Face Recognition Robust to Head Pose Changes Based on the RGB-D Sensor

Cesare Ciaccio, Lingyun Wen, and Guodong Guo*
West Virginia University, Dept. of Computer Science & Electrical Engineering, Morgantown, WV
cciaccio@mix.wvu.edu, lwen@mix.wvu.edu, guodong.guo@mail.wvu.edu

Abstract

Face recognition is still a great challenge in biometrics research, because of the large variations of facial appearance, caused by head pose, lighting, facial expression, aging, etc. Among all possible variations, the biggest change of facial appearance in 2-dimensional (2D) face images probably comes from the three-dimensional (3D) head rotations. With the sensor technology advances, e.g., the recent RGB-D cameras, we study the advantage of using RGB-D images for face recognition, focusing on the challenge of 3D head pose variations. We propose an approach to face recognition robust to head rotations utilizing the RGB-D face images. Unlike the traditional 3D morphable model, our method does not need to learn a generic face model or take a complicated 3D to 2D face registration. We study what is the appropriate scheme to deal with pose variations in order to develop a robust system towards pose-invariant face recognition. Experiments on a public database show that our approach is effective and efficient for face recognition under significant pose changes. Our preliminary result demonstrates the advantages of using the RGB-D sensor for face recognition robust to large pose variations.

1. Introduction

Face recognition is important and useful for many applications, including homeland security, video surveillance, law enforcement, and identity management. Researchers have made significant progresses for face recognition in the biometrics society. A number of face recognition methods have been developed with good performance. However, it is still a great challenge for face recognition in practical applications, because of many variations of facial appearance, caused by head pose, lighting, facial expression, aging, etc. Among all possible variations, the biggest change of facial appearance in 2-dimensional (2D) face images is probably caused by the three-dimensional (3D) head rotations.

Researchers have developed many methods to deal with head pose variations in face recognition. The classical approach is to learn a 3D face model, and then register and match the 3D model to a given 2D face image. For example, the 3D morphable model [4] learns a generic 3D face model from a number of subjects. However, the 3D to 2D registration is not trivial. The optimization for registration is a time-consuming process because it involves many parameters to adjust. It also needs a good initialization of the head pose to start the optimization process. Another representative approach is to use local patches for 2D face matching with pose variations, e.g., [2]. This category of methods is useful for small rotation angles, e.g., up to 30 or 40 degrees of rotations, but it still has difficulty to handle larger head rotations. Another typical approach is to learn the mapping between 2D face images at different angles, and then apply the mapping for test faces for matching between different pose angles, e.g., [11].

The fact is that face recognition under large head pose changes is still an unsolved problem, even though different techniques have been developed in the literature. In this work, we study the advantage of utilizing the recently developed RGB-D sensors to develop a method towards pose-invariant face recognition.

The RGB-D cameras can capture face images with both the color and depth information. Using the 3D depth data, the face images can be rotated directly in 3D to render face images at any pose angles. Then face recognition can be performed at the same or similar pose angles between the probe and gallery faces. As a result, our method does not need to learn a generic face model like the traditional 3D morphable model, or execute a complicated 3D to 2D face registration. This is a great advantage of the RGB-D sensors for face recognition insensitive to head pose changes.

We study what is the appropriate scheme for head rotation in order to develop a robust system towards pose-invariant face recognition. In our approach, only one frontal view face image (with the corresponding depth image) is used as the gallery for each subject, while the probe face can be at any pose angles. In this way, we can fully explore the advantage of the RGB-D images for face recognition

*Corresponding author. This work is partially supported by an NSF CITeR grant.

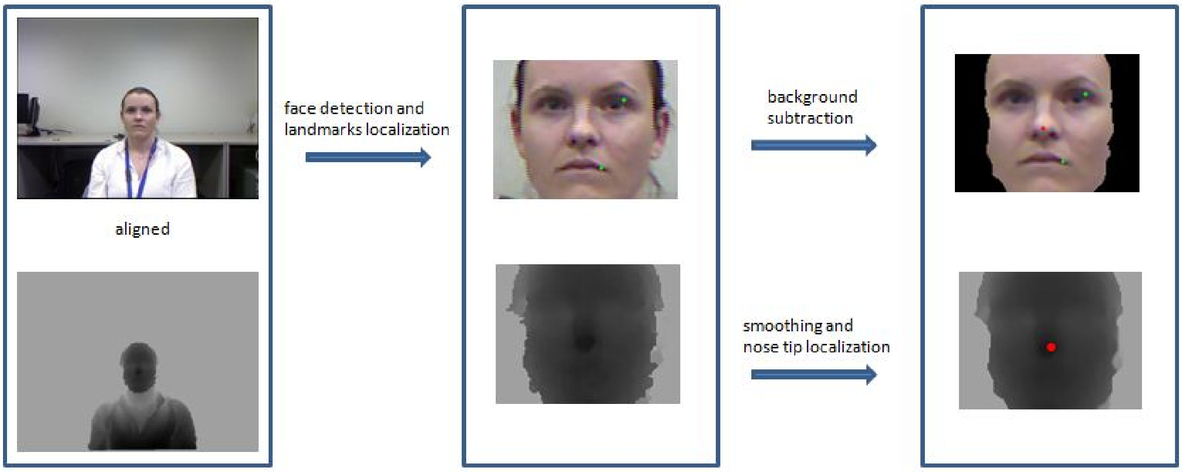

Figure 1. Automated face detection and some landmark localizations in a RGB-D image.

robust to head pose changes.

There are some recent works using the RGB-D sensor for face detection [9, 6], gender classification [7], or face recognition [8, 10]. These works mainly focus on the fusion of features extracted from both color and depth images for an improved detection or recognition, compared to the traditional color-only approaches. However, these works do not explore the advantages of using the depth information for pose-invariant face recognition, which is the major focus of our study. Only the depth images were used for face recognition in [10]. The sparse representation classifier was used for face recognition in [8], which requires multiple face images for each individual in training. It is also slow in their approach using the ICP algorithm [3] to register each face to a face model obtained from a laser scanner.

Our major contributions in this paper include (1) a study of pose-invariant face recognition based on the RGB-D face images; (2) an appropriate scheme for head rotation in RGB-D images; and (3) a new face representation is developed based on an integration of the covariance descriptor [13] with the popular local binary patterns (LBP) [1] for face recognition with improved accuracies.

In the following, we will present our approach in Section 2, and experimental validations in Section 3. Finally, we draw conclusions.

2. Our Approach

Our study focuses on face recognition robust to large head pose changes, discovering the advantages of using the RGB-D face images. We assume that the gallery contains only one example face in a frontal view for each enrolled subject, while the probe faces can be at any pose angles. This is a very challenging problem for traditional 2D image based face recognition. It is quite common to use a single image per person in the gallery for face recognition [12].

Our approach is to utilize the RGB-D face images to render new face images at any pose angles. Multiple face images can be generated even though there is only one frontal view face image in the gallery. The query or test face images can be matched to the face images rendered from the single RGB-D face image in the gallery.

To develop an automated system and achieve a good face recognition result, our approach contains several steps that will be described in the following.

2.1. Face Detection and Landmark Localization

Face detection and some facial fiducial point detection are performed first on the RGB-D images. Figure 1 illustrates the detection process, assuming that the RGB-D images are aligned already by the sensor, i.e., the color and depth face images are registered. The process consists of face detection and some landmark localizations. Face detection is performed on the color image only, using a recently developed technique [15]. The landmark points that we are interested in include the eye corners, the mouth corners and the nose tip, which can also be detected by the method presented in [15]. However, we found that a better result is obtained by using the depth image to localize the nose tip. It is the closest point to the RGB-D camera. We obtained more accurate location of the nose tip from the depth map, which is very important for our head rotation to render new face images.

The depth values are noisy from the cheap RGB-D cameras, e.g., the Microsoft Kinect. We filter the depth image with a median filter and then a Gaussian filter. This filtering is important to improve the landmark detection and other things in the following steps. And also, we used the depth information to perform background subtraction on the RGB face images to remove the background pixels. One example is shown in Figure 1 to illustrate the process for a gallery face image.

2.2. Head Rotations

From each face image (RGB-D) in the gallery, our system generates a set of face images with different pose angles automatically. To do this, we first compute the center of the head based on the nose tip in the depth map. We estimate the center of the head by

$$(x_0, y_0, z_0) = (n_t(x), n_t(y), n_t(z) + \delta)\,, \tag{1}$$

where n_t denotes the "nose tip," and δ is the distance to move from the nose tip to get the center of the head for rotations. We set $\delta = 50mm$ in our experiments, and found that this setting can work well for all faces in the whole database. Each head was rotated around the Y axis of the coordinate system with the center of the head computed by Eq. (1).

Our simple but effective approach to head rotations can be illustrated in Figure 2, and is presented in Algorithm 1 in detail. In the algorithm, T is the transformation, given by

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & x_0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) & z_0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x - x_0 \\ y \\ z - z_0 \\ 1 \end{bmatrix} \tag{2}$$

to rotate a 3D point (x, y, z) on the original face surface to a new location (x', y', z') with a yaw angle θ. The center of the head is at (x_0, y_0, z_0). In theory, the rotations can be any angles in yaw, pitch, and roll directions. In this study, we mainly focus on the rotations in yaw.

Algorithm 1 Rendering of rotated face images

```
procedure FACE IMAGE RENDERING
    for each gallery image do
        estimate center of the head $(x_0, y_0, z_0)$
        for $\theta = 5$ to $\theta = 90$ do
            create a blank image for the target face
            for each foreground pixel $(x, y)$ in the frontal view image do
                $(x', y') = T(x, y, Depth(x, y), \theta)$
                $Image(x', y') \leftarrow Frontal\_view(x, y)$
            end for
            do interpolation on the rendered image
        end for
    end for
end procedure
```

We found that the best rotation is from the frontal view to any other pose angles. To testify this, we tried other ways of head rotations, and compared with this scheme (see experiments). One example of the rendered faces with rotations for every 5 degrees is shown in Figure 3.

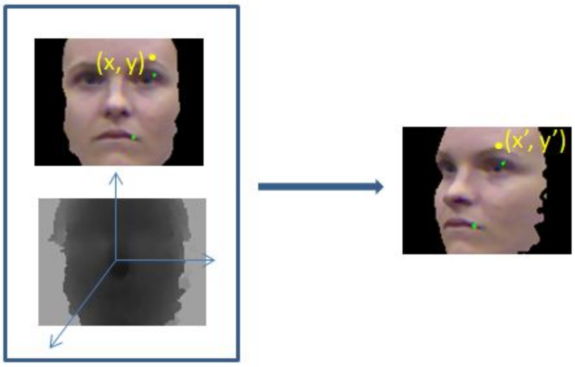

Figure 2. Illustrate the 3D head rotations of a face image to a different angle.

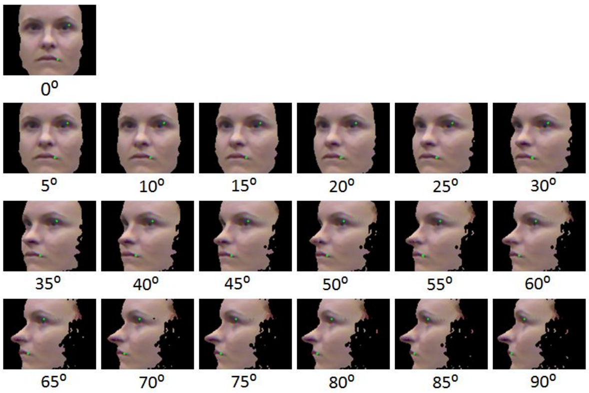

Figure 3. The rendered color faces at different angles with an interval of 5 degrees. The input face is only one frontal view at 0 degree.

2.3. Face Alignment

To align the face images, we use the eye corners and mouth corners. Since the eye and mouth corners are not detected precisely sometimes, especially in the horizontal direction when the head rotations are large (e.g., for test face images), we use the vertical distance between the eyes and mouths for the alignment. All faces will be aligned to have the same vertical distance between the eyes and mouth. Note that traditionally the two eyes are used for face alignment, which is not feasible in our case when the head rotations are close 90 degrees. One eye and one mouth corner cannot be viewed from the profile views for alignment.

After the alignment, each face is cropped and resized to 60×60 for matching.

2.4. Patch Selection

Facial patches are selected for feature extraction and face matching. Given the normalized face size, we use the patch size of 10×10. A step size of 5 pixels is used to sample the patches in face images. The number of selected patches might be different in each face image, depending on the pose angles. When a face is rotated from a frontal view to a large angle, some facial parts will be self-occluded. The rendered face images have the known angles that can be used to control the patch selection.

When a query face is given, it will be matched against each gallery face image with the selected patch locations controlled by the pose angles. The similarity measures over the selected patches will be integrated together, and normalized by the number of patches used to obtain the similarity score between a query and a gallery face image.

2.5. Face Representation

We develop a new representation of the face images, which is a combination of two different descriptors. One is the popular local binary patterns (LBP) [1] feature, and the other is the covariance descriptor [13]. The covariance descriptor has been used successfully for pedestrian detection and object tracking [14], but seldom for facial recognition [5]. We explore whether these two features can be complementary and if a better representation can be obtained by integrating the two different descriptors.

Since the LBP feature [1] is well-known for face recognition, we will not describe it here. In the following, we introduce the covariance descriptor, and then present a probabilistic integration of the two features for face representation.

2.5.1 Covariance Descriptor

The formal definition of the covariance descriptor was first described in [13] for object detection and classification. The usage of covariance matrices as a region descriptor provides some advantages. One of them is that the representation proposes a natural way of fusing multiple features [14].

Given an image (or patch) I, the $W \times H \times d$ dimensional feature image F extracted from I is:

$$F(x, y) = \phi(I, x, y), \qquad (3)$$

where function ϕ can be any mapping such as intensity, color, gradients, filter responses, etc. In our work, we use the same settings as used in [14] to form the feature vector at each pixel, (x, y),

$$\left[x \;\; y \;\; |I_x| \;\; |I_y| \;\; \sqrt{I_x^2 + I_x^2} \;\; |I_{xx}| \;\; |I_{yy}| \;\; arctan\frac{|I_x|}{|I_y|}\right]^T . \qquad (4)$$

The pixel location, intensity derivatives, and the edge orientation are computed as the feature vector for each pixel.

For a given rectangular region $R \subset F$, $\{z_i\}_{i=1..S}$ are the d-dimensional feature vectors at the points inside R. Here $d = 8$. The region R is represented with the $d \times d$ covariance matrix of the feature points,

$$C_R = \frac{1}{S-1} \sum_{i=1}^{S} (z_i - \mu)(z_i - \mu)^T, \qquad (5)$$

where μ is the mean of the points, and S is the number of points in region R.

Covariance matrices are symmetric and positive semi-definite, hence they reside in the Riemannian manifold [14]. The distances between two covariance matrices can be calculated in the Riemannian manifold. The distance $d(X, Y)$ between the points X and Y on the manifold is given by the length of geodesic, which is the minimum length curve connecting two points on the manifold.

The derivatives at X on the manifold lie in the tangent space. From X, there exists a unique geodesic starting with the tangent vector y. The tangent space can be thought of as the set of allowable velocities for a point constrained to move on the manifold. Two operators, namely the exponential exp_X and logarithm maps $log_X = exp_X^{-1}$, are defined over the Riemannian manifold to switch between manifold and tangent space at X [14]. The exponential map maps vector y to the point Y on the manifold surface. So the distance of the geodesic is given by $d(X, Y) = d(X, exp_X(y)) = ||y||_X$. The inner product induces a norm for the tangent vectors in the tangent space such that $||y||_X^2 =< y, y >_X$.

So the covariance descriptor is very different from the classical LBP feature. These two features cannot be concatenated directly into a new feature vector. We propose a probabilistic integration of the two features.

2.5.2 New Representation based on Integration

We develop a probabilistic integration scheme to combine the covariance descriptor with the LBP feature. Our integration is to transform the distance measure from each feature descriptor into probabilities, and then perform a probabilistic multiplication.

Let d_1 represent the distance measure between a query and a gallery face based on the covariance descriptor. Let d_2 be the distance measure based on the LBP feature, then the integration is given by

$$P(d_1, d_2) = A \cdot \exp\left(\frac{-d_1}{\lambda_1}\right) \cdot \exp\left(\frac{-d_2}{\lambda_2}\right), \qquad (6)$$

where the parameters λ_1 and λ_2 are used to balance the contributions from the two different features. In our ex-

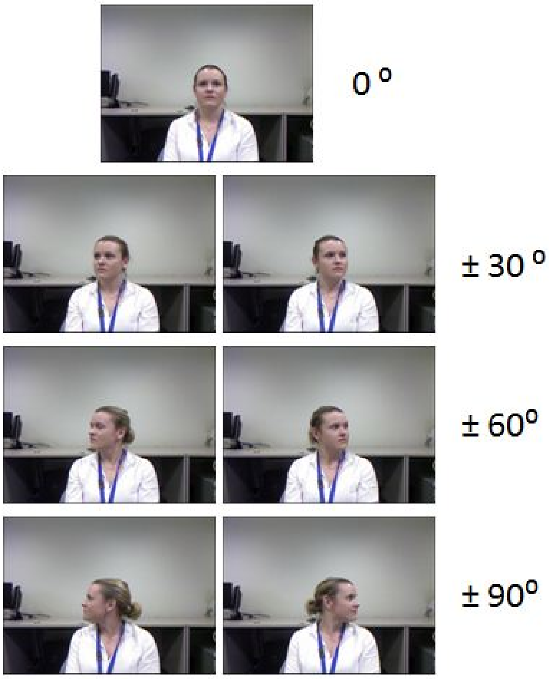

Figure 4. An example subject in the database with different head pose angles.

periments, we set $\lambda_1 = 1$ and $\lambda_2 = 10$. A is a constant to maintain a probabilistic measure.

Based on the probabilistic integration, we derive a new representation of face images for matching that combines different descriptors.

3. Experiments

We conduct experiments to evaluate our proposed approach for face recognition robust to head pose changes. The database is introduced first, and then the experimental results are presented.

3.1. Dataset

There are no standard benchmark databases for RGB-D sensor based face recognition. We used the CurtinFaces database [8] for our experiments, which is publicly available. This dataset contains 52 subjects with variations in pose, illumination, facial expression, and sunglasses disguise. In our study, we use a subset of this database to study pose-invariant face recognition. Specifically, for each subject we use only the frontal view face image with neutral expression as the gallery, while the faces with pose angles different from the 0 degree are used as the query images. The dataset contains pose angles at 6 different yaw directions, i.e., $(\pm 30, \pm 60, \pm 90)$. See Figure 4 for an example subject.

3.2. Results

The experimental results are shown in Tables 1, 2, and 3. Different face recognition experiments are designed to explore the appropriate approach to face recognition robust to head pose changes.

How to rotate the head. In our first experiment, we evaluated two different schemes for head rotations to align the probe and gallery poses for face recognition. One approach is to rotate the query face image to the frontal view (others →front), and then match to the gallery face image in a frontal view; the other approach is the opposite, i.e., each gallery face image is rotated to the other pose angles (front→others in Table 1) to generate face images with pose angles at 30, 60, and 90 degrees, respectively. Then each probe face is matched to all generated face images for identification. For this comparison, we used the LBP feature only.

As shown in Table 1, the second scheme is significantly better than the first one, especially for large pose variations, e.g., 60 and 90 degrees. There are two possible reasons for this result. First, the frontal view of a face image carries more information about the identity compared to other views with large angles. When we rotate the frontal view face image, we simply reduce some information. On the other hand, when we rotate the other views to the frontal, we face the problem of missing information that needs to be generated. The second reason is due to a practical concern. In order to perfectly rotate the profile view to frontal, we need to know the exact pose angle. But this is not trivial to measure from face images. Also, estimating the center of the head in other angles is more difficult than a frontal view. We observe that when rotating the frontal face, we can use the nose tip to get a good estimation of the center of the head, and the nose tip in a frontal view can be estimated accurately in the depth map. When we rotate the profile face, it will be difficult to get the head center.

How dense the rotations. When the head is rotated from the frontal view, i.e., frontal→others, there is a need to do pose estimation for the probe faces in order to rotate the (frontal view) gallery images properly. In our second experiment, we show that this problem can be mitigated by creating multiple face images with many different angles, and then compare the probe face to all these rendered face images. The one with the highest match score will be used as the identification result. Table 2 shows the experimental results for this experiment. Two schemes are compared. One is that the gallery image is rotated to 3 different poses (30, 60, 90 degrees) only, and the probe face is compared to one of these based on the pose angle information provided with the database. Another scheme is to generate 18 different poses with the yaw angles ranging from 5 to 90 degrees with a step size of 5 degrees. The matching is to compare the probe with each of the generated face images. From

Table 1. Face recognition accuracies when different schemes are used for head rotations: (1) from frontal view to other pose angles (i.e., 30, 60, 90 degrees); and (2) from other poses to the frontal view. The three columns are the query face pose angles. Only the LBP feature was used for this experiment.

	30°	60°	90°
LBP (others → front)	75.0%	59.3%	35.4%
LBP (front → others)	76.9%	71.1%	57.6%

Table 2. Face recognition accuracies when different schemes are used to rotate the head from frontal view to other angles: (1) generating all angles with an interval of 5 degrees; and (2) generating only limited pose angles. The three columns are the query face pose angles. Just the LBP feature was used.

	30°	60°	90°
LBP (only limited poses)	76.9%	71.1%	57.6%
LBP (dense poses)	92.3%	73.0%	65.3%

Table 2, we can see that the latter is much better than the former scheme.

How about the new face representation. In our last experiment, we investigate how well our new face representation can perform for face recognition. The usefulness of the covariance descriptor is validated carefully in the context of face recognition, compared with the LBP descriptor. The results are shown in Table 3. We can observe that the covariance descriptor gives better results than the LBP in the pose angles of 60 and 90 degrees. More importantly, the new representation based on the integration of the two features gives a higher accuracy in all poses of the query faces. This demonstrates that the two features can be complementary to each other, and the integrated representation can improve the face recognition accuracies significantly. It is interesting that the recognition accuracies at pose angles of 60 and 90 degrees are not reduced much compared to the 30 degrees.

4. Conclusion and Future Works

We have studied the problem of face recognition with large head pose changes, using the RGB-D face images. Different schemes have been investigated and compared to derive an approach for head rotations. The rendered face images via head rotations can deal with the face recognition problem effectively. A new face representation has been presented that combines the covariance descriptor with the LBP features. High accuracies can be achieved for face recognition with extremely large head pose changes. Our approach is efficient without the need of learning any face models. In future, we will evaluate our method on other RGB-D face databases when available. Another future work is to evaluate the new representation on depth images also that we have not done in this study.

Table 3. Face recognition accuracies using different features: (1) LBP, (2) covariance descriptor (COV), and (3) a new representation combining the two descriptors. The three columns are the query face pose angles. The integrated feature is significantly better than either feature, especially for large pose variations.

	30°	60°	90°
LBP	92.3%	73.0%	65.3%
COV	92.3%	76.9%	69.2%
New (COV+LBP)	**94.2%**	**84.6%**	**75.0%**

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *Eur. Conf. on Comput. Vision*, pages 469–481, 2004.

[2] A. B. Ashraf, S. Lucey, and T. Chen. Learning patch correspondences for improved viewpoint invariant face recognition. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[3] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606, 1992.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[5] M. T. Harandi, C. Sanderson, A. Wiliem, and B. C. Lovell. Kernel analysis over riemannian manifolds for visual recognition of actions, pedestrians and textures. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pages 433–439. IEEE, 2012.

[6] R. Hg, P. Jasek, C. Rofidal, K. Nasrollahi, T. Moeslund, and G. Tranchet. An rgb-d database using microsoft's kinect for windows for face detection. In *Eighth Int'l Conf. on Signal Image Technology and Internet Based Systems*, pages 42–46, 2012.

[7] T. Huynh, R. Min, and J.-L. Dugelay. An efficient lbp-based descriptor for facial depth images applied to gender recognition using rgb-d face data. In *ACCV Workshop on Computer Vision with Local Binary Pattern Variants*, 2012.

[8] B. Y. Li, A. S. Mian, W. Liu, and A. Krishna. Using kinect for face recognition under varying poses, expressions, illumination and disguise. In *IEEE Workshop on Applications of Computer Vision*, pages 186–192, 2013.

[9] R. Mattheij, E. Postma, Y. van den Hurk, and P. Spronck. Depth-based detection using haar-like features. In *Proc. of the BNAIC 2012 conference, Maastricht University, The Netherlands*, pages 162–169, 2012.

[10] R. Min, J. Choi, G. Medioni, and J.-L. Dugelay. Real-time 3d face identification from a depth camera. In *Int'l Conf. on Pattern Recognition*, pages 1739–1742, 2012.

[11] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs. Generalized multiview analysis: A discriminative latent space. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2160–2167, 2012.

[12] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.

[13] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *Europ. Conf. on Computer Vision*, pages 589–600. 2006.

[14] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1713–1727, 2008.

[15] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2879–2886, 2012.