Learning Marginalization through Regression for Hand Orientation Inference

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

Abstract

We present a novel marginalization method for multi-layered Random Forest based hand orientation regression. The proposed model is composed of two layers, where the first layer consists of a marginalization weights regressor while the second layer contains expert regressors trained on subsets of our hand orientation dataset. We use a latent variable space to divide our dataset into subsets. Each expert regressor gives a posterior probability for assigning a given latent variable to the input data. Our main contribution comes from the regression based marginalization of these posterior probabilities. We use a Kullback-Leibler divergence based optimization for estimating the weights that are used to train our marginalization weights regressor. In comparison to the state-of-the-art of both hand orientation inference and multi-layered Random Forest marginalization, our proposed method proves to be more robust.

1. Introduction

In recent years, real-time depth cameras have been the center of attention for novel natural interaction methods [1]. These cameras, however, have limited availability on mobile devices due to the consideration of power consumption, cost and form-factor [2]. In contrast, 2D monocular cameras are readily available in majority of the mobile devices. Therefore, methods that utilize 2D monocular images in new ways can significantly contribute towards novel interaction on such devices.

Furthermore, Augmented Reality (*AR*) based methods can make the interaction experience more natural as the real-world orientation extracted from these methods can be used to *blend in* the virtual objects [3]. A hand orientation based *AR* provides the user direct control of the virtual objects, resulting in more natural interaction [4]. In this paper, we address the inference of hand orientation angles, resulting from flexion/extension of the wrist and pronation/supination of the forearm measured along the azimuth and elevation axes [4,5]. The proposed method infers hand orientation using only a single 2D monocular camera,

(a) (b) (c)

Figure 1: The proposed hand orientation regression enables interaction with augmented object where, (a) hand images with an outstretched pose are acquired, (b) shape features from a hand segmentation are extracted to infer (c) hand orientation angles, enabling a user to control the orientation of the object.

which can be used to interact with augmented objects (see Fig. 1). Furthermore, the inferred hand orientation can also be used to simplify the estimation of hand pose and articulation in 3D model based methods [6, 7].

Hand orientation regression has been previously addressed in [4], which is restricted by limited coverage of orientation angles and the assumption that each orientation angle varies independently. This assumption allowed the use of two independently trained single-variate Random Forest regressors. Existing work on marginalization of multi-layered Random Forest (*ML-RF*) focused on two layered learning, where the first layer is composed of a global classifier and second layer presented expert regressors trained on subsets of a dataset [2, 8, 9]. In these methods, the marginalization is defined as the weighted sum of the posterior probabilities from expert regression layer. The weights for this marginalization come from the posterior probabilities of the global classifier in the first layer. All these *ML-RF* methods rely on posterior probabilities from the first layer which tends to underestimate the true posteriors, making these methods prone to errors [10]. Moreover, as the global classifier in the first layer is trained separately to the expert regressors, these methods do not address for the inaccuracies arising from the posterior probabilities of the expert regression layer.

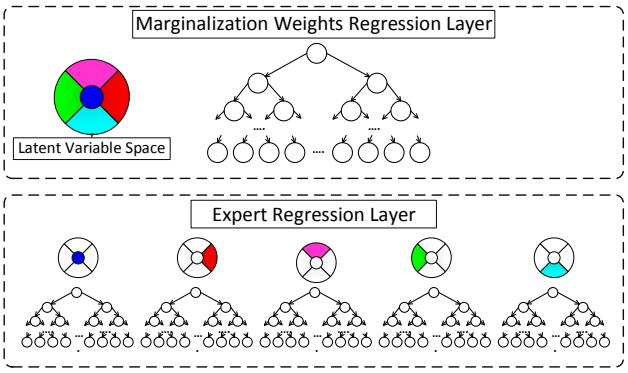


Figure 2: The proposed multi-layered marginalization through regression method utilizes marginalization weights regressor in the first layer to infer the weights for marginalizing posterior probabilities from each expert regressor in the second layer.

To address the hand orientation inference problem and to better couple the multiple layers in a *ML-RF* regressor, we propose a method for learning the marginalization of a *ML-RF* regressor. This method is applied for learning the mapping of hand silhouette images onto hand orientation angles. We extract and use contour distance features from hand silhouette images [4]. The proposed *ML-RF* regressor contains two layers, where the first layer consists of a marginalization weights regressor while the second layer contains expert regressors trained on subsets of our hand orientation dataset (see Fig. 2). We use a latent variable space to divide our hand orientation dataset into subsets, which are then used to train regressors in expert regression layer. Given an input training sample, each expert regressor gives a posterior probability for assigning it to a given latent variable. Our main contribution comes from the regression based marginalization of these posterior probabilities. Given the training data, we use a Kullback-Leibler divergence based optimization for estimating weights for expert regressor. These weights, along with the features from training set, are used to train our marginalization weights regressor. To the best of our knowledge, our proposed method is the first to learn marginalization using regression for *ML-RF* regression.

The rest of the paper is structured as follows. We present an overview of related work on hand pose estimation, orientation inference and marginalization of *ML-RF* in Section 2. Our proposed method, including assumptions, latent variable generation, expert regression layer and marginalization through regression, is detailed in Section 3. Section 4 presents the experimental validation and Section 5 concludes the paper.

2. Related Work

This section presents a review of the previous methods involving hand pose and orientation estimation. We include the review of hand pose estimation methods as these could be related to single-shot hand orientation estimation, where some of these methods also exploit the quantized orientation of the hand [11]. However, accurate hand orientation estimation is addressed only by a few methods [3, 4, 12]. To achieve their goals, researchers have employed different modes of input data, including colored gloves, color and depth images [13]. The following sections present a brief overview of generative and discriminative hand pose estimation methods. This is followed by the presentation of existing work on hand orientation inference. We then present the methods that utilize marginalization of *ML-RF*.

2.1. Generative Methods

Generative methods use a model-based approach to address the problem of hand pose estimation. By optimizing the parameters of a hand model to the input hand image, these methods can simultaneously estimate the articulated hand pose and orientation. A major limitation of 2D monocular cameras is that the projected 2D image loses vital depth information, which gives rise to an ambiguity problem where it becomes difficult to differentiate two different postures with similar 2D image projections. Generative methods are capable of addressing this ambiguity problem in a 2D image by utilizing a fully articulated 3D hand model [6,7]. de La Gorce et al. [6] optimized the texture, illumination and articulations of a 3D hand model to estimate hand orientation and pose from an input 2D hand image. A similar method was proposed in [7], where generative models for both the hand and the background pixels were used to jointly segment and estimate hand pose. Some of the recent generative methods also utilized depth images and advanced optimization techniques such as particle swarm optimization [14–16]. The multi-camera based generative method in [15] recovered hand postures in the presence of occlusion from interaction with physical objects. Although these generative techniques are capable of estimating the underlying articulations corresponding to each hand posture, they are plagued by the drifting problem. The errors in the pose estimation are accumulated over time, which degrades the performance as the model drifts away from the actual hand pose [11]. Moreover optimizing the parameters with up to 27 degrees of freedom (*DOF*) for 3D hand models is computationally expensive [13], and in some cases requires implementation on a GPU to achieve close to real-time execution [14].

2.2. Discriminative Methods

These methods are based on learning techniques and are able to learn the mapping from the feature space to target

parameter space. Their ability to infer a given parameter from a single input [17] has been a major factor in their recent popularity. Furthermore, these methods are computationally lightweight as compared to generative approaches [18].

A number of discriminative methods have been previously proposed to estimate hand pose [9, 11, 19, 20]. Wang et al. [19] used nearest neighbor search to infer hand pose from 2D monocular images. The approach relied on colored glove and a large synthetic dataset of hand poses. In [20], a Random Forest classifier was trained on a large dataset of labelled synthetic depth images to estimate the hand pose. Keskin et al. [9] showed that the performance of the method in [20] can be improved by dividing the dataset into clusters and using a multi-layered Random Forest (*ML-RF*) classification. A major challenge faced by methods relying on synthetic datasets are their lack of generalization for unseen data. Tang et al. [11] addressed this issue by proposing a semi-supervised transductive Regression Forest for articulated hand pose estimation. This approach learned hand pose from a combination of synthetic and realistic dataset of depth images. In [17], generalization for human body pose was addressed by incorporating real scenario based variations into the synthetic data generation method.

2.3. Orientation estimation

There have been a limited number of methods in the literature that estimate hand orientation [3, 4, 12]. Most of these methods use camera calibration and hand features to build a relationship between camera pose and hand orientation. These methods do not address the generalization problem and hence require a calibration step for every new user and camera setup. Regression has only been applied to hand orientation in [4], which used limited orientation angles. This method utilized two single-variate *RF* regressors based on an assumption that the orientation angles vary independently. However, in reality, hand orientation angles are highly dependent on each other. To exploit this dependence, we use a *ML-RF* regression method that uses multivariate regressors to regress the orientation angles together. Additionally, we use a hand orientation dataset that covers a more detailed orientation space. Furthermore, the proposed method does not require camera calibration which renders it suitable for a wider array of applications across different devices. The dataset used for training the proposed method comes from multiple people, which enables it to naturally handles person-to-person hand variations.

2.4. Marginalization of multi-layered Random Forest

Previous work on hand pose estimation have utilized *ML-RF*, where complex problems have been divided and solved by a number of expert regressors trained on simpler subsets of the data. Keskin et al. [9] proposed a *ML-RF* classification for hand pose estimation, which was divided into two classification layers, namely, shape classification and pose estimation layer. Three most significant posterior probabilities from the first layer were used to marginalize the posterior probabilities in the second layer. A similar *ML-RF* regression method was proposed in [2], where the first layer performed coarse classification and the second layer achieved fine regression. Marginalization in this method was done using posterior probabilities from coarse classification layers as weights for predictions at the fine regression layer. Dantone et al. [8] proposed Conditional Random Forest for detecting facial features. This method also used all posterior probabilities from both layers for marginalization. All these methods relied on posterior probabilities from the first layer which tends to underestimate the true posteriors, making these methods prone to errors [10]. Furthermore, as the first layer is trained independent to the second layer, these methods cannont recover from inaccuracies arising from the posterior probabilities of the second layer. To the best of our knowledge, our proposed method is the first to use a marginalization weights regressor that utilizes marginalization weights extracted from posterior probabilities of the expert regressors using the training data.

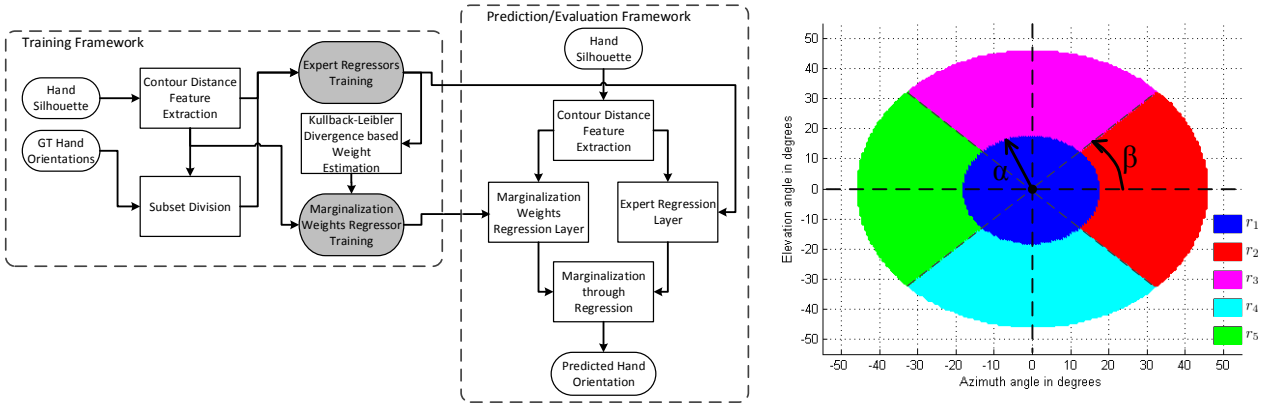
3. Proposed method

Let $\mathcal{S}$ be a dataset of input hand silhouette images captured from an uncalibrated 2D monocular camera such that it contains variations in the hand orientation, shape and size. Given $\mathcal{S}$ we extract the contour distance features $\mathcal{D}$ for each silhouette image [4]. The problem of hand orientation estimation aims at using $\mathcal{D}$ to infer the azimuth (ϕ) and elevation (ψ) angles of the hand along its two major axes, namely, the azimuth and elevation rotation axes.

In our proposed method, the *ML-RF* regressor is split into two layers, namely, marginalization weights and expert regression layer (shown in Fig. 3). A latent variable space is used to split the input data into a number of subsets, that are used to train Random Forest regressors in the expert regression layer. The posterior probabilities corresponding to each sample in the training set are acquired from each of these regressors. Our main contribution comes from the use of a marginalization weights regressor that learns the mapping of silhouette images to marginalization weights. We derive and apply a Kullback-Leibler divergence based optimization technique that estimates the marginalization weights for training data.

3.1. Assumptions

Our proposed hand orientation estimation method is targeted for a 2D monocular camera instead of a 3D depth camera, as most mobile devices have 2D monocular cam-



(a) (b)

Figure 3: (a) Flowchart for training and evaluation of the proposed marginalization through regression for hand orientation inference, and (b) latent variable space showing the different latent variables with different colors.

eras due to the consideration of power consumption, cost and form-factor. Most existing state-of-the-art methods utilize depth data, where the main focus is to infer detailed articulated hand pose [9, 11, 14]. Such methods are not suitable for a mobile scenario where, in addition to the absence of depth sensors, limited computational resources are available. The proposed method for hand orientation estimation assumes only 2D monocular cameras and limited computational resources are available and real-time performance is required. We use a planar hand pose and assume that a hand orientation can be represented with a single 3D normal vector. Skin and hand segmentation have a long history in computer vision and many techniques have been devised [21–23]. As our aim is to present a hand orientation inference method therefore we assume the segmentation problem is already solved.

3.2. Latent variables generation

We define a latent variable space to divide our dataset into subsets. This space is based on the simple observation that the hand orientation can be broadly categorized with respect to the camera as being: (i) fronto-parallel or facing (ii) right, (iii) left, (iv) upwards or (v) downwards, which also corresponds to the maximum distinctive hand shape variations.

Each set of *GT* orientation angles (ϕ_g, ψ_g) are first transformed into polar coordinates (γ, φ) and are then used to generate latent variables for dividing the target space into five different regions as

$$r_a = \begin{cases} r_1 & \text{if } \gamma \leq \alpha^\circ, \\ r_2 & \text{if } \gamma > \alpha^\circ : \varphi \in (0^\circ - \beta, 90^\circ - \beta], \\ r_3 & \text{if } \gamma > \alpha^\circ : \varphi \in (90^\circ - \beta, 180^\circ - \beta], \\ r_4 & \text{if } \gamma > \alpha^\circ : \varphi \in (180^\circ - \beta, 270^\circ - \beta], \\ r_5 & \text{if } \gamma > \alpha^\circ : \varphi \in (270^\circ - \beta, 360^\circ - \beta], \end{cases} \quad (1)$$

where α and β are adjustable parameters defining the radius of the central region and the offset for the latent variable space, respectively, and $r_a \in \{r_1, r_2, r_3, r_4, r_5\}$ are the latent variables dividing the dataset for *ML-RF* regression (Fig. 3(b)).

3.3. Expert regression layer

A set of multi-variate Random Forest regressors are trained on the subset of data to learn the mapping of hand silhouette images onto orientation angles. These regressors use a standard Random Forest regression method [24]. Given an input contour distance feature vector $\mathbf{d}$, the posterior probabilities for orientation angles (ϕ, ψ) for a given latent variable r_a are given by this layer as

$$p(\phi, \psi | r_a, \mathbf{d}) = \frac{1}{T} \sum_t p_t(\phi, \psi | r_a, \mathbf{d}), \quad (2)$$

where $p_t(\phi, \psi | r_a, \mathbf{d})$ is the posterior probability from leaf node of tree t and T are the total number of trees in a given Random Forest model. Then the marginalized posterior probability is defined as

$$p(\phi, \psi | \mathbf{d}) = \sum_a p(\phi, \psi | r_a, \mathbf{d}) \omega_a, \quad (3)$$

where ω_a are weights corresponding to each latent variable such that $\sum_a \omega_a = 1$.

3.4. Marginalization through regression

We formulate prior probability for the training samples using the *GT* orientation angles (ϕ_{gt}, ψ_{gt}) in a multi-variate normal distribution as

$$p\left(\phi_{gt}, \psi_{gt}\right) = \mathcal{N}\left(\left(\phi_{gt}, \psi_{gt}\right), \Sigma\right), \tag{4}$$

where Σ is the covariance that can be adjusted to control the spread of $p\left(\phi_{gt}, \psi_{gt}\right)$.

Given the prior probability $p\left(\phi_{gt}, \psi_{gt}\right)$ and the corresponding posterior probabilities $p\left(\phi, \psi | r_a, \mathbf{d}\right)$, we propose a novel optimization method, where the marginalization error is based on Kullback-Leibler divergence [25]. This error is optimized to estimate the *GT* marginalization weights ω_a for all latent variables $r_a \in \{r_1, r_2, r_3, r_4, r_5\}$. We define this error as

$$E = \iint p\left(\phi_{gt}, \psi_{gt}\right) \log \frac{p\left(\phi_{gt}, \psi_{gt}\right)}{p\left(\phi, \psi | \mathbf{d}\right)} d\phi d\psi. \tag{5}$$

Derivation Here we present the derivation of partial derivatives from Equation 5 that can be used to get optimal weights ω_a.

$$E = \iint p\left(\phi_{gt}, \psi_{gt}\right) \log \frac{p\left(\phi_{gt}, \psi_{gt}\right)}{p\left(\phi, \psi | \mathbf{d}\right)} d\phi d\psi, \tag{6}$$

$$= \iint p\left(\phi_{gt}, \psi_{gt}\right) [\log\ p\left(\phi_{gt}, \psi_{gt}\right) - \log \left[\sum_a p\left(\phi, \psi | r_a, \mathbf{d}\right) \omega_a\right]] d\phi d\psi. \tag{7}$$

The partial derivative w.r.t ω_a can then be defined as

$$\frac{\partial E}{\partial \omega_a} = -\iint \frac{p\left(\phi_{gt}, \psi_{gt}\right) p\left(\phi, \psi | r_a, \mathbf{d}\right)}{\sum_a p\left(\phi, \psi | r_a, \mathbf{d}\right) \omega_a} d\phi d\psi. \tag{8}$$

Optimization We use a standard gradient descent with

$$\nabla E = \left[\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \frac{\partial E}{\partial w_3}, \frac{\partial E}{\partial w_4}, \frac{\partial E}{\partial w_5}\right], \tag{9}$$

for which the optimization is iteratively evolved for a solution given by

$$\omega_a^{n+1} = \omega_a^n - \lambda \nabla E^n, \tag{10}$$

where λ is the step size along the negative gradient direction and n is the iteration number.

Marginalization weights regressor We use a multi-variate Random Forest regressor to learn the mapping of contour distance features to marginalization weights ω_a. This regressor is used during prediction to infer marginalization weights ω_a for marginalizing the posterior probabilities $p\left(\phi, \psi | r_a, \mathbf{d}\right)$ from each expert regressors using Equation 3.

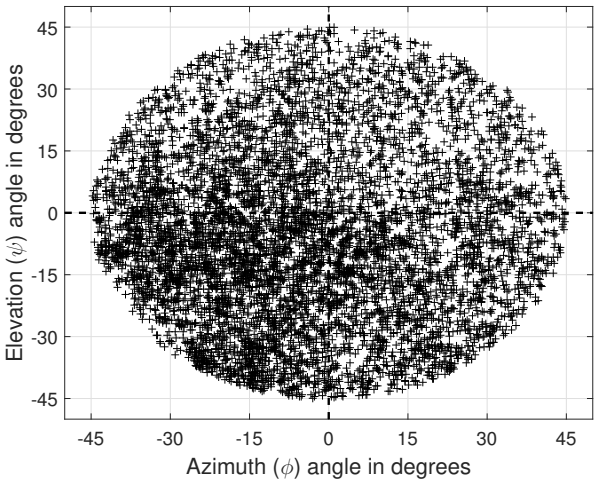


Figure 4: Orientation space plot showing the orientation angles captured by our dataset.

4. Experimental evaluation

To the best of our knowledge, hand orientation regression has only been proposed before in [4]. To this end, while some existing discriminative hand pose estimation methods have used quantized orientation of hands to achieve viewpoint invariance [11], a method for defining accurate hand orientation from 2D monocular images does not exist. Furthermore, the existing datasets for hand pose estimation do not provide accurate *GT* hand orientation [1, 4].

Our dataset contains 7059 samples collected for an open hand pose from 15 different participants. The range of the orientation angles captured by our dataset are restricted to a circular space with a radius of 45° (as shown in Fig. 4). This gives us an appropriate ratio for the number of samples against the variations within this defined orientation space. We show experimental results that demonstrate the ability of our proposed *ML-RF* regression method to apply marginalization through regression for estimating hand orientation on this dataset.

The proposed framework is compared with a previous method for hand orientation regression that uses a single-layered single-variate Random Forest (*SL-SV RF*) with independence assumption on each hand orientation angle. We also compare with three different methods for marginalization of *ML-RF* regressors [2, 8]. These methods are referred to as *ML-RF1*, *ML-RF2* and *ML-RF3* herein, adapted from [2] and [8]. While the methods proposed in [2] and [8] do not originally address hand orientation regression problem, they provide method for marginalizing a *ML-RF* in different domains. In our experimental validation, all three *ML-RF* comparison methods use a two-layered Random Forest with a coarse latent variable classification in the first layer and expert orientation regression in the second layer. These

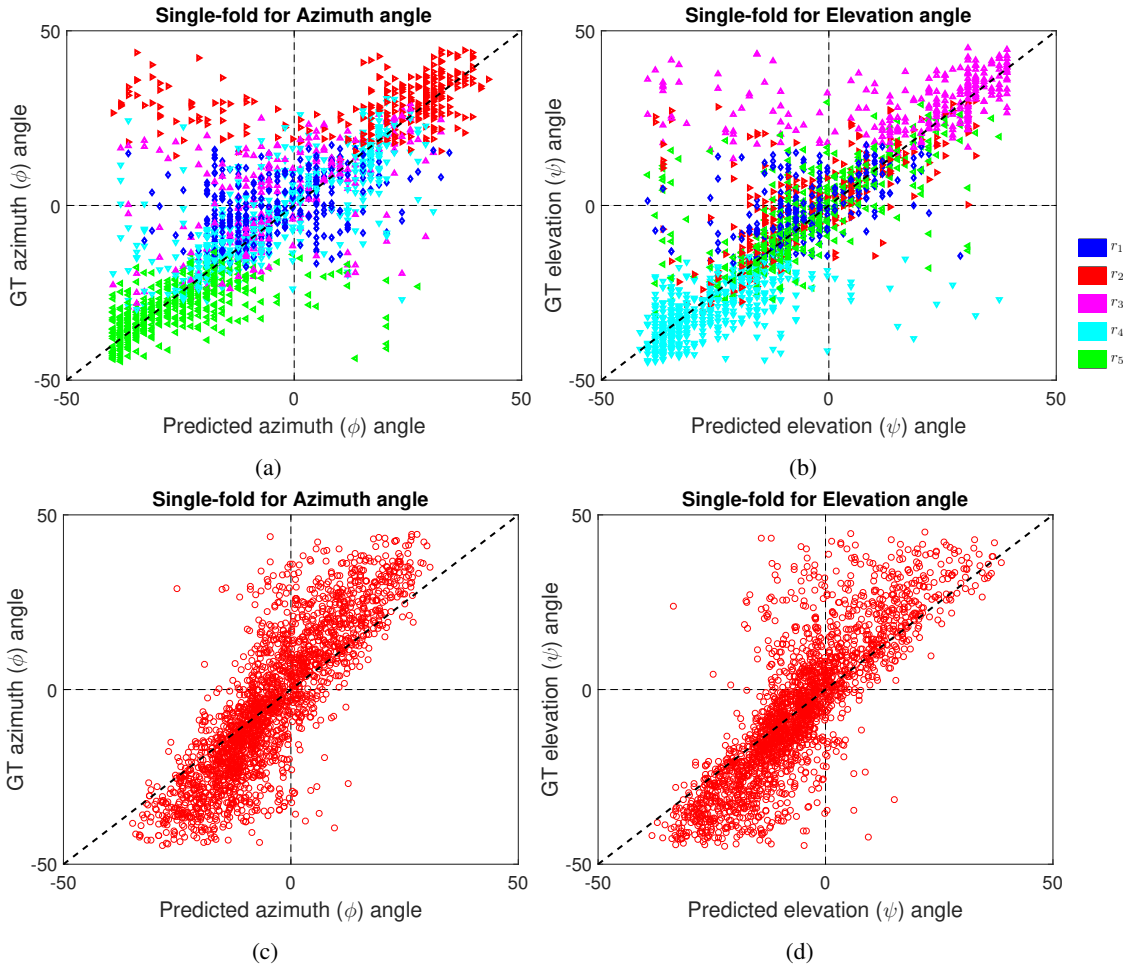


Figure 5: Single-fold validation shows *GT* vs predicted angle results for azimuth and elevation angles using (a)-(b) *ML-RF MtR* (proposed) and (c)-(d) *SL-SV RF* method [4]. It can be seen that the proposed *ML-RF MtR* method infers hand orientation angles without a bias, which is the main source of error in *SL-SV RF*.

methods only differ in marginalization where *ML-RF1* uses the predicted latent variable in the coarse layer to select the corresponding expert regressor for prediction. *ML-RF2* uses posterior probabilities of each latent variable in the coarse layer as marginalization weights for predicted angles from each expert regressor, whereas *ML-RF3* uses posterior probabilities from both the coarse and the expert layers to present the marginalized posterior probability. In this section, we denote our proposed **M**arginalization **t**hrough **R**egression method as *ML-RF MtR*.

4.1. Parameter selection

The proposed *ML-RF* regression has a number of different training parameters. These include the number of trees (T), depth of each tree (δ_t), minimum number of samples in each leaf node (η_j), the number of features selected at each split node (ϵ) and the parameters α and β defining the latent variable based label generation. In our experimental evaluation, we found that the parameters related to the *RF* classifier and regressors simultaneously improves the performance of all the comparison methods. Therefore, we empirically set these parameters to the following values for all experiments, $T = 100$, $\delta_n = 15$, $\eta_j = 5$ and $\epsilon = 1$, $\alpha = 15°$ and $\beta = 45°$.

4.2. Single-fold validation

For this experiment, the dataset is randomly divided with 70% of the data for training and the remaining 30% for testing. Fig. 5 presents the predicted orientation angles using our proposed *ML-RF MtR* method and *SL-SV RF* from [4]. These predicted angles are shown against their corresponding *GT* orientation angles, where in Fig.5(a)-(b) we also show the corresponding latent variables using colors from Fig. 3(b). Furthermore we also present the mean absolute

Table 1: Mean absolute error (*MAE*) in degrees for experiments in Section 4.

Evaluation method	Method used	Azimuth (ϕ)	Elevation (ψ)
Single-fold	*ML-RF MtR* (proposed)	$\mathbf{8.12}^\circ$	$\mathbf{7.36}^\circ$
	SL-RF SV [4]	9.43°	8.60°
	ML-RF1	8.80°	8.18°
	ML-RF2	11.31°	9.58°
	ML-RF3	8.69°	7.79°
User-specific	*ML-RF MtR* (proposed)	$\mathbf{7.89}^\circ$	$\mathbf{7.29}^\circ$
	SL-RF SV [4]	8.19°	7.94°
	ML-RF1	8.11°	7.45°
	ML-RF2	9.20°	8.50°
	ML-RF3	8.12°	7.72°

error (*MAE*) of all comparison methods in Table 1. The proposed *ML-RF MtR* method outperforms the state-of-the-art in hand orientation inference due to its ability to learn expert regressors on subsets of dataset. Furthermore, as opposed to training single-variate regressors for each orientation angles in [4], the proposed method utilizes multi-variate regressors to exploit the interdependence of orientation angles. From Fig. 5 we observe that the *ML-RF MtR* method is able to infer orientation angles without introducing any bias, which is the main source of errors in *SL-SV RF* [4] (as shown in Fig. 5(c)-(d)). Moreover, from Table 1, we note that the proposed *ML-RF MtR* method also outperforms the state-of-the-art in marginalization due to its ability to learn the marginalization weights with a regressor. Furthermore, as the marginalization weights are extracted using posterior probability distributions from expert regressors, therefore they also tend to address inaccuracies in the these posterior probabilities. In contrast, the state-of-the-art directly uses the posterior probabilities for marginalization which tend to underestimate the true posterior [10]. The errors in prediction come from symmetrically opposite latent variable spaces i.e. hand facing left/right or up/down, as can be seen in Fig. 5 (a)-(b) at around -40° and 40° GT orientation angles. This is due to the depth ambiguity of 2D silhouette images where two symmetric hand orientation produce similar results (as shown in Fig. 6). Nevertheless, these errors are few in number and do not affect the overall performance of our method as depicted in Table 1. Furthermore the symmetry problem can be solved by exploiting temporal coherence in a sequence of images using dynamic system models such as Kalman Filter or Particle Filter [26]. We aim to address this in our future work.

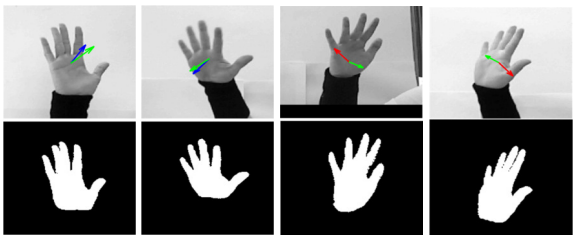


Figure 6: Success and failure cases with normal vector superimposed where green shows *GT* normal vectors, blue shows predicted normal vector for success cases and red shows predicted normal vector for failure cases. It can be seen that the predicted normal vectors for failure cases are symmetrically opposite to the GT normal vectors. Silhouette images show how different orientations can result in similar shape of hand.

4.3. User-specific validation

User-specific validation results of the proposed framework are shown in Table 1, where the training and testing is done using the same participant's data. This depicts an application scenario where a one-time model calibration will require the user to provide a user-specific hand orientation dataset. Once trained, our proposed approach would be able to infer the hand orientation. For this validation, we divide each participants data into training (70%) and testing (30%) sets. From Table 1 we see that the proposed method performs even better than single-fold validation, as now the marginalization is fine-tuned for a particular users hand where variations in shape and size are limited.

5. Conclusion

We proposed a novel marginalization method for multi-layered Random Forest regression of hand orientation. The proposed model was composed of two layers, where the first layer consisted of marginalization weights regressor while the second layer contained expert regressors trained on subsets of our hand orientation dataset. A latent variable space was used to divide the hand orientation dataset into subsets. We used a Kullback-Leibler divergence based optimization to estimate weights that marginalized posterior probabilities from each expert regressor against a *GT* prior probability. Our proposed marginalization weights regressor was trained on these weights that fine-tuned the marginalization of the posterior probabilities during on-line prediction. Our proposed method outperformed the state-of-the-art for both hand orientation inference and multi-layered Random Forest marginalization with an average error of 7.74° for single-fold validation and 7.59° for a user-specific scenario. The depth ambiguity in 2D hand silhouette images produced similar orientation inference for the symmetrically opposite orientations. We aim to address this in our future

work by exploiting temporal coherence using dynamic system models.

References

[1] James Steven Supancic III, Gregory Rogez, Yi Yang, Jamie Shotton, and Deva Ramanan, "Depth-based hand pose estimation: methods, data, and challenges," *arXiv preprint arXiv:1504.06378*, 2015.

[2] Sean Ryan Fanello, Cem Keskin, Shahram Izadi, Pushmeet Kohli, David Kim, David Sweeney, Antonio Criminisi, Jamie Shotton, Sing Bing Kang, and Tim Paek, "Learning to be a depth camera for close-range human capture and interaction," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, pp. 86, 2014.

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

[4] Muhammad Asad and Greg Slabaugh, "Hand orientation regression using random forest for augmented reality," in *Augmented and Virtual Reality*, pp. 159–174. Springer, 2014.

[5] Andrew K Palmer, Frederick W Werner, Dennis Murphy, and Richard Glisson, "Functional wrist motion: a biomechanical study," *The Journal of hand surgery*, vol. 10, no. 1, pp. 39–46, 1985.

[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] Martin de La Gorce and Nikos Paragios, "A variational approach to monocular hand-pose estimation," *Computer Vision and Image Understanding*, vol. 114, no. 3, pp. 363–372, 2010.

[8] Matthias Dantone, Juergen Gall, Gabriele Fanelli, and Luc Van Gool, "Real-time facial feature detection using conditional regression forests," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2578–2585.

[9] Cem Keskin, Furkan Kıraç, Yunus Emre Kara, and Lale Akarun, "Hand pose estimation and hand shape classification using multi-layered randomized decision forests," in *Computer Vision–ECCV 2012*, pp. 852–863. Springer, 2012.

[10] S. Hallman and C. C. Fowlkes, "Oriented edge forests for boundary detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

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

[12] Yoshiaki Mizuchi, Yoshinobu Hagiwara, Akimasa Suzuki, Hiroshi Imamura, and Yongwoon Choi, "Monocular 3d palm posture estimation based on feature-points robust against finger motion," in *International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2013, pp. 1014–1019.

[13] 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.

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

[15] Iasonas Oikonomidis, Nikolaos Kyriazis, and Antonis A Argyros, "Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints," in *2011 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2011, pp. 2088–2095.

[16] Toby Sharp, Cem Keskin, Duncan Robertson, Jonathan Taylor, Jamie Shotton, David Kim, Christoph Rhemann, Ido Leichter, Alon Vinnikov, Yichen Wei, et al., "Accurate, robust, and flexible real-time hand tracking," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 3633–3642.

[17] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[18] Rómer Rosales and Stan Sclaroff, "Combining generative and discriminative models in a framework for articulated pose estimation," *International Journal of Computer Vision*, vol. 67, no. 3, pp. 251–276, 2006.

[19] Robert Y Wang and Jovan Popović, "Real-time hand-tracking with a color glove," in *ACM Transactions on Graphics (TOG)*. ACM, 2009, vol. 28, p. 63.

[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] Cheng Li and Kris Kitani, "Pixel-level hand detection in ego-centric videos," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3570–3577.

[22] M. J Jones and J. M Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[23] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*. Moscow, Russia, 2003, vol. 3, pp. 85–92.

[24] Antonio Criminisi and Jamie Shotton, *Decision forests for computer vision and medical image analysis*, Springer Science & Business Media, 2013.

[25] Solomon Kullback and Richard A Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[26] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.