CITY UNIVERSITY LONDON

IEEE 2016 Conference on Computer Vision and Pattern Recognition

CVPR 2016

Learning Marginalization through Regression for Hand Orientation Inference

Muhammad Asad and Greg Slabaugh
{muhammad.asad.2, gregory.slabaugh.1}@city.ac.uk

1. Introduction

- Real-time depth cameras have limited availability on mobile devices due to power consumption, cost and form-factor considerations
- Accurate hand orientation inference using 2D monocular cameras enables orientation based interactions in augmented reality

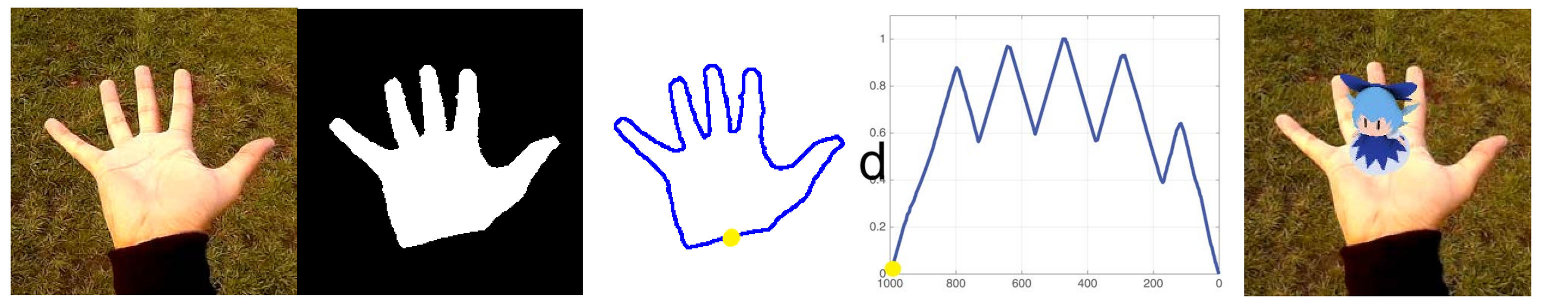

Figure 1: The proposed method uses contour distance features (center), extracted from hand silhouette (left), to infer hand orientation. This enables interaction with augmented object (right)

2. Contribution

- Method for marginalization through regression of a Multi-Layered Random Forest (ML-RF) regressor with two layers, namely, **Expert Regression** and **Marginalization Weights Regression** layers
  - Provides better coupling of multiple layers in a ML-RF regressor

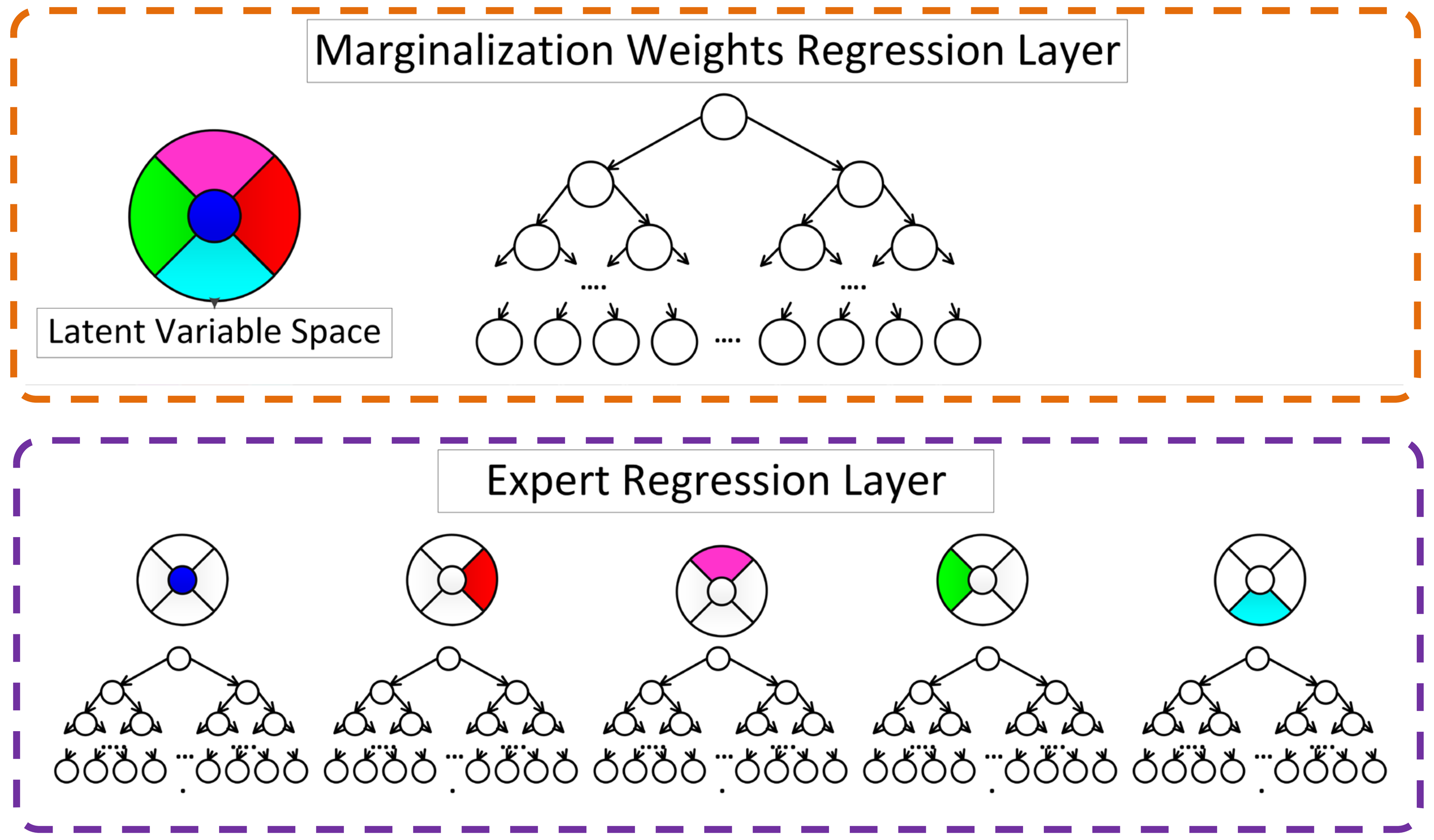

Figure 2: The proposed multi-layered marginalization through regression method.

3. Method

- **Expert Regression Layer** consists of regressors independently trained on subsets of orientation dataset defined by a latent variable r_a
- Given an input feature vector $\mathbf{d}$, the posterior probabilities for hand orientation angles (ϕ, ψ) from this layer are marginalized as

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

- The ground truth (GT) marginalization weights ω_a are estimated by the optimization of a Kullback-Leibler divergence based error between posterior and prior probability distributions

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

- For optimization, gradient descent is used with partial derivative of E w.r.t ω_a

$$\frac{\partial E}{\partial \omega_a} = -\iint \frac{1}{\sum_a p(\phi,\psi|r_a,\mathbf{d})\omega_a} p\left(\phi_{gt}, \psi_{gt}\right) p\left(\phi, \psi|r_a, \mathbf{d}\right) d\phi d\psi$$

- **Marginalization Weights Regressor** is trained against the GT marginalization weights ω_a

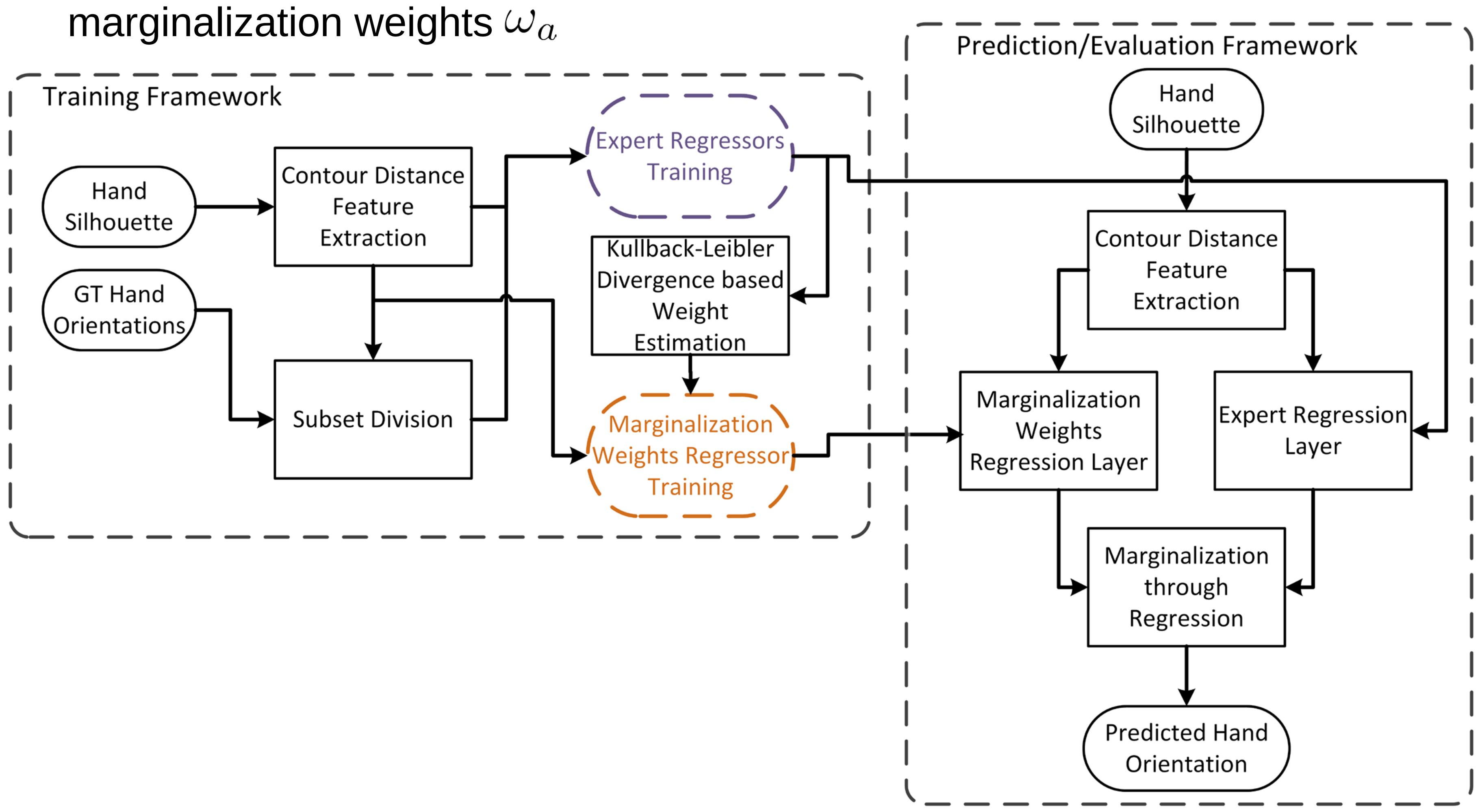

Figure 3: Flowchart of the proposed marginalization through regression method

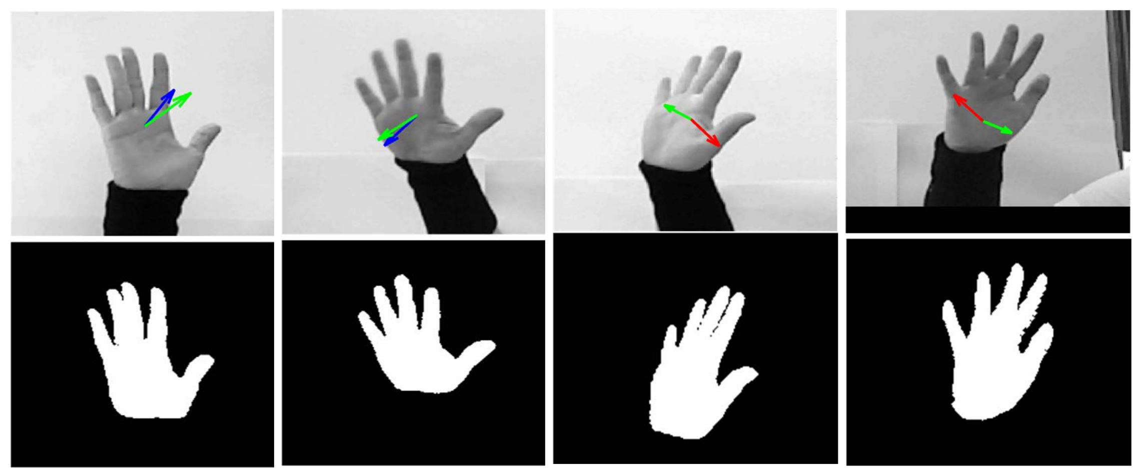

Figure 4: Success and failure cases. Green: GT, blue: success and red failure. It can be seen that failure cases are symmetrically opposite.

4. Results

- The collected dataset contains 7059 samples for an open hand pose from 15 different participants, where the orientation angles are restricted to a circular space of radius 45˚
- Single-fold and user-specific validations are used for comparison with [1] and three ML-RF methods adapted from [2, 3].

Evaluation method	Method used	Azimuth (ϕ)	Elevation (ψ)
Single-fold	*ML-RF MtR* (proposed)	**8.12°**	**7.36°**
	SL-RF SV [1]	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)	**7.89°**	**7.29°**
	SL-RF SV [1]	8.19°	7.94°
	ML-RF1	8.11°	7.45°
	ML-RF2	9.20°	8.50°
	ML-RF3	8.12°	7.72°

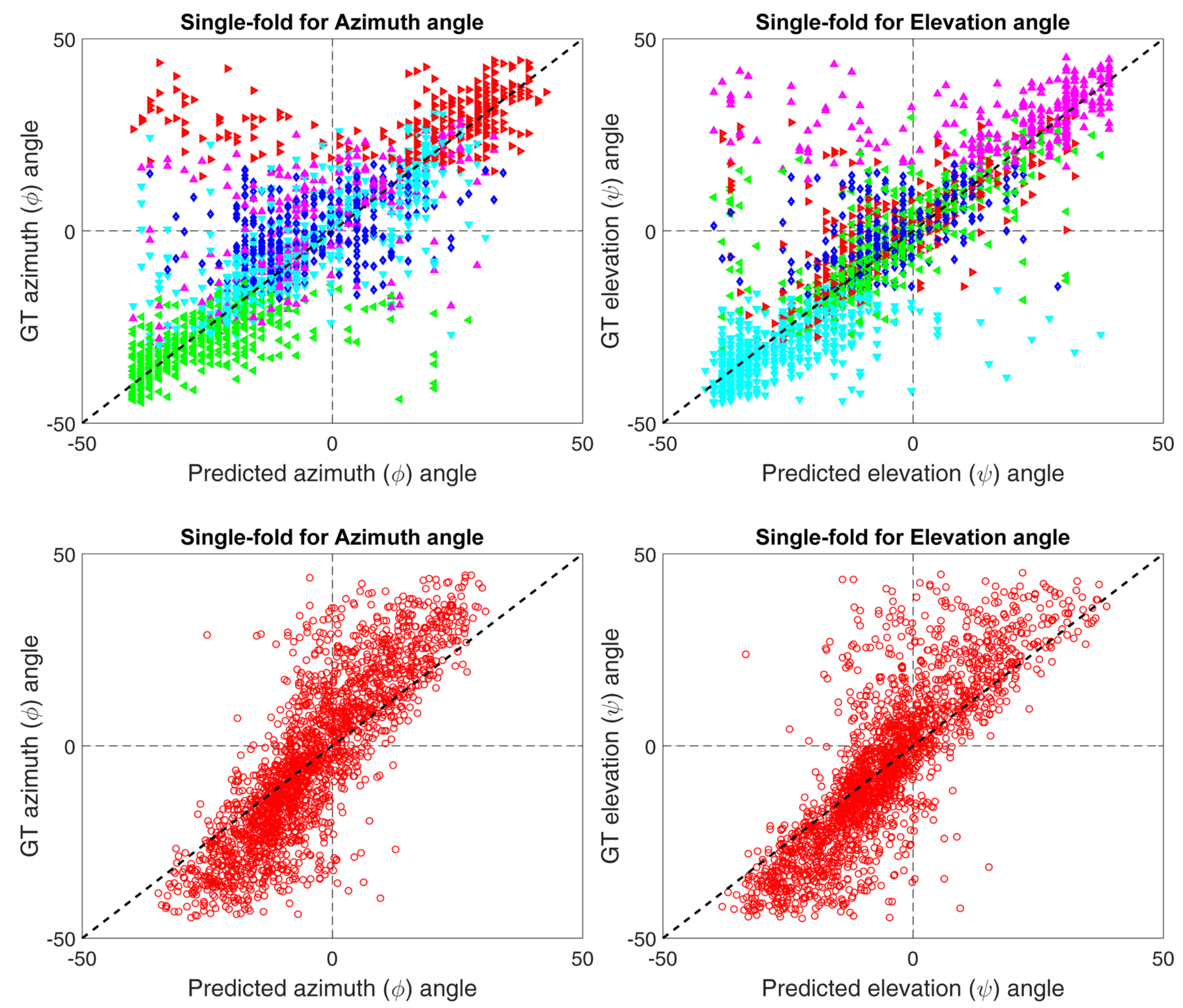

Figure 5: Experimental results showing, (top) table with mean absolute error for azimuth and elevation angles and (bottom) GT vs predicted orientation angles plots using the proposed method and the method from [1].

5. Future work

- Our future work aims at
  - exploiting temporal coherence to improve the accuracy
  - using the inferred hand orientation to simplify hand pose estimation

References

[1] Asad et al. "Hand orientation regression using random forest for augmented reality." Augmented and Virtual Reality. Springer International Publishing, 2014.
[2] Fanello et al. "Learning to be a depth camera for close-range human capture and interaction." ACM Transactions on Graphics (TOG) 33.4 (2014): 86.
[3] Dantone et al. "Real-time facial feature detection using conditional regression forests." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.