fig. 10.

Fig. 10	Predicted		Ground Truth		Abolute Error	
	ϕ_p	ψ_p	ϕ_g	ψ_g	ϕ_e	ψ_e
(a)	$2.96°$	$-7.90°$	$-43.46°$	$-7.26°$	$46.42°$	$0.64°$
(b)	$-0.54°$	$2.29°$	$-43.29°$	$5.85°$	$42.75°$	$3.56°$
(c)	$-3.86°$	$-11.16°$	$-1.15°$	$-41.17°$	$2.71°$	$30.01°$
(d)	$-4.88°$	$-17.85°$	$-7.00°$	$20.97°$	$2.12°$	$38.82°$

While our method works well for most of the cases, it does produce errors. Fig. 10 shows some of the cases where our method fails, while Table 3 presents the corresponding error. These failure cases can easily be identified as outliers in the dataset as they do not have outstretched hand pose. In Fig. 10 (a), (c) and (d) the hand does not follow the planar surface assumption which directly affects the calculation of GT orientation angles, whereas in Fig. 10 (b) the fingers are placed too close together making it impossible to extract a contour distance feature that corresponds to the ones in the training dataset. Furthermore by analysing the absolute prediction errors for each failure case in Table 3 it can be seen that our method only fails for the orientation where these assumptions fail. Since our method is regressing both orientation angles independently therefore, even in these failure cases, the unaffected angle is predicted with good accuracy.

5 Application to Augmented Reality

The proposed method can be applied to a number of different application scenarios. In our work, we present an augmented reality based application for visual inspection of virtual objects (shown in Fig. 11). In this application the digital content is overlaid on an augmented layer. Orientation changes from the hand movements are captured using our method and the corresponding orientation transformations are applied to the augmented object.

This kind of visual inspection of virtual objects is useful in scenarios where user does not have access to the actual object, however they want to view it from different perspectives. When applied to an online shopping scenario, a person can effectively view the object they are going to buy. Using this application they will
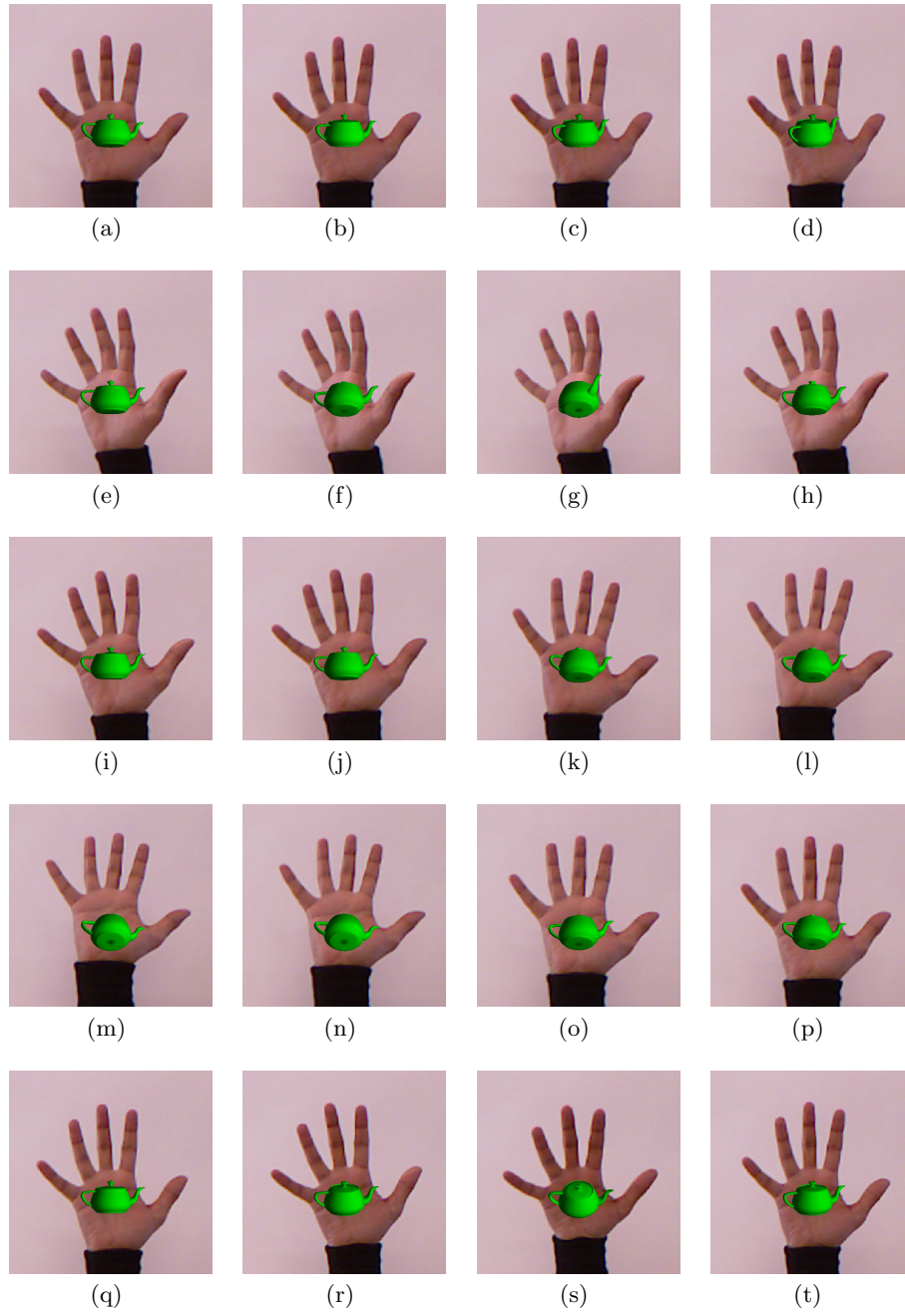
Fig. 11. Application of the proposed approach to augmented reality.

be able to inspect it from different angles in 3D, so as to get the real impression of how the object looks like.

6 Conclusion

A hand orientation regression approach was proposed. This method used a dataset of hand silhouettes only to predict the orientation of the hand in azimuthal and elevational axes. Contour distance features were extracted from hand silhouettes and used along with the GT orientation from depth images to train two Random Forest regressors, one for each angle. The online testing of the system required only a standard 2D image to infer the 3D hand orientation. Comparison of Random Forest with Neural Network based regression shows that the Random Forest is better suited for generalizing the variations in the dataset The system performs well with an average error of $10°$ with single fold evaluation and $12°$ for leave-one-out cross-validation. The proposed method has an average execution time of 16.93 ms per frame in a matlab implementation.

Our future aim is to extend this approach with hand in a number of different poses across different orientations. While contour distance features are able to encode variation related to orientation changes, we endeavour to explore other features as well which might further improve the overall performance of the proposed method. A major challenge for this will be to extract GT orientation. Our existing GT orientation data generation approach can be extended for such scenarios by introducing a palm extraction method. This way, assuming that the palm is rigid, we can again extract the GT orientation of hand. We also envision to use temporal correlation methods such as Kalman filtering to further increase the performance of the proposed method.

References

1. Sara Allison, "Wearable tech - the future, or just a fad?," February 2014, [Online; posted 13-February-2014].
2. Maj Isabelle Olsson, Matthew Wyatt Martin, Joseph John Hebenstreit, and Peter Michael Cazalet, "Wearable device with input and output structures," Sept. 26 2013, US Patent App. 14/037,788.
3. Ali Erol, George Bebis, Mircea Nicolescu, Richard D Boyle, and Xander Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 52–73, 2007.
4. Ying Wu and Thomas S Huang, "Capturing articulated human hand motion: A divide-and-conquer approach," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*. IEEE, 1999, vol. 1, pp. 606–611.
5. Romer Rosales, Vassilis Athitsos, Leonid Sigal, and Stan Sclaroff, "3d hand pose reconstruction using specialized mappings," in *Proceedings of Eighth IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2001, vol. 1, pp. 378–385.
6. Martin de La Gorce, David J Fleet, and Nikos Paragios, "Model-based 3d hand pose estimation from monocular video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1793–1805, 2011.

7. Taehee Lee and Tobias Hollerer, "Handy ar: Markerless inspection of augmented reality objects using fingertip tracking," in *11th IEEE International Symposium on Wearable Computers*. IEEE, 2007, pp. 83–90.

8. Taehee Lee and Tobias Hollerer, "Hybrid feature tracking and user interaction for markerless augmented reality," in *IEEE Virtual Reality Conference, (VR'08)*. IEEE, 2008, pp. 145–152.

9. Taehee Lee and T Hollerer, "Multithreaded hybrid feature tracking for markerless augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 3, pp. 355–368, 2009.

10. Haruhisa Kato and Tsuneo Kato, "A marker-less augmented reality based on fast fingertip detection for smart phones," in *IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2011, pp. 127–128.

11. Ankur Agarwal and Bill Triggs, "Recovering 3d human pose from monocular images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 44–58, 2006.

12. Ahmed Elgammal and Chan-Su Lee, "Inferring 3d body pose from silhouettes using activity manifold learning," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2004, vol. 2, pp. II–681.

13. Jean-Sebastien Franco and Edmond Boyer, "Fusion of multiview silhouette cues using a space occupancy grid," in *Tenth IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2005, vol. 2, pp. 1747–1753.

14. Albert Causo, Etsuko Ueda, Yuichi Kurita, Yoshio Matsumoto, and Tsukasa Ogasawara, "Model-based hand pose estimation using multiple viewpoint silhouette images and unscented kalman filter," in *The 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*. IEEE, 2008, pp. 291–296.

15. Leo Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

16. 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.

17. Michael E Leventon, W Eric L Grimson, and Olivier Faugeras, "Statistical shape influence in geodesic active contours," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2000, vol. 1, pp. 316–323.

18. Marin Šarić, "Libhand: A library for hand articulation," 2011, Version 0.9.

19. Jamie Shotton, Ross Girshick, Andrew Fitzgibbon, Toby Sharp, Mat Cook, Mark Finocchio, Richard Moore, Pushmeet Kohli, Antonio Criminisi, Alex Kipman, et al., "Efficient human pose estimation from single depth images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2821–2840, 2013.

20. 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.

21. Toby Sharp, "Implementing decision trees and forests on a gpu," in *Computer Vision–ECCV 2008*, pp. 595–608. Springer, 2008.