Localization of Skin Features on the Hand and Wrist from Small Image Patches

Lee Stearns*, Uran Oh*, Bridget J. Cheng†, Leah Findlater*, David Ross‡, Rama Chellappa*, Jon E. Froehlich*
* University of Maryland, College Park, Maryland, United States
† Cornell University, Ithaca, New York, United States
‡ Atlanta VA R&D Center for Vision & Neurocognitive Rehabilitation, Atlanta, Georgia, United States

Abstract*—Skin-based biometrics rely on the distinctiveness of skin patterns across individuals for identification. In this paper, we investigate whether small image patches of the skin can be localized on a user's body, determining not "who?" but instead "where?" Applying techniques from biometrics and computer vision, we introduce a hierarchical classifier that estimates a location from the image texture and refines the estimate with keypoint matching and geometric verification. To evaluate our approach, we collected 10,198 close-up images of 17 hand and wrist locations across 30 participants. Within-person algorithmic experiments demonstrate that an individual's own skin features can be used to localize their skin surface image patches with an F_1 score of 96.5%. As secondary analyses, we assess the effects of training set size and between-person classification. We close with a discussion of the strengths and limitations of our approach and evaluation methods as well as implications for future applications using a wearable camera to support touch-based, location-specific taps and gestures on the surface of the skin.

***Keywords*— skin texture classification; biometrics and computer vision applications; on-body input**

I. Introduction

Previous work in skin classification has largely been in the context of biometrics—that is, determining the *uniqueness* of a user's skin patterns for identification purposes (*e.g.,* [1–7]). In this paper, rather than identifying *who* an image represents, we seek to identify *where* an image is located on a single user's body. More specifically, we investigate to what extent are surface image patches of the hand and wrist localizable?

Being able to determine the location of a small patch of skin could enable a wearable camera to support a range of *on-body interactions*, an emerging paradigm in human-computer interaction (HCI) where users tap or gesture on their own body to control a computing device (*e.g.*, [8–16]). One advantage is that this type of input is always available, allowing the user to, for example, quickly tap or swipe on their palm to answer a phone call or listen to new emails (Figure 1a). On-body interaction is also useful when visual attention is limited because the skin's tactile perception allows for more accurate input than is possible with a touchscreen [17, 18].

Sensing these on-body taps and gestures, however, is a challenging problem. Researchers have investigated a variety of wearable cameras (*e.g.*, [11, 19]) and other sensors (*e.g.*, bio-acoustics [8], ultrasonic rangefinders [9]). While promising, these approaches are limited by the placement and range of the sensor [14, 19], suffer from occlusion [19] or precision [8] problems, or cover the user's skin [15], reducing tactile sensitivity. Instead, we envision using close-up images from a small *finger-mounted camera* (*e.g.*, [20, 21]) to sense and localize user input (Figure 1b). By instrumenting the gesturing finger with a camera, our approach extends the user's interaction space to anything within reach and can support fairly precise location-based input.

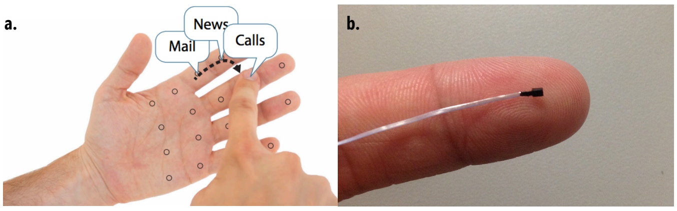

Figure 1. (a) Conceptual visualization of on-hand input to control a mobile phone, as in [17]. (b) Cameras developed for minimally invasive surgeries are small enough to mount on the finger. Shown: AWAIBA NanEye ($1\times1\text{mm}^2$, 250×250px resolution) used in [20, 21].

Localizing small (~1–2 cm) image patches within the larger skin surface is similar to partial finger and palm print recognition in forensic applications; however, high-resolution, high-contrast images of ridge impressions are typically needed to reliably extract distinctive point and line features. In contrast, cameras small enough to be mounted on the finger (Figure 1b) are low resolution and low contrast, making it difficult to detect minute ridge features. Several recent biometric systems recognize finger and palm prints using lower-quality images [2, 5–7, 22–27]. Unfortunately, these approaches are frequently designed to align and process the finger or palm image as a whole, and cannot reliably recognize a small portion of the print. To our knowledge, no work has attempted to recognize or localize a small skin patch from live camera images, which we do here.

To ultimately support on-body localization using a finger-mounted camera, we investigate the classifiability of 17 locations on the front and back of the palm, fingers, wrist, nails, and knuckles. We introduce a hierarchical texture classification approach to first estimate the approximate touch location on the body given close-up images of the skin surface and then refine the location estimate using keypoint matching and geometric verification. To evaluate our approach, we collected a skin-surface image dataset consisting of 30 individuals and the 17 hand and wrist locations (10,198 total images).

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617

When testing and training on an individual's own skin data (within-person experiments), our results show that skin patches are classifiable by location under controlled conditions with 96.6% recall and 96.4% precision, suggesting that finger-mounted cameras may be feasible for sensing on-body interactions.

In summary, the contributions of this paper include: (i) a robust algorithmic pipeline for recognizing several different locations on the hand from small patches of skin; (ii) classification results for a dataset consisting of 30 individuals, achieving accuracy above 96% on average for within-person experiments; and (iii) analysis of hand distinctiveness and similarities among users, which may impact accuracy and scalability (*e.g.,* between-person training feasibility).

II. Related Work

Our work applies and extends research in biometrics and on-body interaction, which we describe below.

A. Finger and Palm Biometrics

Work in biometrics has demonstrated that the skin of the hand—specifically, the palm and fingers—contains a large number of highly distinctive visual features that can be used to identify individuals [1, 2]. As noted in the Introduction, we borrow and extend biometrics algorithms, including those for partial fingerprint and palmprint recognition used in forensics [3, 4] as well as techniques that use relatively low resolution and low contrast webcam and mobile phone camera images for person identification and verification [2, 5–7].

Fingerprint and palmprint recognition requires matching an image against a set of stored templates, either directly using the image intensities [22], using texture representations [23, 24], or using minutiae and other point features [3, 4, 6, 27, 28]. A direct match requires precise alignment of the finger or palm as a whole, which is infeasible for the close-up, partial images used in our work. However, we can support partial matching by applying some of the same methods for preprocessing, representing texture, and extracting point features.

To preprocess the images, most approaches apply some filter (*e.g.,* Gabor filters [23]) that enhances the skin's ridges and principal lines. We use a method inspired by Huang *et al.*'s palmprint verification work [29]. To represent texture, Local Binary Pattern (LBP) histograms are a common choice in biometrics [24, 30]. While we explored other texture-based methods, such as Gabor histograms [25] and wavelet principal components [26], we found that they offered negligible improvements over LBP despite their increased computational complexity. To detect point features, some systems use Harris corners [27] or scale-invariant feature transform (SIFT) key points [7, 27]; however, the comparisons between feature descriptors are challenging due to the repetitive nature of finger and palm images. Thus, we instead extract custom features that are based upon Gabor filter response, inspired by [29].

B. On-Body Interaction

A wide variety of wearable sensors have been used to support touch-based input on the user's own body, from arm-worn bio-acoustic sensors [8] and ultrasonic rangefinders [9] to infrared reflectance sensors [14] and touch-sensitive skin overlays [15]; these approaches, however, are limited by occlusion and precision problems, sensor placement and range issues, and/or impeding the user's sense of touch—as noted in the Introduction. Most relevant are approaches that use body-mounted cameras—often depth and/or infrared cameras worn on the chest, shoulder, or head [11–13, 16, 19, 31]. These devices detect on-body gestures from camera images using low-level computer vision techniques such as histogram [19] and motion-based segmentation and tracking [31], and Support Vector Machine (SVM)-based classifiers [31]. For instance, Harrison *et al.*'s [11] *OmniTouch* enables touch-based interactions on a variety of surfaces via a shoulder-mounted depth camera for hand tracking and a pico-projector for visual feedback. While similar to using a finger-mounted camera, body-mounted camera hardware is subject to occlusion issues and limits the interaction space to the *fixed* camera's field of view [19].

In contrast, we envision using a finger-mounted camera that can be freely pointed to interaction targets. Such cameras have been used by *FingerReader* [32] and *HandSight* [21, 33] for reading printed text, and by *Magic Finger* [20], which combines a high-speed optical mouse sensor to track finger movement and a slower but higher-resolution camera to capture detail for texture classification. We were particularly inspired by Magic Finger, which achieved high accuracy (99%) in classifying a set of 22 textures that included table surfaces, white paper, clothing, and two body locations (hand, thumb). The first stage of our surface classification approach is similar to theirs, using an LBP texture representation and an SVM classifier.

III. Touch Localization Pipeline

Robust localization of close-up skin images from a finger-mounted camera is challenging due to the limited field of view (~1–2 cm) and relatively low contrast of the ridges and other skin surface features. To estimate the user's touch location from close-up images, we developed a hierarchical classifier with four stages: (i) preprocessing, (ii) coarse-grained classification, (iii) fine-grained classification, (iv) geometric verification and refinement. The coarse-grained stage classifies an input image into one of five regions: *palm*, *fingers*, *nail*, *knuckle*, and *other* (wrist and back of hand). The fine-grained stage further classifies the image into a discrete location within that region (17 locations in all; see Figures 2 and 5). These locations were selected because previous work has shown that users can reliably locate them with high accuracy even without sight [18]. While our four-stage pipeline integrates multiple known approaches in fingerprint and palmprint enhancement, texture classification, and 2D keypoint matching, our primary innovation is in their novel combination and application towards *localization* rather than identification.

Stage 1: Preprocessing. Images are first preprocessed to remove noise and emphasize ridge features. We apply an efficient median filter [34] to reduce the effect of dirt and other camera noise while preserving the edge information necessary for processing finger and palm prints (Figure 2).

To emphasize the ridgelines, we adapt a technique from Huang *et al.* [29]. However, while they use a modified version of the finite radon transform to emphasize the principal lines and creases of the palm, these features are not as prominent in our images due to the narrow field of view. We instead use Gabor

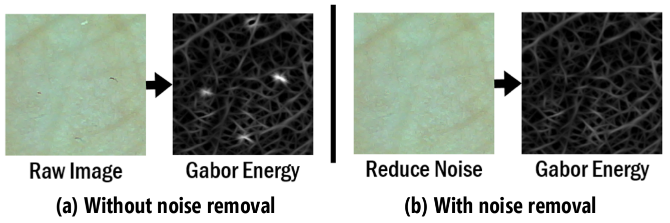

Figure 2. Stage 1 preprocessing first removes dirt and other noise before emphasizing ridge features using the energy of a set of Gabor filters with different orientations. Shown: an example image from the left side of the palm, scaled and cropped to demonstrate the effect that surface artifacts can have on the Gabor energy image.

filters. We compute the Gabor energy image defined as the maximum response at each pixel from a set of Gabor filters with different orientations. Specifically, the energy at pixel location (*x, y*) is:

$$E_{x,y} = \left|\max_{\theta}\left[G_{\theta} * \left(\bar{I}_{x,y} - I_{x,y}\right)\right]\right| \tag{1}$$

where $I_{x,y}$ is the gray-scale pixel value at (*x, y*) and $\bar{I}_{x,y}$ is the local mean in a window around that location (estimated using a Gaussian smoothing function), G_θ is a discrete Gabor filter with orientation θ, and $*$ is the convolution operator. In our experiments, we use 18 uniformly distributed orientations, with a fixed scale and bandwidth that were chosen empirically based upon the average ridge frequency in our preliminary experiments with a separate set of pilot data. Example energy images are shown in Figure 2, 3, and 4.

Stage 2: Coarse-Grained Classification. After preprocessing, we obtain a rough classification of the image's location using the visual texture, which we represent using LBP histograms. We chose LBP because of its computational efficiency and natural invariance to illumination variations. To improve accuracy and achieve rotation invariance, we use only the uniform patterns alongside the variance of the neighboring values as suggested in [35]. Our implementation uses a 2D histogram with 14 uniform pattern bins and 12 variance bins ($LBP_{12,2}^{riu2}$ and $VAR_{12,2}$, as defined in [35]), computed at 3 scales. These parameters were selected because they provided a balance between classification accuracy and computational efficiency on our pilot data. The histograms for each scale are flattened and concatenated together to produce a 672-element feature vector, which is then normalized. To classify the LBP histograms into coarse-grained body regions, we train a support vector machine (SVM)—commonly used in texture classification (*e.g.*, [20, 22, 36]).

Stage 3: Fine-Grained Classification. We compare the LBP histogram using a template matching approach against *only* the training templates from the coarse-grained region identified in Stage 2. This hierarchical approach reduces the number of possible match locations and enables us to prioritize different features for each region individually (*e.g.*, for the palm we can automatically weight the palmprint texture features that best discriminate the five different palm locations). For template comparisons, we use the χ^2 distance metric, which is known to perform well with LBP histograms (*e.g.,* [37]). Stage 3 produces a sorted list of templates, with the lowest distance representing the most likely match.

Stage 4: Geometric Verification and Refinement. Stage 4 ensures the validity of the texture match and refines the precise touch location using a set of keypoint matching and geometric

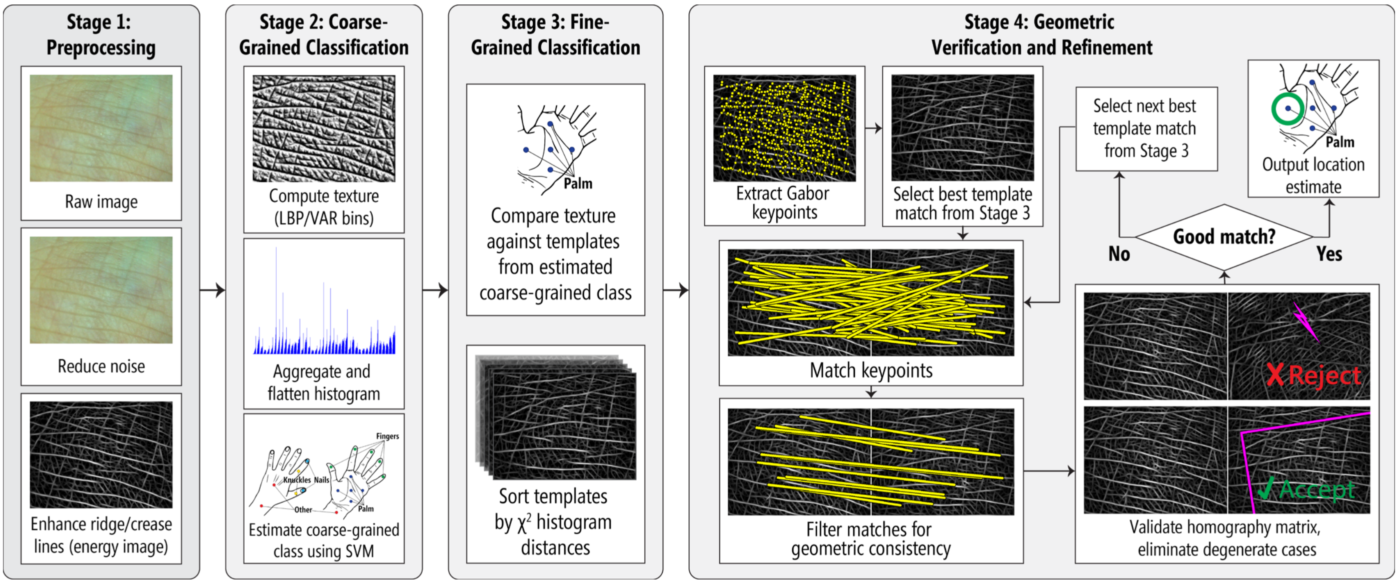
Stage 1: Preprocessing
Raw image
Reduce noise
Enhance ridge/crease lines (energy image)
Stage 2: Coarse-Grained Classification
Compute texture (LBP/VAR bins)
Aggregate and flatten histogram
Fingers
Knuckles
Nails
Other
Palm
Estimate coarse-grained class using SVM
Stage 3: Fine-Grained Classification
Palm
Compare texture against templates from estimated coarse-grained class
Sort templates by χ² histogram distances
Stage 4: Geometric Verification and Refinement
Extract Gabor keypoints
Select best template match from Stage 3
Select next best template match from Stage 3
Palm
Output location estimate
Good match?
No
Yes
Match keypoints
Filter matches for geometric consistency
✗Reject
✓Accept
Validate homography matrix, eliminate degenerate cases

Figure 3. The four stages of our localization algorithm, as applied to an example image from the left side of the palm. First, the image is preprocessed to remove surface artifacts and camera noise before calculating the Gabor energy to emphasize ridge and crease lines. Second, the image is classified into one of five coarse-grained locations (in this case, the palm) using a 2D texture histogram of LBP and pixel variances. Third, the image's texture is compared against the templates from the predicted coarse-grained class, which are sorted by their χ^2 histogram distances to prioritize matching for the next stage. Finally, the image is compared geometrically against images from the predicted coarse-grained class, using a set of custom Gabor keypoints and descriptors. The image is compared against individual templates starting with the most likely match (as predicted in Stage 3), proceeding in order until a template with sufficient geometrically consistent keypoint matches is found. If a geometrically consistent match is found, then the fine-grained location can be estimated with a high degree of certainty (in this case, the left side of the palm); otherwise, the algorithm falls back upon the closest texture match from Stage 3.

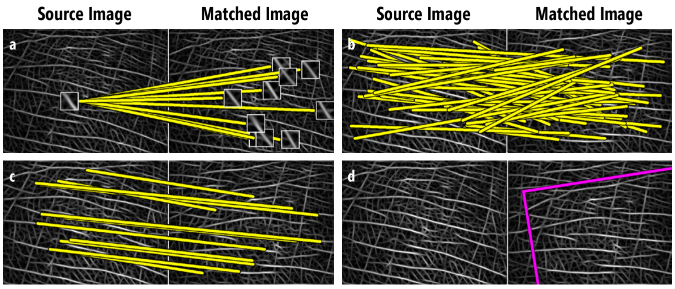

Figure 4. Keypoints in the Gabor energy images frequently appear visually similar (a), leading to a high percentage of mismatches (b). We filter outliers using a series of verification steps to ensure geometric consistency (c and d).

verification steps. We investigated SIFT keypoints [6, 7, 27], Harris corners [27], and fingerprint minutiae [3, 4, 28], but found them too unreliable in preliminary tests. Instead, we use keypoints with a high Gabor filter response at two or more orientations, which tend to lie at the intersections of ridgelines or creases. The Gabor energy values in the 16×16px neighborhood surrounding the keypoint serve as a reliable descriptor. To achieve rotation invariance, we generate multiple descriptors at each keypoint location, rotating the neighborhood for each using the orientation of the filters with locally maximum response strength. We keep a list of keypoints for each training image.

These image patches, however, are frequently visually similar (*e.g.*, Figure 4a), leading to a high percentage of mismatches between the keypoints (Figure 4b). We address this issue using a series of geometric verification steps. First, we filter the matches for orientation consistency, eliminating matches that do not agree with the majority vote for the relative rotation between images (*i.e.*, any more than 20° from the average rotation across all matches). Second, we compute a homography matrix using random sample consensus (RANSAC), identifying inliers and ensuring that there are sufficient geometrically consistent feature matches (*i.e.,* more than the minimum necessary to define a homography; in our experiments, we required 16 consistent matches). Although the palm and fingers are not rigid planar surfaces, in the close-up images we gathered they appear nearly so; we compensate for any irregularities by allowing a greater than usual inlier distance of 10 pixels. Third, we verify that the homography matrix is well behaved using the following constraints, which ensure that the match preserves orientation and does not have extreme variations in scale or perspective:

$$\begin{array}{ll} 1. & H_{11}H_{22} - H_{21}H_{12} > \frac{1}{2} \\ 2. & \frac{1}{2} < \sqrt{{H_{11}}^2 + {H_{21}}^2} < 2 \\ 3. & \frac{1}{2} < \sqrt{{H_{12}}^2 + {H_{22}}^2} < 2 \\ 4. & \sqrt{{H_{31}}^2 + {H_{32}}^2} < \frac{1}{1000} \end{array} \qquad (2)$$

These constraints were selected empirically to eliminate most degenerate cases that could lead to false-positive matches. Fourth and finally, to avoid further degenerate cases, we ensure that the inlier features are not collinear and that they have sufficient spread. We do this by calculating the standard deviation along the two principal directions computed using principal component analysis; if $\sigma_1 < 25$ or $\sigma_1/\sigma_2 > 4$, we declare the match invalid (these numbers were also selected empirically and validated on a separate set of pilot data). If a template match is declared invalid, we proceed to the next best texture match, stopping once we find one that passes all conditions. If a valid match is not found, then we fall back upon the best Stage 3 texture match.

Participant Demographics

Gender	23 female, 7 male	
Age	Mean = 30.6, SD = 11.5, Min = 18, Max = 59	
Race	Black, Afro-Caribbean, or Afro-American	6
	East Asian or Asian-American	5
	Latino or Hispanic American	1
	Non-Hispanic White or Euro-American	14
	South Asian or Indian American	2
	Other or Multiple	2
Palm Size	Mean = 98.3 mm, SD = 10.3 mm, Min = 79.7 mm, Max = 129.5 mm	

Table 1. Our dataset captures variations in gender, age, race, and palm size. Palm size was measured diagonally from the base of the thumb to base of the smallest finger while the fingers were spread and fully extended.

The output of our hierarchical algorithm is an estimated classification of a query image into one of 17 locations, along with a confidence score based upon the texture similarity and the number of inliers for the best template match. From the computed homography matrix, we also obtain a more precise location estimate relative to the matched training templates, potentially enabling finer localization for future explorations.

IV. Data Collection and Dataset

To evaluate our approach, we created an image dataset collected from 30 volunteers (23 female) recruited via campus email lists. The participants were on average 30.6 years old

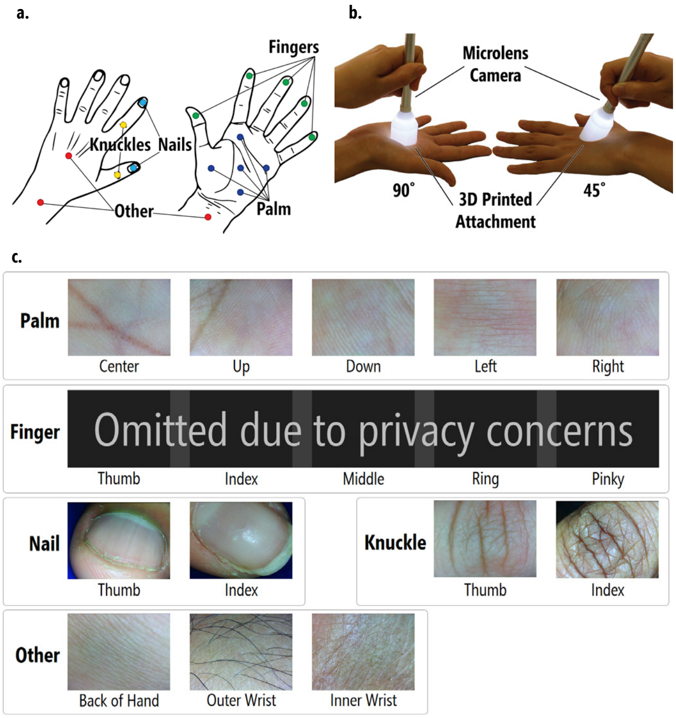

Figure 5. Data collection setup: (a) the 17 close-up image locations on the left hand in 5 coarse-grained regions–coded with different colors; (b) the pen-based camera and physical constraints (one angled at 45° and one at 90°) used for close-up image capture. (c) representative images from our dataset for each of the 17 locations, selected across 12 participants.

(*SD*=11.5, *range*=18–59), and represented a variety of skin tones and palm sizes (Table 1). For each participant, we collected close-up images of 17 locations (Figure 5) using a small 0.3-megapixel micro-lens camera in the shape of a pen.

The micro-lens camera is self-illuminated with a manually adjustable focal length, enabling us to capture clean 640×640px images of the hand from as close as 1cm. We controlled for distance and perspective using two 3D-printed camera attachments that place the camera approximately 2.5cm from the surface of the hand, at either a 90° or 45° angle (Figure 5b). Compared to a finger-mounted camera, this form factor enabled us to more easily control for variables such as distance, perspective, focus, and lighting, while still capturing images that are representative of our target domain. Ultimately we expect to use a smaller camera similar to Figure 1b.

Participants used the camera to point to 17 locations on the hand and palm, with 10 trials for each location and two perspectives (45° and 90°) resulting in 340 images per person. Rather than point 10 times in a row to the same location, the order of trials was randomized to provide natural variation in translation, rotation, and pressure (which impacts scale and focus). In total, we have 10,198 close-up micro-lens images across the 30 participants (one participant skipped two trials). While we would like to release this dataset publicly, we cannot do so without risking the privacy of our participants.

V. Experiments and Results

We first describe results related to coarse- and fine-grained hand classification performance before presenting secondary analyses related to the effect of training sample size on performance and between-person classification. Our analyses report standard measures including precision, recall, and F_1 scores. These metrics are more informative than accuracy due to the uneven number of training examples per class our hierarchy defines.

A. Within-person Classification

To evaluate the overall location-level classifiability of the hand, we conducted a within-person classification experiment. We used an *n*-fold, leave-one-out cross-validation approach. Our results are the average across all 20 folds for each of the 30 participants. We first present aggregate results before examining performance by location and by participant.

Stage 2: Coarse-grained Classification Confusion Matrix

	Palm	Finger	Nail	Knuckle	Other
Palm	**99.0%**	0.5%			0.5%
Finger	0.6%	**99.3%**	0.1%		
Nail	0.2%	0.1%	**99.7%**	0.1%	
Knuckle			0.2%	**99.1%**	0.7%
Other	0.6%		0.1%	0.5%	**98.8%**

Table 2: Classification percentages for classes at the coarse-grained level. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row).

At the coarse-grained level (Stage 2), the average precision is 99.1% (*SD*=0.9%) and average recall is 99.2% (*SD*=0.8%). At the fine-grained level (Stage 3), the average precision is 88.2% (*SD*=4.4%) and recall is 88.0% (*SD*=4.5%). After performing geometric validation and refinement (Stage 4), fine-grained classification increases to 96.6% precision (*SD*=2.2%) and 96.4% recall (*SD*=2.3%). The high precision and recall values demonstrate the feasibility of using close-up images to classify locations on the hand and wrist. Stage 2 precision and recall are very high (above 99%), which is important because errors in estimating the coarse-grained region will propagate to the next stage (a limitation of our hierarchical approach). Across all stages, we observed classification errors that were caused primarily by similarities between the locations' visual textures, poor image quality, and insufficient overlap between the training and testing images, although the high accuracies meant that there was not enough data for statistical analysis of the errors.

To examine the impact of different hand/wrist locations on performance, we created confusion matrices for Stage 2 (coarse-grained) and Stage 4 (fine-grained) classifications. See Tables 2 and 3 respectively. The locations with the lowest F_1 score were those on the back of hand (*M*=92.3%; *SD*=10.1%) and wrist (*M*=91.8%; *SD*=8.4%), which appear visually similar (Figure 6). This was true to a lesser extent across all coarse-grained regions, with the textures of different locations within each region appearing similar. While Stage 4 geometric validation reduced misclassifications, it was not always successful. For example, in some cases, an image for a participant did not sufficiently overlap any other image in the dataset, preventing

Stage 4: Fine-grained Classification Confusion Matrix

	Palm					Fingers					Nails		Knuckles		Other		
	C	U	D	L	R	1st	2nd	3rd	4th	5th	1st	2nd	1st	2nd	BH	OW	IW
Palm Center (C)	**98.3%**					0.2%		0.2%		0.2%					0.2%	0.5%	0.5%
Palm Up (U)	0.2%	**98.5%**		0.2%	0.2%		0.2%			0.3%						0.2%	0.3%
Palm Down (D)	0.3%	1.2%	**95.7%**	0.2%	1.7%	0.3%	0.2%						0.2%			0.2%	0.2%
Palm Left (L)	0.3%	0.3%	0.2%	**98.7%**	0.3%	0.3%											0.7%
Palm Right (R)	0.7%	0.5%	0.3%	0.5%	**97.5%**	0.2%	0.2%			0.2%							
1st Finger		0.5%	0.2%	0.2%	0.7%	**96.3%**	0.3%	0.5%	0.5%	0.7%					0.2%		
2nd Finger		0.3%			0.2%	0.3%	**95.8%**	1.7%	0.5%	1.2%							
3rd Finger			0.3%			0.2%	1.3%	**95.4%**	2.2%	0.7%							
4th Finger			0.2%				0.3%	1.8%	**95.3%**	2.3%							
5th Finger		0.2%		0.2%			0.3%	0.5%	1.5%	**97.0%**	0.3%						
1st Nail				0.2%							**98.2%**	1.7%					
2nd Nail			02%						0.2%		0.5%	**99.0%**		0.2%			
1st Knuckle											0.2%		**97.3%**	1.2%	0.2%	0.2%	1.0%
2nd Knuckle												0.2%	0.8%	**98.8%**		0.2%	
Back of Hand (BH)				0.2%									0.5%	0.2%	**92.2%**	4.7%	2.3%
Outer Wrist (OW)	0.2%												0.2%		6.0%	**90.2%**	3.5%
Inner Wrist (IW)	0.7%	0.7%	0.2%	0.2%	0.3%						0.2%		0.7%	0.2%	0.8%	0.5%	**96.2%**

Table 3: Classification percentages for classes at the fine-grained level (Stage 4 output), averaged across 20 trials and 30 participants. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row).

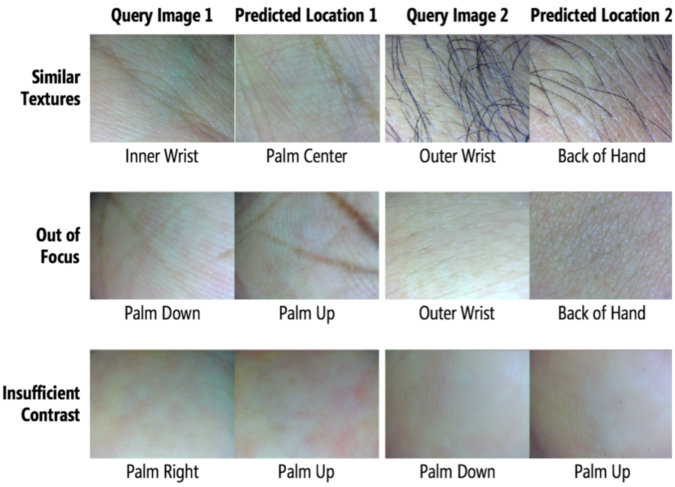

Figure 6: Classification errors were caused primarily by similarities between the locations' visual textures and poor image quality. Each set of images shows, in order, two examples (from different participants) of an incorrectly classified test image along with a training image from the predicted location.

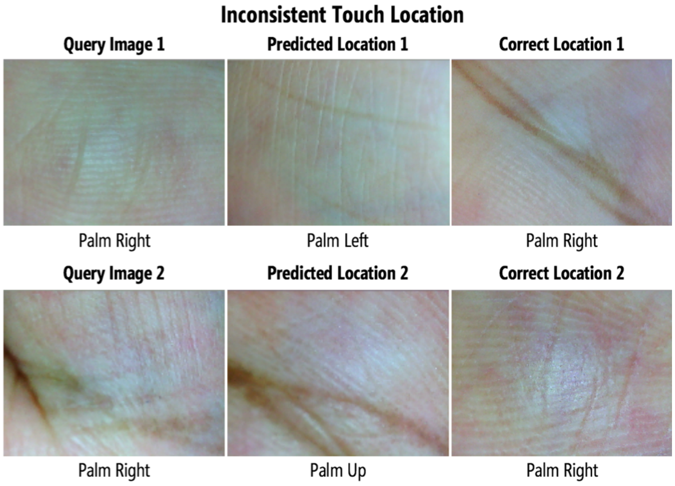

Figure 7: Classification errors for several participants were also caused by inconsistent touch locations. Shown are two examples (from two different participants) where the locations were far enough apart to appear as entirely unrelated images.

geometric keypoint matching (Figure 7). In these cases the algorithm fell back to the best Stage 3 texture match.

To examine how performance varies across individuals, Figure 8a shows F_1 scores broken down by participant. F_1 scores ranged from 95.9% to 100.0% at the coarse-grained level (Stage 2) and 86.5% to 99.7% at the fine-grained level (Stage 4). Participant 29 performed the worst, with a Stage 4 F_1 score of 86.5%—4.4 standard deviations below the mean. Based on a qualitative examination, we found decreased skin contrast with fewer distinctive finger and palm features, as well as significant variations in translation, rotation, and image focus for each location. In comparison, the top performing participants had high contrast skin textures, more consistent pressure (resulting in fewer variations in lighting and focus), and greater consistency in returning to the same touch location each trial.

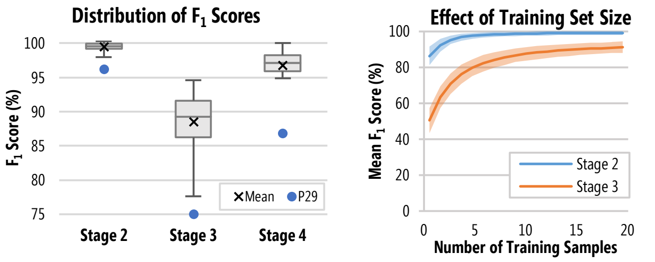

Figure 8: (a) Distribution of F_1 scores by participant, with outlier P29 marked by the blue dot; (b) Effect of the number of training examples on mean texture classification F1 score at coarse-grained (Stage 2, blue) and fine-grained (Stage 3, orange) levels.

Our dataset captured some variations in age, skin tone, and hand size, although not in large enough numbers to confidently determine the effect those factors may have on classification performance (see Table 1). We found no significant correlations across our 30 participants with any of these variables; however further work with a larger, more diverse participant pool is needed.

B. *Effect of Training Set Size on Performance*

To explore performance as a function of training set size, we tested our algorithms again using *n*-fold cross-validation but this time varying the number of training samples from $m = 1$ to 19. Specifically, we randomly selected from the 20 images per class available for each participant, with one image set aside for testing. Figure 8b shows the average texture classification accuracy at the coarse-grained (Stage 2) and fine-grained (Stage 3) levels when increasing the number of training examples. To reduce the effect of selecting the images randomly and obtain a more representative estimate, we averaged the results of 10 randomized trials. Each point represents the average F_1 score across all participants, locations, and trials when trained using *m* examples. Accuracy begins to level off above five training images per location, especially at the coarse-grained level (which approaches 100% accuracy). However, performance at both levels steadily improves as the number of training images is increased. We did not evaluate Stage 4 for this experiment as its performance depends largely upon the amount of spatial overlap between the training and testing images rather than the number of training samples.

C. *Between-person Classification*

To potentially bootstrap the training set and to identify similarities across individuals, we conducted a secondary classification experiment in which the training set and testing set consisted of images from different participants (*i.e.,* between-person experiments). More specifically, we employed *n*-fold cross-validation, where each fold trained on data from 29 participants and tested on the remaining participant. We did not expect this approach to yield a high accuracy, especially at the fine-grained level since finger and palm prints can vary significantly person to person (which is the basis of biometric identification). However, we hoped to discover textural similarities across participants that could be used to boost future classifiers to either improve accuracy or reduce the amount of per user training.

Between-person Coarse-grained Classification Confusion Matrix

	Palm	*Finger*	*Nail*	*Knuckle*	*Other*
Palm	**55.2%**	16.8%	7.8%	4.0%	38.8%
Finger	8.1%	**85.5%**	10.4%	2.1%	2.3%
Nail	0.2%	3.4%	**85.3%**	4.4%	0.9%
Knuckle	1.2%	0.2%	1.3%	**67.8%**	18.2%
Other	12.4%	4.1%	0.1%	18.2%	**60.3%**

Table 4: Between-person classification percentages for classes at the coarse-grained level. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row).

As expected, the between-person classification results are lower than the within-person results. At the coarse-grained level, our classification algorithms achieve an average precision of 72.6% (*SD*=12.9%) and recall of 70.8% (*SD*=12.3%). Still, these results are considerably higher than chance for five classes (20%) or majority-vote for the palm class (29.4%). See Table 4 for a confusion matrix. Average precision at the fine-grained level is 27.1% (*SD*=7.5%) and recall of 26.1% (*SD*=5.8%), which are also well above chance for 17 classes (5.9%). Although these accuracies are clearly too low to support a reliable user interface without an individual training procedure, they may provide enough classification information to allow for bootstrapping.

VI. Discussion

Our controlled experiments explored the distinguishability of small image patches on the surface of the hand and wrist for localization purposes. In our within-person experiments we were able to achieve an average F_1 score above 99% at the coarse-grained level (Stage 2) and above 96% at the fine-grained level (Stage 4), which suggests that skin-surface image patches can be classified and localized on the body with high levels of accuracy. While an end-to-end deep learning approach may be more elegant, our more heuristic approach requires substantially less training data, and our performance results suggest that an on-body input system applying our algorithms is feasible. Here, we reflect on the implications of our findings as well as challenges for implementing a real-time system.

A. Expanding On-Body Input

While we only evaluated locations on the hand and wrist, our finger-mounted approach should support a range of input locations within the user's reach, including on-body and off-body surfaces (*e.g.*, tabletops). This is in contrast to most previous on-body input approaches that are more limited by their fixed sensor placements and range. Although recognition accuracy may drop as the number of locations increases (*e.g.*, thigh, forearm), we expect to boost performance through improvements to our hierarchical approach. Performance was particularly high at the first level of the hierarchy, with an F_1 score above 99%. Thus, for each region we could apply different preprocessing and matching approaches at the second level that are tuned specifically to distinguish the fine-grained locations within that region. For example, we could extract knuckle-specific features (*e.g.*, [2]) to distinguish knuckle locations, which may require completely different algorithms than the palm locations. Similarly, it will be important to explore the feasibility of extending the localization hierarchy further, for regions that can support an even finer level of granularity beyond the locations studied (*e.g.*, palm, fingers); such granularity could enable highly precise on-body interactions (*e.g.*, sliding your finger along your palm to trace a map route).

B. Training a Camera-Based On-Body Localization System

The procedure for training a new user may impact both algorithmic performance and user perceptions toward the system. As shown in Figure 7b, classification performance improves with the number of training examples, but begins to level off after five examples per class. However, it may be possible to boost accuracy while simultaneously reducing the number of training examples that are required of a new user. The images in our dataset relied on natural variations that were introduced through randomization during data collection. To potentially improve performance, the training interface could prompt the user to vary rotations, poses, and perspectives– similar to Apple's iPhone training procedure for their fingerprint sensor. In addition, as our preliminary experiments indicate, it may be possible to bootstrap the system using between-person data and reduce the amount of training required for a new user. This approach would work especially well in our first stage of classification to recognize surface classes that appear similar across many users (*e.g.*, skin, knuckles, clothing).

C. Limitations and Future Work

Our experiments were conducted under controlled conditions, but a real-time system would likely need to deal with greater variations in image quality. Although we randomized trial order to introduce natural variation in translation, rotation, and pressure, we carefully controlled for other variables such as distance, lighting, and perspective. A finger-worn camera will likely constrain this complexity, potentially mitigating these concerns. For example, distance will remain relatively constant during touch-based interactions since the camera can be positioned at a fixed location on the finger and lighting can be controlled via a self-illuminated camera. While perspective may vary considerably, our results show that our algorithm functions well for both 90-degree and 45-degree perspectives. Further work is necessary to explore variations under less controlled conditions, including potential changes over time (*e.g.*, due to differences in humidity/dryness), as well as other variations in skin surface textures and features due to age, skin tone, and hand size. The above mitigating factors suggest that our approach should still be applicable.

Our work focused solely on RGB camera-based sensing using static images. Future research should explore other imaging and non-imaging sensors as well as combining video and multiple sensor streams (sensor fusion). For example, hyperspectral imaging would expose veins and other sub-dermal features that could be used for localization as well as improve the contrast of surface features across a wider range of skin tones (*e.g.*, [38]). Depth sensors could provide 3D geometry of the hand and ridges, potentially improving robustness to variations in perspective and allowing us to more reliably extract finger and palm print features to use for localization (*e.g.*, [39]). Finally, non-imaging sensors (*e.g.,* infrared reflectance [14] or inertial motion [40]) could provide complementary information to help resolve visual ambiguities and better integrate localization with gesture recognition.

VII. Conclusion

This paper introduces an algorithmic pipeline for recognizing low-resolution, close-up images of several different locations on the hand/wrist with an average F_1 score of 96.5% for within-person skin patch classification. While future work will need to address potential implementation challenges with a real-time system, our results suggest that a finger-mounted computer vision approach to support location-based on-body interaction should be feasible and that the system training process may be able to be bootstrapped using a dataset of hand images collected from multiple individuals.

References

[1] A. Meraoumia, S. Chitroub, and A. Bouridane, "Fusion of Finger-Knuckle-Print and Palmprint for an Efficient Multi-Biometric System of Person Recognition," in *Proc. of ICC '11*, 2011, pp. 1–5.

[2] M. Choraś and R. Kozik, "Contactless palmprint and knuckle biometrics for mobile devices," *Pattern Anal. Appl.*, vol. 15, no. 1, pp. 73–85, Feb. 2012.

[3] A. K. Jain and J. Feng, "Latent Palmprint Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1032–1047, Jun. 2009.

[4] Eryun Liu, A. K. Jain, and Jie Tian, "A Coarse to Fine Minutiae-Based Latent Palmprint Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2307–2322, Oct. 2013.

[5] M. O. Derawi, B. Yang, and C. Busch, "Fingerprint Recognition with Embedded Cameras on Mobile Phones," in *Security and Privacy in Mobile Info. and Com. Sys.*, no. Jan 2012, Springer, 2012, pp. 136–147.

[6] A. Morales, M. A. Ferrer, and A. Kumar, "Improved palmprint authentication using contactless imaging," in *IEEE Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2010, pp. 1–6.

[7] X. Wu, Q. Zhao, and W. Bu, "A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors," *Pattern Rec.*, vol. 47, no. 10, pp. 3314–3326, Oct. 2014.

[8] C. Harrison, D. Tan, and D. Morris, "Skinput: Appropriating the Body As an Input Surface," in *Proceedings of CHI '10*, 2010, pp. 453–462.

[9] R.-H. Liang, S.-Y. Lin, C.-H. Su, K.-Y. Cheng, B.-Y. Chen, and D.-N. Yang, "SonarWatch: Appropriating the Forearm as a Slider Bar," in *SIGGRAPH Asia 2011 Emerging Technologies*, 2011, p. 5.

[10] N. Dezfuli, M. Khalilbeigi, J. Huber, F. Müller, and M. Mühlhäuser, "PalmRC: Imaginary Palm-Based Remote Control for Eyes-free Television Interaction," in *Proc. of EuroITV '12*, 2012, p. 27.

[11] C. Harrison and A. D. Wilson, "OmniTouch: wearable multitouch interaction everywhere," in *Proc. of UIST '11*, 2011, pp. 441–450.

[12] E. Tamaki, T. Miyaki, and J. Rekimoto, "Brainy Hand: an earworn hand gesture interaction device," in *Extended Abstracts of ACM CHI 2009*, 2009, pp. 4255–4260.

[13] P. Mistry and P. Maes, "SixthSense: A wearable gestural interface," in *Proc. of ACM SIGGRAPH Asia*, 2009, p. Article No. 11.

[14] M. Ogata, Y. Sugiura, Y. Makino, M. Inami, and M. Imai, "SenSkin: Adapting Skin as a Soft Interface," in *Proc. UIST '13*, 2013, pp. 539–544.

[15] M. Weigel, T. Lu, G. Bailly, A. Oulasvirta, C. Majidi, and J. Steimle, "iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing," in *Proc. CHI'15*, 2015, pp. 2991–3000.

[16] U. Oh and L. Findlater, "Design of and subjective response to on-body input for people with visual impairments," in *Proc. of ASSETS '14*, 2014, pp. 115–122.

[17] S. G. Gustafson, B. Rabe, and P. M. Baudisch, "Understanding Palm-based Imaginary Interfaces: The Role of Visual and Tactile Cues when Browsing," in *Proc. of CHI '13*, 2013, pp. 889–898.

[18] U. Oh and L. Findlater, "A Performance Comparison of On-Hand versus On-Phone Non-Visual Input by Blind and Sighted Users," *ACM Trans. Access. Comput.*, vol. 7, no. 4, p. 14, 2015.

[19] S. Gustafson, C. Holz, and P. Baudisch, "Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device," in *Proc. of UIST '11*, 2011, pp. 283–292.

[20] X.-D. Yang, T. Grossman, D. Wigdor, and G. Fitzmaurice, "Magic finger: always-available input through finger instrumentation," in *Proc. of UIST '12*, 2012, pp. 147–156.

[21] L. Stearns, R. Du, U. Oh, Y. Wang, R. Chellappa, L. Findlater, and J. E. Froehlich, "The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind," *Proc. ECCV '14, Workshop on Assistive Computer Vision and Robotics*, 2014.

[22] J. Doublet, M. Revenu, and O. Lepetit, "Robust GrayScale Distribution Estimation for Contactless Palmprint Recognition," in *IEEE Conference on Biometrics: Theory, Applications, and Systems 2007*, 2007, pp. 1–6.

[23] V. Kanhangad, A. Kumar, and D. Zhang, "A Unified Framework for Contactless Hand Verification," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3, pp. 1014–1027, Sep. 2011.

[24] G. K. Ong Michael, T. Connie, and A. B. Jin Teoh, "Touch-less palm print biometrics: Novel design and implementation," *Image Vis. Comput.*, vol. 26, no. 12, pp. 1551–1560, 2008.

[25] Wai Kin Kong and D. Zhang, "Palmprint texture analysis based on low-resolution images for personal authentication," in *Proc. Pattern Recognition '02*, 2002, vol. 3, pp. 807–810.

[26] M. Ekinci and M. Aykut, "Palmprint Recognition by Applying Wavelet-Based Kernel PCA," *Comput. Sci. Technol.*, vol. 23, no. 107, pp. 851–861, 2008.

[27] A. S. Parihar, A. Kumar, O. P. Verma, A. Gupta, P. Mukherjee, and D. Vatsa, "Point based features for contact-less palmprint images," in *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, 2013, pp. 165–170.

[28] M. Laadjel, A. Bouridane, F. Kurugollu, O. Nibouche, and W. Yan, "Partial Palmprint Matching Using Invariant Local Minutiae Descriptors," in *Transactions on Data Hiding and Multimedia Security*, 2010, pp. 1–17.

[29] D.-S. Huang, W. Jia, and D. Zhang, "Palmprint verification based on principal lines," *Patt. Recog.*, vol. 41, no. 4, pp. 1316–1328, Apr. 2008.

[30] L. Nanni and A. Lumini, "Local binary patterns for a hybrid fingerprint matcher," *Patt. Recog.*, vol. 41, no. 11, pp. 3461–3466, Nov. 2008.

[31] Harrison, S. Ramamurthy, and Hudson, "On-body interaction: armed and dangerous," *Proc. TEI '12*, pp. 69–76, 2012.

[32] R. Shilkrot, J. Huber, M. E. Wong, P. Maes, and S. Nanayakkara, "FingerReader: a wearable device to explore printed text on the go," in *Proc. of CHI '15*, 2015, pp. 2363–2372.

[33] L. Findlater, L. Stearns, R. Du, U. Oh, D. Ross, R. Chellappa, and J. E. Froehlich, "Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch," in *Proc. of ASSETS '15*, 2015, pp. 383–384.

[34] K. Kanagalakshmi and E. Chandra, "Performance evaluation of filters in noise removal of fingerprint image," in *2011 3rd International Conf. on Electronics Computer Technology*, 2011, vol. 1, pp. 117–121.

[35] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[36] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *Pattern Recognition*, vol. 42, no. 7, pp. 1408–1418, 2009.

[37] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," in *Proc. of ECCV '04*, 2004, pp. 469–481.

[38] M. Goel, S. N. Patel, E. Whitmire, A. Mariakakis, T. S. Saponas, N. Joshi, D. Morris, B. Guenter, M. Gavriliu, and G. Borriello, "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications," in *Proc. of UbiComp '15*, 2015, pp. 145–156.

[39] W. Li, D. Zhang, L. Zhang, G. Lu, and J. Yan, "3-D Palmprint Recognition With Joint Line and Orientation Features," *IEEE Trans. Syst. Man, Cybern. Part C*, vol. 41, no. 2, pp. 274–279, Mar. 2011.

[40] L. Jing, Y. Zhou, Z. Cheng, and T. Huang, "Magic ring: A finger-worn device for multiple appliances control using static finger gestures," *Sensors*, vol. 12, no. 5, pp. 5775–5790, 2012.