Personalized 3D-Aided 2D Facial Landmark Localization

Zhihong Zeng, Tianhong Fang, Shishir K. Shah, and Ioannis A. Kakadiaris

Computational Biomedicine Lab
Department of Computer Science
University of Houston
4800 Calhoun Rd, Houston, TX 77004

Abstract. Facial landmark detection in images obtained under varying acquisition conditions is a challenging problem. In this paper, we present a personalized landmark localization method that leverages information available from 2D/3D gallery data. To realize a robust correspondence between gallery and probe key points, we present several innovative solutions, including: (i) a hierarchical DAISY descriptor that encodes larger contextual information, (ii) a Data-Driven Sample Consensus (DDSAC) algorithm that leverages the image information to reduce the number of required iterations for robust transform estimation, and (iii) a 2D/3D gallery pre-processing step to build personalized landmark metadata (i.e., local descriptors and a 3D landmark model). We validate our approach on the Multi-PIE and UHDB14 databases, and by comparing our results with those obtained using two existing methods.

1 Introduction

Landmark detection and localization are active research topics in the field of computer vision due to a variety of potential applications (e.g., face recognition, facial expression analysis, and human computer interaction). Numerous researchers have proposed a variety of approaches to build general facial landmark detectors. Most methods follow a learning-based approach where the variations in each landmark's appearance as well as the geometric relationship between landmarks are characterized based on a large amount of training data. These methods are typically generalized to solve the problem of landmark detection for any input image and not necessarily use any information from the gallery images. In contrast to the existing methods, we propose a personalized landmark localization framework. Our approach provides a framework wherein the problem of training across large data sets is alleviated, and instead focuses on the efficient encoding of landmarks in the gallery data and their counterparts in the probe image.

Our framework includes two stages. The first stage allows for off-line processing of the gallery data to generate an appearance model for each landmark and a 3D landmark model that encodes the relationship between the landmarks. The second stage is performed on-line on the input probe image to identify potential

R. Kimmel, R. Klette, and A. Sugimoto (Eds.): ACCV 2010, Part II, LNCS 6493, pp. 633–646, 2011.

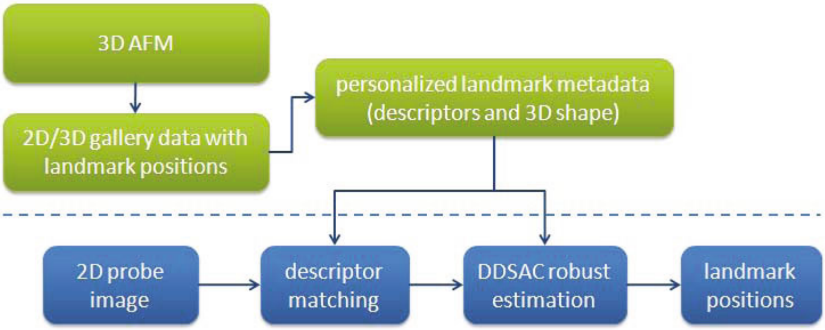
3D AFM
2D/3D gallery data with landmark positions
personalized landmark metadata (descriptors and 3D shape)
2D probe image
descriptor matching
DDSAC robust estimation
landmark positions

Fig. 1. Diagram of our personalized landmark localization pipeline. The upper part represents the gallery data processing and the bottom part depicts the landmark localization step in a probe image.

landmarks and establish a match between the models derived at the first stage using a constrained optimization scheme.

We propose several computational methods to improve the accuracy and efficiency of landmark detection and localization. Specifically, we extend the existing DAISY descriptor [1–3] and propose a hierarchical formulation that incorporates contextual information at multiple scales. Motivated by robust sampling approaches, we propose a Data-Driven SAmple Consensus (DDSAC) method for estimating the optimal landmark configuration. Finally, we introduce a hierarchical landmark searching scheme for efficient convergence of the localization process.

We test our method on the CMU Multi-PIE [4] and the UHDB14 [5] databases. Our results obtained are compared to the latest version of a method based on the Active Shape Model [6, 7] and a state-of-the-art classifier-based facial feature point detector [8, 9]. The experimental results demonstrate the advantages of our proposed solution.

2 Related Work

The problem of facial landmark detection and localization has attracted much attention from researchers in the field of computer vision. Most of the existing efforts toward this task follow the spirit of the Point Distribution Model (PDM) [10]. The implementation of a PDM framework includes two main subtasks: local search for PDM landmarks in the neighborhood of their current estimates, and optimization of the PDM configuration such that the global geometric compatibility is jointly maximized.

For the first subtask, most of the existing efforts are based on an exhaustive independent local search for each landmark using feature detectors. A common theme is to build a probability distribution for the landmark locations in order to reduce sensitivity to local minima. Both parametric [11, 12] and non-parametric representations [13] of the landmark distributions have been proposed. For the second subtask, most of the studies are based on a 2D geometric model with similarity or affine transformation constraints (e.g., [6, 11, 14–16]). Few studies (e.g., [17, 18]) have investigated 3D generative models and weak-perspective projection constraints. For the optimization process, most existing approaches use gradient descent methods (e.g., [6, 10]). A few researchers have also used MCMC [19, 20] and RANSAC [17] that theoretically are able to find the global optimal configuration, but in practice these methods are hampered by the curse of dimensionality. Furthermore, classifier-based (e.g., [8, 21]) and regression-based (e.g., [15, 22]) approaches have also been used. Across all methods, a large training database has been the key to enhancing the power of detecting landmarks and modeling their variations. For example, Liu *et al.* [19] use 280 distributions modeled by GMM to learn the landmark variations in both position and appearance. Similarly, Liang *et al.* [21] use 4,002 images for training classifiers and 120,000 image patches to learn directional classifiers for each facial component.

Our objective is to explore the possibility of leveraging the gallery data to reduce the cost of training. The study closest to our work is the recent investigation by Li *et al.* [23], where they use both the training data and a gallery image to build a person-specific appearance landmark model to improve the accuracy of landmark localization. However, there are several differences between our work and their work: (i) our method employs only one gallery image whereas Li *et al.* [23] also use additional training data and (ii) our method uses the 3D geometric relationship of landmarks to constrain the search.

3 Method

3.1 Framework

Our personalized landmark localization framework (URxD-L) includes two stages (Fig. 1). In the gallery processing step, personalized landmark metadata (i.e., landmark descriptor templates and a 3D landmark model) are computed. When only 2D images are available in the gallery, we use a statistical landmark model and 2D-3D fitting method to obtain a 3D personalized landmark model. Next, the landmark descriptor templates are computed on the projected landmark positions. If 3D data is also available in the gallery, we can obtain a personalized 3D landmark model directly from the annotated landmarks on the 3D data. We also use an Annotated Face Model (AFM) [24] to align the 3D face data and generate multiple images from a variety of viewpoints, which are then used to compute multi-view landmark descriptors. Landmark localization in a probe image is achieved through correspondence between the landmark descriptors in the gallery image and those in the probe image, constrained by the geometric relationship derived for the specific individual. This process includes efficient descriptor matching and robust estimation.

3.2 Matching Problem Formulation

In our framework, landmark localization is posed as an energy minimization problem. The energy function $E(X)$ depends on $X = \{x^j, j = 1, \cdots, n\}$, which denotes the configuration to be estimated, whereas x^j denotes the coordinates of the j^{th} landmark. Specifically, $E(X)$ is defined as a weighted sum of two energy terms:

$$E(X) = \lambda^a E^a(X) + \lambda^g E^g(X), \tag{1}$$

where λ^a and λ^g are non-negative scalar weights ($\lambda^a + \lambda^g = 1$), and the energy terms $E^a(X)$ and $E^g(X)$ denote the associated appearance and geometric terms, respectively. The energy term $E^a(X)$ is a sum of unary terms that favors correspondence between similar appearance descriptors across two images:

$$E^a(X) = \sum_{k \in K} \varphi_k, \tag{2}$$

where φ_k is the L_2 distance of the descriptor pair between two images, and K is the set of all correspondences between the gallery and probe images. The energy term $E^g(X)$ is a measure of geometric compatibility of correspondences between the gallery image and the probe image and it favors the correspondences that are compatible to the geometric relationship of facial landmarks (the shape of the landmarks) with a predefined transformation. In this paper, we use weak-perspective projection of the 3D landmark model to form the geometric constraints. Thus, $E^g(X)$ penalizes any deviation from these constraints and is given by:

$$E^g(X) = ||X - U||, \tag{3}$$

where U is the target configuration and X is estimated configuration. We define E^g as the L_2 distance between the projected landmarks from the 3D landmark template and their estimated positions in the probe image.

3.3 Personalized 3D Facial Shape Model

The objective of the gallery processing is to obtain a personalized 3D landmark model representing the geometry of the facial landmarks (of the gallery data), which can be detected automatically or through manual annotations. If the gallery includes 3D data, then this geometric relationship is constructed directly from the 3D landmark positions, obtained either manually or automatically [25]. In the case that the gallery contains only 2D images, the geometry is recovered through the use of a statistical 3D landmark model and a 2D-3D fitting process.

3D Statistical Landmark Model: A generic 3D statistical shape model is built based on the BU-3DFE database [26]. The BU-3DFE is a 3D facial expression database collected at State University of New York at Binghamton. We select only the datasets with neutral expressions, one per each of the 100 subjects. Considering the ambiguity of the face contour landmarks, we discard those and keep the 45 landmarks from different facial regions. By aligning the selected 3D landmarks and applying Principal Component Analysis, the statistical shape model is formulated as follows:

$$S = S_0 + \sum_{j=1}^{L} \alpha_j S^j, \tag{4}$$

where S_0 is the 3D mean landmark model, S^j is the j^{th} principal component, and L denotes the number of principal components retained in the model.

2D-3D landmark fitting process: A regular 3D-to-2D transformation can be approximated with a translation, a rotation and a weak-perspective projection:

$$P = \left[\begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \end{pmatrix} R_\gamma R_\zeta R_\phi \begin{matrix} t_x \\ t_y \end{matrix} \right], \tag{5}$$

where f is the focal length, t_x, t_y are the 2D translations, and $R_\gamma, R_\zeta, R_\phi$ denote image plane rotation, elevation rotation, and azimuth rotation, respectively. The configuration X is given by:

$$X = P(S_0 + \sum_{j=1}^{L} \alpha_j S^j). \tag{6}$$

Based on the correspondences between the 3D landmark model and the 2D gallery image with known landmark positions U, we can estimate the parameters of the transformation matrix P and the shape coefficients $\alpha = (\alpha^j, j = 1, \cdots, L)$ according to the maximum-likelihood solution as follows:

$$\{\hat{P}, \hat{\alpha}\} = \underset{P,\alpha}{\operatorname{argmin}} ||P(S_0 + \sum_{j=1}^{L} \alpha_j S^j) - U||, \hat{P} = U(S_0)^+, \hat{\alpha} = A^+(U - \hat{P}S_0), \tag{7}$$

where A is a matrix which is constructed by PS^j. The details are presented by Romdhani *et al.* [17].

When detecting landmarks in a probe image, the shape coefficients α are considered to be person-specific and are fixed, while only the transformation P needs to be estimated. The deformation of the personalized 3D face model due to expression variations is not considered in this work.

3.4 Hierarchical DAISY Representation of a Facial Landmark

A local feature descriptor is used to established a match between landmarks in the gallery and probe images. Inspired by the DAISY descriptor [1–3] and ancestry context [27], we propose a hierarchical DAISY descriptor that augments the discriminability of DAISY by including larger context information via a multi-resolution approach (Fig. 2). The contextual cues of facial landmarks are naturally encoded in the face image with this hierarchical representation that relates different components of a face (i.e., nose, eyes and mouth).

We construct the hierarchical DAISY descriptor of a facial landmark by progressively enlarging the window of analysis, which is effectively achieved by decreasing image resolution. Given a face image and a landmark location, we

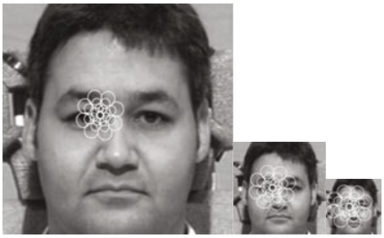

Fig. 2. The hierarchical daisy descriptor for the left inner eye corner. (a-c) Images of different resolutions with the corresponding DAISY descriptors.

define the ancestral support regions of a landmark as the set of regions that are obtained by sequentially decreasing the image resolution, covering not only the local context information but also global context information. Keeping the patch size (pixel units) of the DAISY descriptor constant, we compute the DAISY descriptors in these ancestral support regions. Then, these descriptor vectors are concatenated to form a hierarchical DAISY descriptor.

3.5 Data-Driven SAmple Consensus (DDSAC)

We propose a Data-Driven SAmple Consensus (DDSAC) that is inspired by the principle of RANSAC (RANdom SAmple Consensus) and importance sampling. Although RANSAC has been successfully used for many computer vision problems, several issues still remain open [28]. Examples include the convergence to the correct estimated transformation with fewer iterations, and tuning the parameter that distinguishes inliers from outliers (this parameter is usually set empirically).

With these open issues in mind, we propose a Data-Driven SAmple Consensus (DDSAC) approach, which follows the same hypothesize-and-verify framework but improves the traditional RANSAC by using (i) data-driven sampling and (ii) a least median error criterion for selecting the best estimate. The image observations are used as a proposal sampling probability to choose a subset of landmarks, in order to use fewer iterations to obtain the correct landmark configuration. The proposal probability is designed so that it is more likely to choose a landmark subset that is more compatible with the geometric shape of the landmarks as well as their appearance.

Given a probe image I, we first search for the best descriptor matching candidates of the landmarks $\{y^j = (u^j, v^j), j = 1, \cdots, n\}$ where u^j and v^j are the position and the descriptor of the candidate of the j^{th} landmark, respectively. If we construct a template tuple by $B^j = (x^j, \Gamma^j)$, where Γ^j is the descriptor template and x^j is the initial position (or estimated position from the previous iteration), the proposal probability of choosing j^{th} landmark $P(B^j; y^j)$ can be formulated as:

$$P(B^j; y^j) = P(x^j, \Gamma^j; u^j, v^j) = \beta p(x^j|u^j) + (1 - \beta)q(\Gamma^j|v^j), \tag{8}$$

where β is the weight of the probability derived from the geometry-driven process $p(x^j|u^j)$, and $(1-\beta)$ is the weight for the probablity based on appearance $q(\Gamma^j|v^j)$.

The term $p(x^j|u^j)$ in Eq. 8 is given by:

$$p(x^j|u^j) = \frac{N(u^j; x^j, \sigma)}{\sum_{r=1,\cdots,n} N(u^r; x^r, \sigma)}, \tag{9}$$

where $N(u^j; x^j, \sigma)$ is the probability that the j^{th} landmark point is located at u^j given the landmark initial position x^j. In this work, we model $N(u^j; x^j, \sigma)$ using a Gaussian distribution with mean the initial position x^j. In order to reduce the sensitivity to initialization, we use robust estimation to compute x^j as follows. First, a large search area is defined to obtain the best descriptor match of the landmarks. Second, we compute the distances between these match locations and the initial landmark positions. Finally, the initial configuration X is translated based on the match that yields the median of these distances.

Algorithm 1. Data-Driven SAmple Consensus (DDSAC)

```
Input: A tentative set of corresponding points from descriptor template
matching and a geometric constraint (i.e., personalized 3D landmark model and
a weak-perspective transformation)
  1. Select elements with the proposal probability defined by Eqn. 8
  2. Combine the selected points and delete the duplicates to obtain a subset of
     samples
  3. Estimate the transformation parameters (i.e., generate a hypothesis)
  4. Project the template shape and compute the median of errors between the
     projected locations and locations of landmark candidates
  5. Repeat Steps 1-4 for $m$ iterations
  6. Select the estimated configuration with the least median error, and sort the
     landmark candidates by the distances between their positions and this
     configuration
  7. Re-estimate the configuration based on the first $k$ landmark candidates in the
     list sorted in ascending order
Output: The estimated transformation and matching results
```

The second term in Eq. 8, $q(\Gamma^j|v^j)$, defines the uncertainty of the best descriptor match position for the j^{th} landmark based on the descriptor similarity. It can be formulated as follows:

$$q(\Gamma^j|v^j) = \frac{P(v^j; \Gamma^j, \Omega^j)}{\sum_{r=1,\cdots,n} P(v^r; \Gamma^r, \Omega^r)} \tag{10}$$

where $P(v^j; \Gamma^j, \Omega^j)$ is the probability of j^{th} landmark point at position u^j given the landmark descriptor template Γ^j. A Gaussian distribution is also used here to define $P(v^j; \Gamma^j, \Omega^j)$ with the mean being the descriptor template from the gallery image. The pipeline for estimating the landmark matches and the shape transformation is presented in Alg. 1.

Algorithm 2. 2D gallery processing

```
Input: A 2D face image associated with a subject ID and the annotated landmarks
  1. Build a personalized 3D landmark model (geometric relationship of landmarks) by using landmark correspondences between the 2D images and the 3D statistical landmark model (Eq. 7)
  2. Compute the personalized descriptors of the annotated landmarks from the 2D image
Output: Personalized metadata (landmark descriptors and a 3D landmark model).
```

Algorithm 3. 3D gallery processing

```
Input: A 3D textured data associated with a subject ID and the annotated landmarks
  1. Align the 3D face data with the Annotated Face Model (AFM) [24]
  2. Orthographically project the 3D texture surface and landmark positions to predefined viewing angles
  3. Compute the descriptors at the projected locations of the landmarks on the projections
Output: Personalized metadata (multi-view landmark descriptors and a 3D landmark model).
```

In our experiments, we empirically set $m = 50$, $\beta = 0.8$, $\sigma = 0.1W$ (W being the width of the face image) and $\Omega^j = 10$. Figure 5(b) depicts the effect of selecting k or the ratio k/n on the results. Based on the sampled subset of the tentative correspondences, we are able to estimate the parameters in the weak-perspective matrix P with the maximum-likelihood solution described in Section 3.3.

3.6 Gallery/Probe Processing

The algorithms for 2D gallery, 3D gallery, and probe processing are presented in Alg. 2, Alg. 3, and Alg. 4, respectively. For 2D gallery image processing (frontal face image without 3D data), we run the ASM algorithm [6, 7] to automatically obtain 68 landmarks in order to reduce the annotation burden. We selected among them 45 landmarks from eyebrows, eyes, nose and mouth that have correspondences in our 3D statistical landmark model. Then, we manually correct the erroneous annotations from ASM to obtain accurate ground truth. For 3D gallery face data, we manually annotate each of the landmarks on the 3D surface. Given a probe image, we use the hierarchical optimization method to localize the landmarks from coarse level to fine level as presented in Alg. 4.

Algorithm 4. Hierarchical landmark localization on a probe image

```
Input: A 2D probe face image and the corresponding gallery metadata

  1. Scale the gallery landmarks to match the size of the face region
  2. Align the geometric center of the landmarks with that of the face image
  3. Optimize from coarse to fine resolution:

     a. Generate candidates by the descriptor template matching
     b. Estimate the translation by the median of distances between candidates
        and the initialization
     c. Run DDSAC to obtain matching results and estimate the transformation

Output: Estimated landmark locations
```

4 Experiments

We evaluated our method on the Multi-PIE [4] and the UHDB14 [5] databases. Both databases include facial images with pose and illumination variations. The Multi-PIE dataset includes only 2D images while the UHDB14 includes both 3D data and the associated texture images for all subjects. We compare the results of our method to those obtained by the latest ASM-based approach [6, 7] and a classifier-based facial landmark detector [8, 9].

4.1 Multi-PIE

The Multi-PIE database collected at CMU contains facial images from 337 subjects in four acquisition sessions over a span of six months, and under a large variety of poses, illumination and expressions. In our experiments, we chose the data from 129 subjects who attended all four sessions. We used the frontal neutral face images in Session 1 under Lighting 16 (top frontal flash) condition as the gallery, and the neutral face images of the same subjects with different poses and lighting conditions in Session 2 as the probes. The probe data were divided into cohorts defined in Table 1, each with 129 images from 129 subjects.

We obtained 45 landmarks in gallery images by using the method described in Section 3.6. In addition, eight landmarks on the probe images were manually annotated as the ground truth for evaluation. The original images were cropped to 300 × 300 pixels, ensuring that all landmarks are enclosed in the bounding box. Examples of gallery and probe images are depicted in Fig. 3.

Table 1. Probe subset definitions of the Multi-PIE database

		Pose 051 (0°)	Pose 130 (30°)
lighting	16 (top frontal flash)	C1	C2
	14 (top right flash)	C3	C4
	00 (no flash)	C5	C6

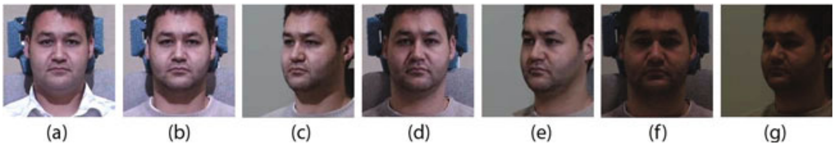

Fig. 3. Examples of gallery and probe images from Multi-PIE. (a) Gallery image from Session 1 with Light 16 and Pose 051 (b-g) Probe images from Session 2 with same ID as the gallery. Example images with Light 16 and Pose 051 (b), Light 16 and Pose 130 (c), Light 14 and Pose 051 (d), Light 14 and Pose 130 (e), Light 00 and Pose 051 (f), Light 00 and Pose 130 (g).

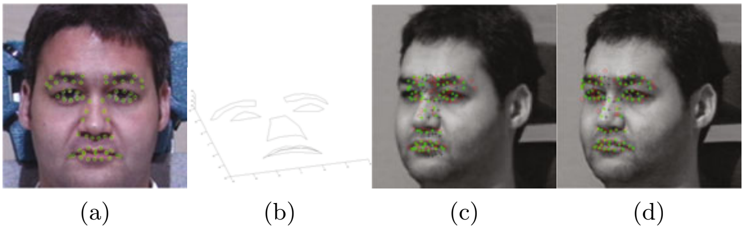

Fig. 4. Gallery image processing and landmark localization on a probe image. (a) Gallery image with annotations (yellow circles) and fitting results (green crosses). (b) Personalized 3D landmark model. (c-d) Matching results [initialization (black crosses), best descriptor matches (red circles), geometrical fitting (green crosses)] on two levels of resolutions, respectively.

The matching process is depicted in Fig. 4. First, the landmark descriptors are computed on the gallery image and the 3D personalized landmark model is constructed from fitting the 3D statistical landmark model, as illustrated in Figs. 4a and 4b. Then, from coarse to fine resolution, the landmarks are localized by descriptor matching and geometrical fitting, as illustrated in Figs. 4c and 4d.

The experimental results are shown in Figs. 5 and 6. All of the results are evaluated in terms of the mean of the normalized errors, which are computed as the ratio of the landmark errors (i.e., distance between the estimated landmark position and the ground truth) and the intra-pupillary distance. The average intra-pupillary distance is about 72 pixels among the frontal face images (cohorts: C1, C3 and C5), and about 62 pixels among images with varying poses (cohorts: C2, C4 and C6).

Figure 5(a) depicts the results of our method (URxD-L) using different levels of the DAISY descriptor. It can be observed that the two-level descriptor with larger context region results in an improvement over the one-level descriptor in all probe subsets. Figure 5(b) depicts the effect of ratio (k/n) of landmarks used in DDSAC on the estimation results, where k and n are the number of selected landmarks and the total number of landmarks, respectively. This finding suggests that robust estimation (ratio 0.9-0.4) is better than non-robust estimation (ratio 1.0) in dealing with outliers. Note that the results on probe subsets with frontal

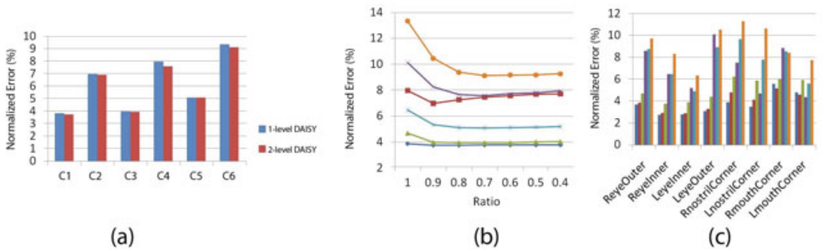

Fig. 5. (a) Normalized error of URxD-L using different levels of DAISY descriptor (b) Effect of ratio (k/n) of landmarks on the estimation results (c) Means of normalized errors of different landmarks under different poses and lighting conditions.

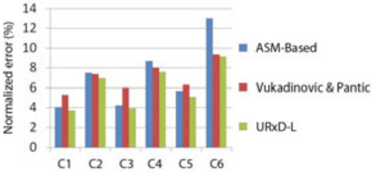

Fig. 6. Comparison among ASM-based [6, 7], Vukadinovic & Pantic detector [8, 9] and URxD-L on Multi-PIE

pose (same as the gallery) have "flatter" curves with lower error rates than other probe subsets. This can be attributed to the fact that frontal face images have less outliers than the non-frontal face images. Figure 5(c) illustrates the normalized localization errors of 8 landmarks (right/left eye outer/inner corners, right/left nostril corners, and right/left mouth corners). We can observe that the eye inner corners are more reliable landmarks in frontal face images. As expected, for face images that are non-frontal (e.g., 30° yaw) the landmarks on the left face region are more reliable than those on the right side except for left eye outer corner, which is often occluded by glasses.

Figure 6 presents a comparison of landmark localization using three different methods, including the latest ASM-based approach [6, 7], a classifier-based landmark detector (Vukadinovic & Pantic) [8, 9], and URxD-L. The results show that our method can achieve similar or better results than the ASM-based and Vukadinovic & Pantic detector. They also suggest that a gallery image and robust descriptors are very helpful in reducing training burden and achieving good results.

4.2 UHDB14 Database

To evaluate the performance of URxD-L for the case where 3D data are available in the gallery, we collected a set of 3D data with associated 2D texture images using two calibrated 3dMD cameras [29] and a set of 2D images of the same subjects using a Canon XTi DSLR camera. This database, which we name

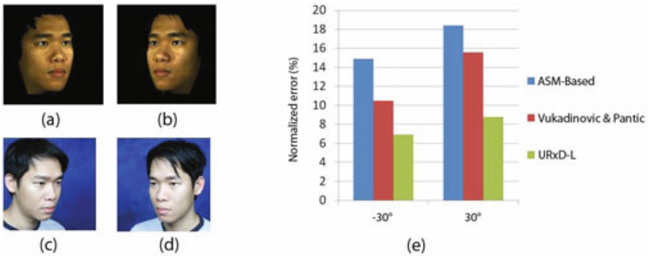

Fig. 7. (a-d) Examples of gallery and probe cohorts from UHDB14. Gallery images generated from the 3D data (a) -30° yaw and (b) 30° yaw. Probe images (c) -30° yaw and (d) 30° yaw. (e) Comparison among ASM-based [6, 7], Vukadinovic & Pantic detector [8, 9] and URxD-L on UHDB14.

UHDB14 [5], includes 11 subjects (3 females and 8 males) with varying yaw angles (-30°, 0°, 30°). We use the textured 3D data as the gallery, and the 2D images with yaw angles -30° and 30° as our probe cohorts. Since the data were obtained with different sensors, the illuminations of the gallery and the probe images are different. Some sample images from the gallery and probe cohorts are shown in Figs. 7(a-d). We manually annotated 45 landmarks on the gallery 3D data around each facial components (eyebrows, eyes, nose and mouth), and 6 landmarks on the probe images as the ground truth for evaluation. The average intra-pupillary distances of the two probe cohorts with different yaw angles are 72 and 54 pixels, respectively.

We evaluated our method on the 6 annotated landmarks, and compared the results against the ones from ASM-based [6, 7] and Vukadinovic & Pantic detector [8, 9] in terms of average normalized error. Figure 7(e) presents the comparison in terms of the mean of normalized landmark errors. Note that our method outperforms both methods.

5 Conclusions

In this paper, we have proposed a 2D facial landmark localization method URxD-L that leverages a 2D/3D gallery image to carry out a person-specific search. Specifically, a framework is proposed in which descriptors of facial landmarks as well as a 3D landmark model that encodes the geometric relationship between landmarks are computed from processing a given 2D or 3D gallery dataset. Online processing of the 2D probe, targeted towards the face authentication problem [30], focuses on establishing correspondences between key points on the probe image and the landmarks on the corresponding gallery image. To realize this framework, we have presented several methods, including a hierarchical DAISY descriptor and a Data-Driven SAmple Consensus algorithm. Our approach is compared with two of the existing methods and it exhibits a better performance.

Acknowledgement. This work was supported in part by the US Army Research Laboratory award DWAM80750 and UH Eckhard Pfeiffer Endowment. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of our sponsors.

References

1. Tola, E., Lepetit, V., Fua, P.: DAISY: An efficient dense descriptor applied to wide baseline stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 815–830 (2010)
2. Winder, S., Brown, M.: Learning local image descriptors. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, pp. 1–8 (2007)
3. Winder, S., Hua, G., Brown, M.: Picking the best DAISY. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 178–185 (2009)
4. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands (2008)
5. UHDB14, `http://cbl.uh.edu/URxD/datasets/`
6. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
7. STASM, `http://www.milbo.users.sonic.net/stasm/`
8. Vukadinovic, D., Pantic, M.: Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In: Proc. IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, Hawaii, USA, pp. 1692–1698 (2005)
9. FFPD, `http://www.doc.ic.ac.uk/~maja/`
10. Cootes, T., Taylor, C.: Active shape models: Smart snakes. In: Proc. British Machine Vision Conference, Leeds, UK, pp. 266–275 (1992)
11. Gu, L., Kanade, T.: A generative shape regularization model for robust face alignment. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 413–426. Springer, Heidelberg (2008)
12. Wang, Y., Lucey, S., Cohn, J.: Enforcing convexity for improved alignment with constrained local models. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, AK (2008)
13. Saragih, J., Lucey, S., Cohn, J.: Face alignment through subspace constrained mean-shifts. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 1034–1041 (2009)
14. Cootes, T., Walker, K., Taylor, C.: View-based active appearance models. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 227–232 (2002)
15. Cristinacce, D., Cootes, T.: Boosted regression active shape models. In: Proc. British Machine Vision Conference, University of Warwick, United Kingdom, pp. 880–889 (2007)
16. Liang, L., Wen, F., Xu, Y., Tang, X., Shum, H.: Accurate face alignment using shape constrained Markov network. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, pp. 1313–1319 (2006)

17. Romdhani, S., Vetter, T.: 3D probabilistic feature point model for object detection and recognition. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, MN (2007)
18. Gu, L., Kanade, T.: 3D alignment of face in a single image. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, pp. 1305–1312 (2006)
19. Liu, C., Shum, H.Y., Zhang, C.: Hierarchical shape modeling for automatic face localization. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 687–703. Springer, Heidelberg (2002)
20. Tu, J., Zhang, Z., Zeng, Z., Huang, T.S.: Face localization via hierarchical condensation with fisher boosting feature selection. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 719–724 (2004)
21. Liang, L., Xiao, R., Wen, F., Sun, J.: Face alignment via component-based discriminative search. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 72–85. Springer, Heidelberg (2008)
22. Valstar, M., Martinez, B., Binefa, X., Pantic, M.: Facial point detection using boosted regression and graph models. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, pp. 2729–2736 (2010)
23. Li, P., Prince, S.: Joint and implicit registration for face recognition. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 1510–1517 (2009)
24. Kakadiaris, I., Passalis, G., Toderici, G., Murtuza, M., Lu, Y., Karampatziakis, N., Theoharis, T.: Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 640–649 (2007)
25. Perakis, P., Passalis, G., Theoharis, T., Toderici, G., Kakadiaris, I.: Partial matching of interpose 3D facial data for face recognition. In: Proc. 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems, Arlington, VA, pp. 439–446 (2009)
26. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.: A 3D facial expression database for facial behavior research. In: Proc. 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, pp. 211–216 (2006)
27. Lim, J., Arbelaez, P., Gu, C., Malik, J.: Context by region ancestry. In: Proc. 12th IEEE International Conference on Computer Vision, Kyoto, Japan, pp. 1978–1985 (2009)
28. Proc. 25 Years of RANSAC (Workshop in conjunction with CVPR), New York, NY (2006)
29. 3dMD, http://www.3dmd.com/
30. Toderici, G., Passalis, G., Zafeiriou, S., Tzimiropoulos, G., Petrou, M., Theoharis, T., Kakadiaris, I.: Bidirectional relighting for 3d-aided 2d face recognition. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 2721–2728 (2010)