Robust Hand Geometry Measurements for Person Identification using Active Appearance Models

Ralph Gross, Yiheng Li, Latanya Sweeney, Xiaoqian Jiang, Wanhong Xu, and Daniel Yurovsky

Abstract— The increased demand for tighter border and building security has renewed public interest in biometric identification and verification systems. With fingerprint recognition being socially stigmatized, hand geometry-based recognizers have emerged as niche solutions. However, systems currently available in the marketplace require direct contact with the device, raising, among others, significant hygiene concerns. In this paper we introduce a novel approach to hand geometry-based identification. The proposed method employs Active Appearance Models to track the hand inside the capture device and to extract geometry features for identification. The AAM fitting algorithm runs faster than real-time, enabling robust system performance. In experiments on a small-scale database of hand images, the accuracy of our system exceeds 90% using as little as five features.

I. Introduction

The increased demand for tighter border and building security has renewed public interest in biometric systems for identification (one-to-many matching) and verification (one-to-one matching). With fingerprint recognition being socially stigmatized due to its connection to crime scene analysis, hand geometry-based recognizers have emerged as niche solutions [7]. However, systems currently available in the marketplace require usage of physical restraints such as pegs to guarantee consistent hand positioning. In order to interact with the system, users therefore have to place their hand directly onto the device, raising significant hygiene concerns.

In this paper we introduce a novel approach to hand geometry-based person identification. Our system employs Active Appearance Models (AAMs [3], [12]) to track the hands of subjects inside the capturing device. Using the location of landmark points on the hand determined from the fitted AAM, we extract distance features for identification. Unlike the devices currently available on the market, our system allows subjects to put their hand anywhere in the field of view of the camera while minimizing physical contact to the device and thereby eliminating hygiene concerns. In experiments on a database of 18 subjects the accuracy of our system exceeds 90% using as little as five features.

This paper is organized as follows. Section II provides an overview over prior work in the field. Section III describes our system and all of its components. We introduce the AAM framework in Section IV and discuss the algorithm used to fit a model to a new image. In Section V we describe the hand database that we collected to facilitate development and testing of our system. Section VI discusses a set of candidate distance features and a feature consistency measure used to

determine the most reliable features for inclusion into the classifier. In Section VII we report the results of model fitting and classification experiments before we conclude with discussions and future work in Section VIII.

II. Background: Hand Geometry-Based Biometric Systems

While the biometrics market is dominated by devices based on fingerprint recognition, hand geometry-based systems for person verification have been used for many years. Despite its comparatively small share of the overall biometrics market (10% in 2003 in comparison to 52% for fingerprint based systems [7]) hand geometry-based systems are popular for access-control and time-and-attendance applications [6]. Hand geometry systems measure various distances on the hand, including the overall length, width and thickness of the hand as well as the dimensions of the fingers. Since distances computed on the hand vary substantially with pose and finger configuration, sensing devices often use physical restraints to ensure consistent hand placement. In [8] 5 pegs are used to guide positioning of the hand. The system then extracts 16 distance features from the fingers and the hand. Using a database of 50 people, the authors report a verification rate of over 90% at a false acceptance rate of 0.01. Using a comparable hardware setup Sanchez-Reillo et al. employ six pegs to position the hand in the capture device [13]. The system however extracts 31 features for comparisons. On a small dataset of 200 images recorded from 20 subjects recognition accuracies between 88% and 97% are reported. While pegs help overall in hand positioning they also alter the hand shape and fail to prevent small variations in finger placement (shown in [16]). Systems proposed more recently therefore employ alternative means for consistent feature acquisition. [16] uses a flatbed scanner, requiring subjects to separate the fingers. With the help of heuristics, landmark points are located on the hand which serve as reference points to extract five length and eight width measurements together with the contour of three fingertips. On a dataset of 288 images recorded from 22 subjects the system achieves a 89% verification rate at a 2.2% false acceptance rate. In [11] images are recorded with a digital camera. The users are instructed to ensure that the back of their hands touches the device table and that their fingers are well separated. After extraction of the hand region through thresholding, the hand position is normalized using the axes of a fitted ellipse. The system then extracts 16 features. On a database of 1000 images recorded from 100 users the best system operating point was achieved at a 5.29% false
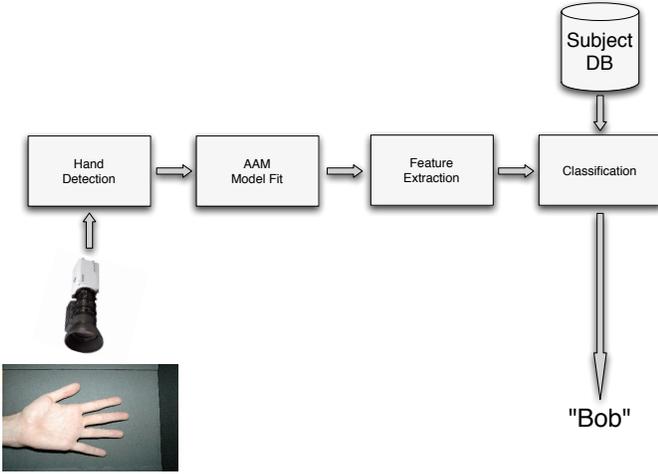
Fig. 1. Overview of the proposed hand-geometry based identification system. A video camera records the hand movements in front of a static background. After hand localization, the AAM hand model is fitted to the image. The model parameters resulting from the fit are used to compute geometry measurements of the hand which are matched against a database of previously recorded measurements.

acceptance rate and 8.34% false rejection rate. In [15] Active Appearance Models were used for hand appearance based matching. Good results are reported for both identification as well as verification settings, however, no details about the underlying hand database are given.

III. HAND-GEOMETRY BASED IDENTIFICATION SYSTEM

Our system employs a static video camera within an enclosure to acquire image sequences of the hand to be identified. See Figure 1 for an illustration. Since the background is static and controlled, we use background subtraction (see e.g. [2], [17]) to detect the presence of a hand in the field of view. The position of the centroid of the hand region along with the approximate scale then serves as initialization point for the Active Appearance Model fitting algorithm. After a sufficient fit has been established (as measured by the model reconstruction error) the AAM fitting algorithm is used to update the model position in each image. Once the hand position stabilizes (as determined from the frame-to-frame model position distance) hand geometry measurements are extracted from the fitted AAM model and compared against a database. The current version of the system assumes hands to be at a fixed distance from the camera with the fingers fully extended. If a given application requires more flexiblity, alternative solutions can be employed, e.g. estimating the camera-hand distance using stereo vision.

In an identification application the system outputs the name of the closest match from the subject database if the distance to the model satisfies a pre-set matching criterium. During operation the quality of the AAM fit is monitored constantly. If the reconstruction error exceeds a predetermined threshold the model is re-initialized or the user is alerted in case of deviations from the target hand configuration. Since the AAM fitting algorithm runs comfortably at

more than 230 frames per second [12], real-time processing is still achieved. Due to the robustness of the AAM fitting algorithm the system can handle arbitrary hand rotations in the image as well as non-separated fingers. Usage of a video camera as input device trades off image resolution with overall system speed. In order to achieve both, a hybrid still camera/video camera setup can be used. The hand is tracked in the image sequence acquired from the video camera using the combination of background subtraction and AAM model fit. Once a stable hand position is established, the high resolution still camera is triggered. Both cameras are calibrated so that the model fit achieved in the images from the video camera can be used to more accurately fit the model in the high resolution image [10]. The system can be augmented to include one or multiple controlled flashes.

IV. ACTIVE APPEARANCE MODELS

We begin with a review of Active Appearance Models (AAMs) [3]. We define them, explain model construction from training data, and describe the efficient "Project Out" fitting algorithm introduced by Matthews and Baker [12].

A. Definition and Model Construction

The *2D shape* of an AAM is defined by a 2D triangulated mesh and in particular the vertex locations of the mesh. Mathematically, we define the shape $\mathbf{s}$ of an AAM as the 2D coordinates of the n vertices that make up the mesh: $\mathbf{s} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^{\mathrm{T}}$. AAMs allow linear shape variation. This means that the shape matrix $\mathbf{s}$ can be expressed as a base shape $\mathbf{s}_0$ plus a linear combination of m shape matrices $\mathbf{s}_i$:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} p_i \, \mathbf{s}_i \qquad (1)$$

where the coefficients p_i are the shape parameters. AAMs are computed from training data consisting of a set of images with the shape mesh usually hand marked on them [3]. The training shapes are then geometrically aligned using the *Procrustes* algorithm [3], [12]. Principal Component Analysis (PCA) [9] is applied to the aligned training meshes. The base shape $\mathbf{s}_0$ is the mean shape and the matrices $\mathbf{s}_i$ are the (reshaped) eigenvectors corresponding to the m largest eigenvalues.

The *appearance* of the AAM is defined within the base mesh $\mathbf{s}_0$. Let $\mathbf{s}_0$ also denote the set of pixels $\mathbf{u} = (u, v)^{\mathrm{T}}$ that lie inside the base mesh $\mathbf{s}_0$. The appearance of the AAM is then an image $A(\mathbf{u})$ defined over the pixels $\mathbf{u} \in \mathbf{s}_0$. AAMs allow linear appearance variation. This means that the appearance $A(\mathbf{u})$ can be expressed as a base appearance $A_0(\mathbf{u})$ plus a linear combination of l appearance images $A_i(\mathbf{u})$:

$$A(\mathbf{u}) = A_0(\mathbf{u}) + \sum_{i=1}^{l} \lambda_i \, A_i(\mathbf{u}) \qquad (2)$$

where the coefficients λ_i are the appearance parameters. The appearance images A_i are usually computed by applying PCA to the shape normalized training images [3], [12].

B. Model Fitting

Fitting an AAM may be formulated as minimizing the sum of squares difference between the appearance $A(\mathbf{u}) = A_0(\mathbf{u}) + \sum_{i=1}^{l} \lambda_i A_i(\mathbf{u})$ and the input image warped back onto the base mesh $I(\mathbf{N}(\mathbf{W}(\mathbf{u}; \mathbf{p}); \mathbf{q}))$ [12]:

$$\sum_{\mathbf{u} \in \mathbf{s}_0} \left[A_0(\mathbf{u}) + \sum_{i=1}^{l} \lambda_i A_i(\mathbf{u}) - I(\mathbf{N}(\mathbf{W}(\mathbf{u}; \mathbf{p}); \mathbf{q})) \right]^2 \quad (3)$$

In this equation, the warp $\mathbf{W}$ is the piecewise affine warp defined by the mesh triangulation from the base mesh $\mathbf{s}_0$ to the current AAM shape $\mathbf{s}$ and $\mathbf{N}$ is a 2D similarity transformation used to normalize the shape of the AAM. The goal of AAM fitting is to minimize the expression in Equation (3) simultaneously with respect to the appearance parameters $\boldsymbol{\lambda}$, the linear shape parameters $\mathbf{p}$, and the similarity transform parameters $\mathbf{q}$.

One algorithm for fitting an AAM to an image is the "project-out" inverse compositional algorithm proposed in [12]. This algorithm performs the non-linear optimization of Equation (3) in two steps (similar to Hager and Belhumeur [5]). First the shape and linear transformation parameters $\mathbf{p}$ and $\mathbf{q}$ are found through a non-linear optimization in a subspace in which the appearance variation can be ignored. The second step is then a closed form linear optimization with respect to the appearance parameters λ. The algorithm is very fast, running at over 230 frames per second on standard hardware [12]. An alternative algorithm for fitting an AAM to an image is the Simultaneous Inverse Compositional Algorithm [1] which minimizes Equation (3) by performing Gauss-Newton gradient descent simultaneously on the warp and appearance parameters. The resulting fitting algorithm performs better then the project out algorithm described above, albeit at a much slower speed. Usage of the simultaneous algorithm in a real-time environment therefore has to be limited.

V. HAND DATABASE

To facilitate development and testing of the hand-geometry based identification system we recorded a database of hand images. We collected 54 images from 18 subjects. Each person was imaged three consecutive times within the span of approximately 60 seconds [14]. The majority of subjects changed the position of the hand in the image between recordings with some also changing the hand configuration. Figure 2 shows example images from the database.

VI. HAND GEOMETRY FEATURES FOR PERSON IDENTIFICATION

In this section we describe a set of candidate features for our system (Section VI-A) as well as a selection criterion to identify promising features (Section VI-B).

A. Candidate Features

For any classifier an ideal set of features shows little *intra* subject variation and large *inter* subject variation, leading to well separated class representations. Following this goal and

established practices in the field (see Section II) we define 14 distances on the hand as candidate features for our system. See Figure 3 for an illustration. The features include the lengths of the fingers as well as the length and width of the palm. In our framework, each feature is defined with respect to AAM model points. Therefore, after the AAM fit has been established, the feature representation for the hand can be computed very accurately, assuming a good model fit.

B. Feature Selection

Out of the 14 distance-based features, we are only interested in those that are consistent, i.e. features that are insensitive to hand pose variations. The standard deviation of a given feature across different images of the same subject provides useful information, but it is not normalized: a medium large standard deviation might be good for large-distanced features, but certainly would disqualify a small-distanced one. Thus, we introduced the following feature consistency metric, rms_i, for feature i:

$$rms_i = \frac{\frac{1}{n} \sum_{j=1}^{n} \mu^{n_j}(f_i(sub_j))}{\frac{1}{n} \sum_{j=1}^{n} \sigma^{n_j}(f_i(sub_j))} \quad (4)$$

where $f_i(sub_j)$ refers to the i-th feature of subject j (with n_j examples for subject j in the dataset). We furthermore define

$$\mu^{n_j}(f_i(sub_j)) = \frac{1}{n_j} \sum_{s=1}^{n_j}(f_i(sub_{j,s})) \quad (5)$$

as the average of the values of feature i for subject j. Here $sub_{j,s}$ refers to the example s for subject j. Finally we define

$$\sigma^{n_j}(f_i(sub_j)) = \frac{1}{n_j} \sum_{s=1}^{n_j}(f_i(sub_{j,s}) - \mu^{n_j}(f_i(sub_{j,s})))^2 \quad (6)$$

as the standard deviation of feature i for subject j. For each feature i the rms_i metric therefore computes an averaged version of the ratio of the magnitude of the feature and the corresponding deviation. The higher the rms_i value, the more consistent we consider the feature i to be.

VII. EXPERIMENTS

In this section we evaluate the performance of the two main system components, AAM fitting and classification using the hand database introduced in Section V. We first discuss the results of synthetic fitting experiments using the AAM model constructed from the images of our hand database (Section VII-A). In Section VII-B we then evaluate the performance of a classifier using the features derived from the fitted AAM model.

A. Model Fitting

We use all images and labels in the hand database described in Section V to build a first AAM model[1]. In order

---

[1]The fitting experiments are designed to evaluate only the accuracy of the fitting algorithm, not the capability of the model to handle unseen data. For this reason, all labels are included in the dataset.
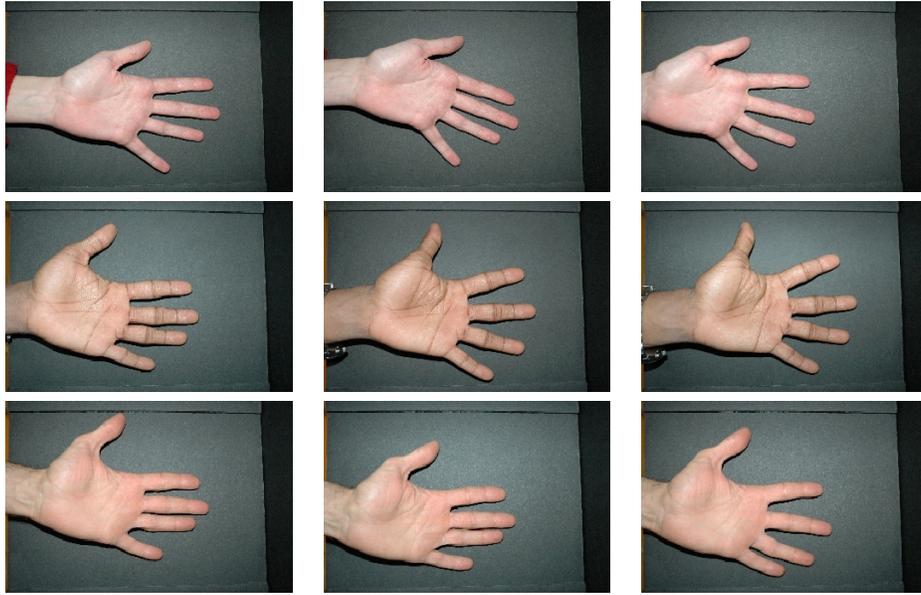
Fig. 2. Example images from the hand database. Hand positions and poses vary significantly between users. Subjects were encouraged to change the hand pose between recordings.
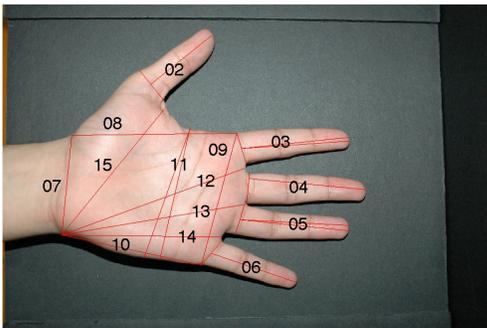

Fig. 3. Location of the 14 features used as candidates in our feature selection process. The distances are defined with respect to the AAM model points. In cases where two lines are shown for a particular feature (features "03", "05", and "11"), the longer of the two lines is chosen.

Fig. 4. Results from fitting experiments comparing the original and refitted AAM models for different ground-truth perturbations (referred to as "10/1" and "10/3").

to reduce noise in the manually established labels we apply refitting [4]. In this process the AAM fitting algorithm is initialized with the ground-truth labels and run until convergence. The vertex locations of the fitted mesh are used as ground-truth labels to build a new model. It was shown in [4] that using refitting improves the fitting accuracy on face data, especially for unseen subjects. In order to evaluate the accuracy of the fitting algorithm we randomly perturb the ground-truth shape and similarity transform parameters and initialize the fitting algorithm with it. We then run the fitting algorithm until convergence and measure the Euclidean distance between the ground-truth vertex locations and the fitted vertex locations.
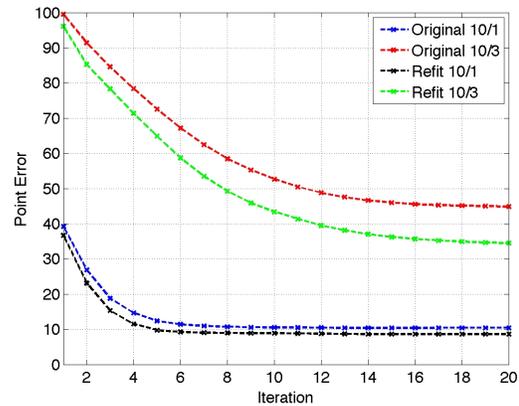
Figure 4 shows the Euclidean distance between the fitted model and the ground truth across iterations of the fitting algorithm for both the original and refitted model, averaged across all 54 images. Refitting reduces the error for both small and large initial perturbations by 17% and 23%, respectively. Figure 5 shows examples of the refitted model before and after fitting. The fitting algorithm is able to accurately retrieve the ground-truth labels for small perturbations and for most of the large perturbations.

B. Classification Accuracy

In order to evaluate the system performance in a realistic setting, we determine identification accuracies for two sets of hand-geometry measures: (a) distances computed from the manually established ground-truth labels (referred to as
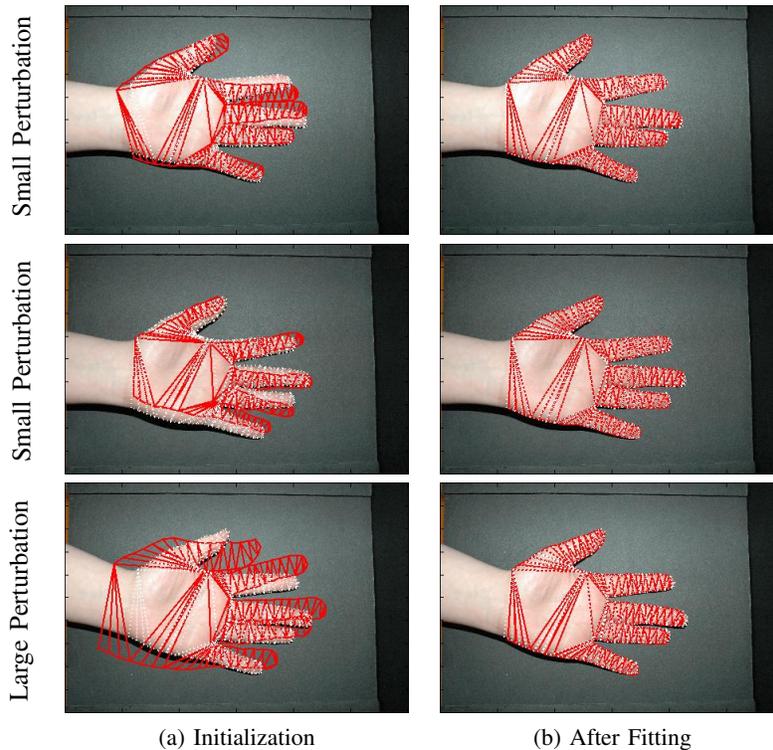
(a) Initialization   (b) After Fitting

Fig. 5. Results of applying the fitting algorithm using the refitted model. The images in the (a) column show the ground-truth labels (white) and the model after initial perturbation (red). The images in the (b) column show the model after convergence of the fitting algorithm. The algorithm is able to accurately fit the model for all of the small perturbations and for most of the large perturbations.

		M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
GTD	AvgMean	434.1935	645.9435	728.7413	643.9215	536.8944	620.0989	933.5206	817.7293	758.3496	850.1235	1102.598	1033.476	908.2513	957.7213
	AvgStdev	14.1436	10.13595	8.843408	7.926835	8.058131	9.953717	14.65154	8.797033	16.53992	6.910976	11.7292	12.89964	12.96122	9.296884
	rms	30.69893	63.72795	82.40503	81.23311	66.62767	62.29822	63.71483	92.95512	45.84966	123.0106	94.00454	80.11671	70.07451	103.0153

		M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
FTD	AvgMean	430.7407	643.8694	725.4004	639.788	532.6261	617.4728	933.72	816.0743	757.3106	848.5833	1101.475	1031.718	908.9726	954.4163
	AvgStdev	10.78056	6.804586	6.784188	9.13805	8.076536	8.229159	14.01989	9.075007	25.39715	5.953137	21.29947	22.46883	19.3817	8.510045
	rms	39.95532	94.62287	106.9252	70.01362	65.94734	75.03474	66.59968	89.92547	29.81873	142.5439	51.71372	45.91774	46.8985	112.1517

Fig. 6. Feature consistency measure rms for the 14 candidate features shown in Figure 3. In the first row we show results for features derived from ground-truth labels (GTD). The features extracted after perturbation from ground-truth and fitting of the model are shown in the second row (FTD). The 7 most consistent features are highlighted in color.

GTD) and (b) distances extracted from the fitted model after initial perturbation from ground-truth (referred to as FTD). Using the rms metric introduced in Section VI-B we evaluate the feature consistency for the 14 measures described in Section VI-A. Figure 6 lists the results for both the GTD and FTD measurements with the most consistent features highlighted. The results show large differences in rms values between features. Feature $M2$ (length of the thumb) is far less consistent than feature $M11$ (hand width), a plausible result since $M2$ is much more likely to change across hand poses than $M11$. $M11$ is therefore a better candidate to be used for classification. The average rms value for GTD (75.7) and FTD (74.1) is similar, with FTD values being slightly higher when averaged over the 7 most consistent features (98.7) in comparison to the GTD values (93.8).

To determine identification performance we perform 3-fold cross-validation experiments with a 1-Nearest-Neighbor classifier using the 3 images available for each subject. Using the 7 most consistent features we exhaustively search for the best performing 1-, 2-, 3-, etc. feature combination, treating each feature as a dimension in the corresponding Euclidean space. For reference purposes we also report the corresponding Top2 match rate (correct match within 2-Nearest-Neighbors). The results are shown graphically in Figure 7 and numerically in Figure 8. The identification performance is high for both distance measures, 94.4% for GTD and 90.7% for FTD, in each case using only 5 features. We therefore conclude that on our database differences in performance between using ground-truth labels and fitted labels are small.

VIII. SUMMARY AND FUTURE WORK

In this paper we introduced a novel approach to hand geometry-based person identification. We successfully used Active Appearance Models to robustly extract distance measurements from hand images. Unlike the devices currently
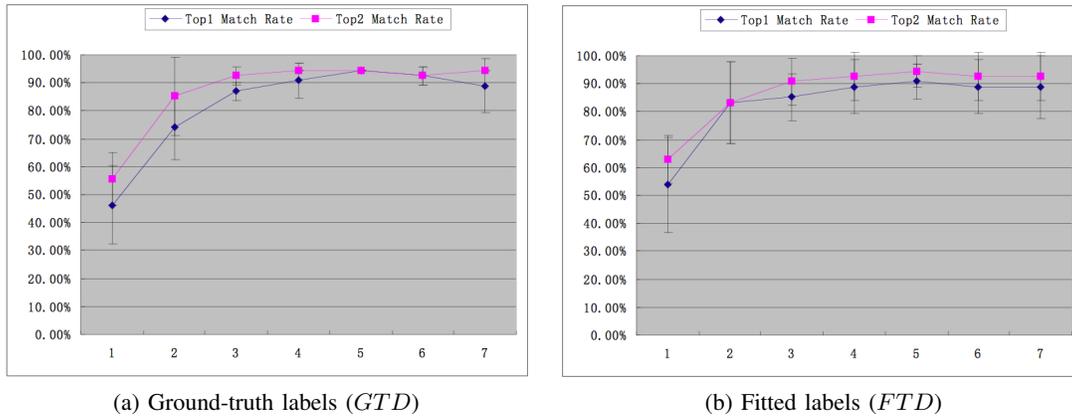
(a) Ground-truth labels (GTD)   (b) Fitted labels (FTD)

Fig. 7. Classification accuracy obtained using 3-fold cross-validation with different numbers of features for ground-truth measurements (GTD) and features extracted after fitting the AAM (FTD). The performance achieved by the two methods is very much comparable.

	Number of Measures	1	2	3	4	5	6	7
GTD	Best Measure Set	11	9,15	4,9,12	4,5,9,12	4,9,11,12,15	4,5,9,11,12,15	4,5,9,11,12,13,15
	Top1 Match Rate	46.30%	74.07%	87.04%	90.74%	94.44%	92.59%	88.89%
	Standard Deviation	13.98%	11.57%	3.21%	6.42%	0.00%	3.21%	9.62%
	Top2 Match Rate	55.56%	85.19%	92.59%	94.44%	94.44%	92.59%	94.44%
	Standard Deviation	9.62%	13.98%	3.21%	0.00%	0.00%	3.21%	0.00%

	Number of Measures	1	2	3	4	5	6	7
FTD	Best Measure Set	11	4,15	4,11,15	3,9,11,15	3,4,9,11,15	3,4,7,9,11,15	3,4,5,7,9,11,13
	Top1 Match Rate	53.70%	83.33%	85.19%	88.89%	90.74%	88.89%	88.89%
	Standard Deviation	16.97%	14.70%	8.49%	9.62%	6.42%	9.62%	11.11%
	Top2 Match Rate	62.96%	83.33%	90.74%	92.59%	94.44%	92.59%	92.59%
	Standard Deviation	8.49%	14.70%	8.49%	8.49%	5.56%	8.49%	8.49%

Fig. 8. Numerical results corresponding to the performance graphs in Figure 7. The difference in Top1 classification accuracy between ground-truth measurements (GTD) and features extracted after fitting the AAM (FTD) is small with the Top2 results for the best performing setting of five features being identical.

available on the market, our system allows subjects to put their hand anywhere in the field of view of the camera while minimizing physical contact to the device, thereby eliminating hygiene concerns. We proposed a feature consistency measure and used it to extract features suitable for recognition from a larger set of candidates. In experiments on a small-scale database, our system achieved identification accuracies in excess of 90% using as little as five features.

In order to extend the experiments reported on here we plan on collecting a significantly larger database of hand images. This dataset will be used to evaluate the ability of AAMs to generalize to unseen subjects, similiar to our previous work on faces [4].

IX. ACKNOWLEDGEMENTS

REFERENCES

[1] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Carnegie Mellon University Robotics Institute, 2003.

[2] S.-C. Cheung and C. Kamath. Robust background subtraction with foreground validation for urban traffic video. *Journal on Applied Signal Processing*, 14:2330–2340, 2005.

[3] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE PAMI*, 23(6):681–685, June 2001.

[4] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(12):1080–1093, 2005.

[5] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE PAMI*, 20(10):1025–1039, 1998.

[6] D. Hakala. Buying into biometrics. In *VARBusiness*, June, 7 2002.

[7] International Biometric Group. Biometrics market report 2003-2007, 2003.

[8] A.K. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. In *AVBPA*, March 1999.

[9] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE PAMI*, 12(1):103–108, 1990.

[10] S. Koterba, S. Baker, I. Matthews, C. Hu, J. Xiao, J. Cohn, and T. Kanade. Multi-view AAM fitting and camera calibration. In *ICCV*, volume 1, pages 511–518, 2005.

[11] A. Kumar, D.C.M. Wong, H.C. Shen, and A.K. Jain. Personal verification using palmprint and hand geometry biometric. In *AVBPA*, 2003.

[12] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[13] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos. Biometric identification through hand geometry measurements. *IEEE PAMI*, 22(10), October 2000.

[14] L. Sweeney. Hand shake: a database of naturally posed hands. Carnegie Mellon University, Data Privacy Lab, Working paper 31, 2006.

[15] X. Teng, Y. Liu, and C. Liu. AAM based matching of hand appearance for user verification. In *Advances in Biometric Person Authentication*, 2004.

[16] A.L.N. Wong and P. Shi. Peg-free hand geometry recognition using hierarchical geometry and shape matching. In *IAPR Workshop on Machine Vision Applications*, 2002.

[17] C.R. Wren, A. Azarbayejani, T.J. Darrell, and A.P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE PAMI*, 19(7):780–785, July 1997.