Using Kinect for Face Recognition Under Varying Poses, Expressions, Illumination and Disguise

Billy Y.L. Li[1] Ajmal S. Mian[2] Wanquan Liu[1] Aneesh Krishna[1]

[1]Curtin University
Bentley, Western Australia
y.li2@postgrad.curtin.edu.au
{w.liu, a.krishna}@curtin.edu.au

[2]The University of Western Australia
Crawley, Western Australia
ajmal@csse.uwa.edu.au

Abstract

We present an algorithm that uses a low resolution 3D sensor for robust face recognition under challenging conditions. A preprocessing algorithm is proposed which exploits the facial symmetry at the 3D point cloud level to obtain a canonical frontal view, shape and texture, of the faces irrespective of their initial pose. This algorithm also fills holes and smooths the noisy depth data produced by the low resolution sensor. The canonical depth map and texture of a query face are then sparse approximated from separate dictionaries learned from training data. The texture is transformed from the RGB to Discriminant Color Space before sparse coding and the reconstruction errors from the two sparse coding steps are added for individual identities in the dictionary. The query face is assigned the identity with the smallest reconstruction error. Experiments are performed using a publicly available database containing over 5000 facial images (RGB-D) with varying poses, expressions, illumination and disguise, acquired using the Kinect sensor. Recognition rates are 96.7% for the RGB-D data and 88.7% for the noisy depth data alone. Our results justify the feasibility of low resolution 3D sensors for robust face recognition.

1. Introduction

Face recognition has attracted significant research interest in the past two decades due to its broad applications in security and surveillance. Sometime, face recognition can be performed non-intrusively, without the user's knowledge or explicit cooperation. However, facial images captured in an uncontrolled environment can have varying *poses*, *facial expressions*, *illumination* and *disguise*. Since the type of variations are unknown for a given image, it becomes critical to design a face recognition algorithm that can handle all these factors simultaneously.

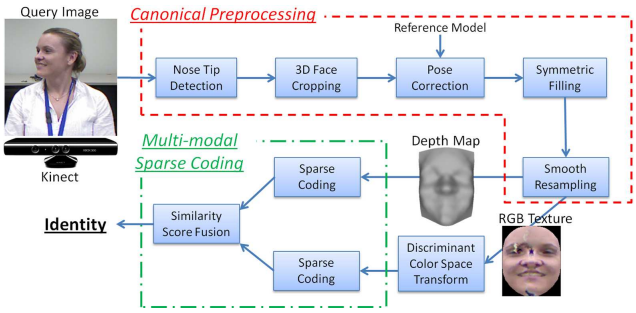

Figure 1. The proposed face recognition framework.

Simultaneously dealing with different variations is a challenging task for face recognition. Traditional approaches have tried to tackle one challenge at a time using 2D images i.e. texture. The illumination cone method [5] models *illumination* changes linearly. They have shown that the set of all images of a face under the same pose but different illuminations lies on a low dimensional convex cone which can be learned from a few training images. Although this technique can be used to generate facial images under novel illuminations, it assumes that faces are convex and requires training images to be taken with point light source. The Spare Representation Classifier (SRC) method [19] can handle face images with *disguise* (e.g. wearing sunglasses) by removing/correcting the outlier pixels. However, these pixels may have similar texture intensity to the human face and thus can not be identified. Some researchers have also tried to solve the *pose* problem using 2D images. For example, Gross *et al.* [6] construct the Eigen-light fields which are the 2D appearance models of a face from all viewpoints. This method requires many training images under different poses and dense correspondences between them which are difficult to achieve. Similarly, Sharma and Jacobs [16] use Partial Least Squares (PLS) to linearly map facial images in different poses to a common linear subspace where they

are highly correlated. However, such a linear subspace may not exist. In fact, pose variations are highly non-linear and can not be modeled by linear methods. This is why the performance of the above methods drops dramatically with extreme pose variations.

The most reliable way to address the *pose* problem is with 3D face models. Facial geometry is invariant to illumination and various imaging conditions whereas 2D images are a direct function of the lighting conditions (direction and spectrum). Although, the 3D imaging process can be influenced by lighting conditions, the 3D data itself is illumination invariant. Facial images under different illumination conditions can be generated using a 3D face model [17]. In addition, it can be used to correct the facial pose or to generate infinite novel poses. For example, the Iterative Closest Point (ICP) [1] algorithm finds the optimal transformation to minimize the closest point distance between two point clouds. Some approaches [9, 12] use the final ICP registration error for face recognition. However, this point-to-point error is sensitive to expression variations. To handle the *expression* problem, Bronstein *et al.* [3] proposed an expression-invariant representation of the facial surfaces based on isometric deformations. Mian *et al.* [12] proposed a multi-modal part-based method that utilizes texture information and focuses on the rigid parts of the face. Kakadiaris *et al.* [8] proposed the Annotated Face Model (AFM) to register the input 3D face to an expression-invariant deformable model. Recently, Passalis *et al.* [13] further extended the AFM method with facial symmetry to handle missing data caused by self-occlusion in non-frontal poses. A comprehensive survey on 3D face recognition methods is outside the scope of this paper and is given in [2]. Existing 3D face recognition methods are not designed to handle *disguise*. More importantly, they all assume the availability of high resolution 3D face scanners. Such scanners are costly, bulky in size and have slow acquisition speed which limit their applications.

Table 1. Comparison of 3D data acquisition devices.

Device	Speed (sec)	Charge Time	Size (inch^3)	Price (USD)	Acc. (mm)
3dMD	0.002	10 sec	N/A	>$50k	<0.2
Minolta	2.5	no	1408	>$50k	∼0.1
Artec Eva	0.063	no	160.8	>$20k	∼0.5
3D3 HDI R1	1.3	no	N/A	>$10k	>0.3
SwissRanger	0.02	no	17.53	>$5k	∼10
DAVID SLS	2.4	no	N/A	>$2k	∼0.5
Kinect	0.033	no	41.25	<$200	**∼1.5-50**

Some common 3D acquisition devices are compared in Table 1. High quality 3D scanners, for instance, the Minolta used in the well-known Face Recognition Grand Challenge (FRGC) [15], requires 2.5 seconds to capture a single 3D scan. Such a slow speed is not practical for scanning non-static faces. Requesting the subject to sit perfectly still for 2.5 seconds is not practical in many cases. High cost and slow acquisition time also make the collection of training data difficult. A single sample per person is usually not sufficient to represent all possible variations in the face [4], and many techniques such as Linear Discriminant Analysis (LDA) and Sparse Representation Classifier (SRC) [19] require many training samples. On the other hand, some 3D data acquisition devices trade data quality for high speed and low cost. For example, the Kinect sensor is extremely low cost, high in capture speed and compact in size. These properties are more appealing for user non-intrusive face recognition applications. On the downside, the 3D data provided by Kinect is very noisy and has low depth resolution. One can see from Figure 2 that, compared to the Minolta 3D scan, the Kinect 3D face model is hardly recognizable as human face and most of the popular face landmarks such as eye or mouth corners are not precisely locatable even manually. In this paper, we look into the feasibility of using the Kinect depth data for face recognition under varying pose, expressions, illumination and disguise.

Figure 2. Sample texture and 3D faces acquired with Minolta and Kinect. 3D faces are rendered as smooth shaded surfaces.

We propose a face recognition algorithm designed specifically for low resolution 3D sensors. It consists of novel preprocessing steps for estimating a canonical frontal view from non-frontal views. Our algorithm requires only the nose tip position. An efficient Iterative Closest Point (ICP) method and facial symmetry are used to canonicalize non-frontal faces. A multi-modal sparse coding based approach is proposed to utilize Kinect color texture and depth information (RGB-D). Ultimately, we can recognize faces under different poses, expressions, illumination and disguise using a single algorithm, which is compact and scalable. The proposed system is evaluated on a publicly available dataset namely CurtinFaces, which contains over 5000 samples of 52 subjects captured using the Kinect sensor. High recognition rates are achieved under challenging experiments. These results justify the feasibility of performing non-intrusive face recognition using a low-cost sensor.

To the best of our knowledge, the proposed algorithm is the first to utilize low quality 3D data from a consumer level sensor for addressing the challenging pose invariant face recognition problem. In addition, the idea of harnessing facial symmetry to estimate missing data and subsequently correcting the estimation error with sparse coding is an in-

novative combination.

2. Robust Face Recognition using Kinect

Figure 1 shows a block diagram of the proposed algorithm. Details of each component are given below.

2.1. Canonical Preprocessing

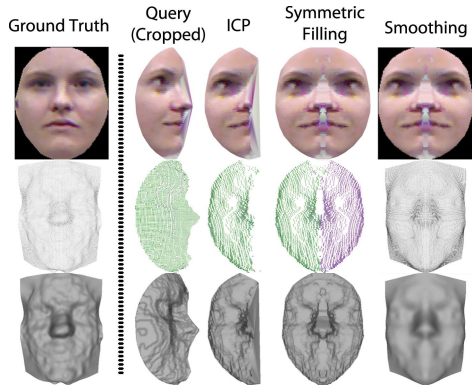

Figure 3. An example of canonical preprocessing on profile view.

Given a 6D point cloud (XYZ-RGB), the proposed preprocessing algorithm canonicalizes the face model and produces a depth map and RGB texture image. Unlike common range data preprocessing which only aims to remove holes and spikes, the proposed algorithm also aims to achieve view-point invariant representation. In fact, all the data we obtained from the Kinect sensor do not have spikes[1]. Holes are filled during the resampling step. An example is shown in Figure 3 and each preprocessing step is detailed below.

2.1.1 Nose Tip Detection

Due to the level of noise in Kinect depth data (as illustrated in Figure 2), the nose tip is the most reliable landmark that can be located on such 3D face model. Some approaches [12, 14] can detect the nose tip on a 3D face under different expressions and poses. However, the 3D data is high resolution in those cases. In this work, we assume that the approximate nose tip location has been detected. Since the nose tip is required only for face cropping and rough alignment, the system can work as long as the detected point is close enough to the true location.

2.1.2 3D Face Cropping and Pose Correction

Given the nose tip position, face cropping can be easily done in 3D. Following [12], we use a sphere of 8cm radius to crop the face. Specifically, we first translate the point cloud such that the nose tip is at the origin. Then points that are more than 8cm away from the origin are removed. As a result, a 6D point cloud (XYZ-RGB) of only the face surface area are obtained.

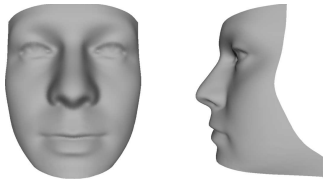

Figure 4. The reference face model.

The Iterative Closest Point (ICP) algorithm [1] is an accurate technique for alignment. However, it is known to be computationally expensive, and hence registering the query face to every frontal gallery face in search of the best alignment is not feasible. Instead, we register the query (XYZ only) to a reference model. Since different subjects have different face shape, the reference face model must be a reliable representation of common 3D faces. Such a reference face can not be constructed from the Kinect 3D data due to its noise level. Therefore, we build the reference face using face models (with no expression) from the FRGC [15] and the UWA database [11]. The reference face is constructed by aligning the scans, resampling them on a uniform grid and then taking their mean. The reference face has 64 points between the centers of the eyes. The number of points from the center of the lip to the line joining the eyes is also 64. The complete face has 128x128 points. Figure 4 shows the reference face used in our experiments. All faces including the training data and query face are registered to this reference face with six ICP iterations.

2.1.3 Symmetric Filling

Although the left and right regions of human face are not perfectly symmetric, the variations caused by facial asymmetry are less than the variations caused by different identities [13]. Unlike the work in [13] which mirrors the AFM external forces from one side to the other and then generates two different fitted AFMs for recognition, we utilize facial symmetry in the preprocessing stage at the point cloud level.

After pose correction, we create a mirrored point cloud by replacing the X values in the original point cloud by their opposite numbers (-X). However, not all the mirrored points are useful because we only want to fill in the missing data. Ideally, no point should be added on a frontal face, while all points should be mirrored on a profile view. To this end, for each mirrored point, we compute its Euclidean distance using (XY values only) to the closest point in the original point cloud. If this distance is less than δ, the mirrored point is removed. The idea is to add the mirrored point only if there is no neighboring point at that location. Note that Z is not used when calculating the distance, because the difference in Z is usually caused by facial asymmetry rather than

[1] It is possible that filtering is done inside the Kinect hardware or API.

missing data. The remaining mirrored points are then combined with the original point cloud. A sample symmetric filling is illustrated in Figure 3.

The threshold δ can be chosen based on the spatial resolution of the sensor or the point cloud itself. In our experiments, it was user defined. Depending on the original sample density, high values of δ will lead to a noisy surfaces while values too low will not benefit from symmetric filling. We empirically found that a good balance can be achieved with $\delta = 2$mm, however the performance is not affected much when setting δ to values between 1-5mm.

2.1.4 Smooth Re-sampling

There are three main objectives of re-sampling. First, it smooths out the noisy surface generated by the Kinect sensor and symmetric filling. Second, it fills up holes that still remain after symmetric filling. Lastly, it reduces the effect of face misalignment on the 2D grid caused by ICP registration. To this end, we fit a smooth surface to the point cloud (XYZ) using a publicly available code[2]. This algorithm fits a surface to the points using approximation as opposed to interpolation. Surface fitting is performed using a smoothing (or stiffness) factor that does not allow the surface to bend abruptly thereby alleviating the effects of noise and outliers. For each face, 128×128 points are re-sampled uniformly from its minimum X and Y to the maximum X and Y values. The advantage of re-sampling from min to max is that it can align the face on a 2D grid. Notice that we do not smooth the RGB texture since it is not noisy and smoothing will only blur it. Instead, we just re-sample it to the same XY location with interpolation. After re-sampling, the X and Y grids are discarded and four 128×128 matrices of depth and RGB are obtained. These are down-sampled to 32×32 for further processing.

2.2. DCS Transform

Color information is proven to be useful especially in the absence of shape clue [23]. In other words, color can improve recognition robustness. Recent research shows that color can improve face recognition performance significantly [21, 22, 18, 20]. Color images are usually modeled using RGB space, which is a weak space for face recognition due to its high inter-component correlation [22]. Therefore, researchers have focused on seeking a better color space such as the Discriminat Color Space (DCS) where faces are better separated [21][18, 20]. DCS finds a set of linear combinations for the R, G and B components in order to maximize class separability similar to the idea of LDA. We use the original DCS method [21] which is effective and reliable. We apply the DCS transform on the RGB texture image after preprocessing.

[2] mathworks.com/matlabcentral/fileexchange/8998

2.3. Multi-modal Sparse Coding

The Sparse Representation Classifier (SRC) is shown to be robust against disguise [19]. More importantly, it can correct small portion of errors or missing data. Our proposed canonical preprocessing algorithm usually results in small errors due to the fact that the human face is not perfectly symmetric and because sometimes less than half the face is visible in a profile view. Some data is completely missing when the profile view is slightly larger than 90 degrees in which case there are no reference points for mirroring. See Figure 3 as an example, which shows an error line in the middle of the resulting canonical face image. This kind of error can be effectively corrected by SRC.

In our proposed framework, we employ a multi-modal SRC algorithm for face recognition. Specifically, SRC is applied on the preprocessed depth map and the DCS color texture separately. Since DCS texture consists of three channels, they are first stacked into one augmented vector before SRC can be applied. Following the classification strategy in [19], two set of similarity scores are obtained based on individual class reconstruction error for the depth and texture. These two scores are normalized using the z-score technique [7] and summed for final decision. The query is assigned the label of the class with the highest similarity score.

In this paper, we formulate the sparse coding as the LASSO problem with ℓ_1 penalty:

$$\min_x \|Ax - y\|_2 + \lambda \|x\|_1 \qquad (1)$$

where A is the dictionary i.e. the training samples, y is the query face, x is the coding parameters vector and λ is a constant that controls the coding sparsity. Through out this paper, we set λ equal to 0.05 and Eq. (1) is solved using the SPAMS package [10].

3. Experiment

In this section, the CurtinFaces dataset is introduced, several experimental results are reported.

3.1. CurtinFaces Dataset

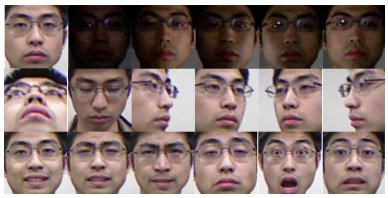

Figure 5. Sample un-preprocessed training images of a subject.

We use the online CurtinFaces database[3], which is publicly available, in our experiments. This dataset contains

[3] impca.curtin.edu.au/downloads/datasets.cfm

over 5000 images of 52 subjects. We use a subset which consists of 4784 images of 52 individuals with variations in poses(P), illumination(I), facial expressions(E) and sunglasses disguise. The database contains facial images with and without glasses. However, the first 3 images per subject in frontal pose, left and right profile view are without glasses. Additionally, for each subject, there are 49 images at $7P \times 7E$ and 35 images at $5I \times 7E$. Images with sunglasses are under five conditions (i.e. $3P$ and $2L$). The full set for each person consists of 92 images.

For training, we select 18 images per subject (see Figure 5). Each training image contains only one of the three variations (illumination, pose and expression). These images are used to compute DCS projection as well as the coding dictionary. Note that these images are also preprocessed with the proposed algorithm prior to use. Note that the use of multiple training images is practically feasible in the case of Kinect as it can acquire RGBD data instantly at 30 frames per second.

3.2. Pose and Expression Robust Face Recognition

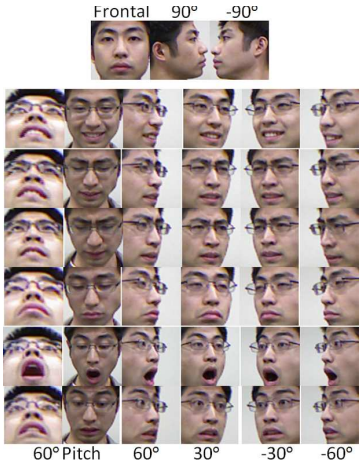

Figure 6. Sample un-preprocessed test images with simultaneous variation in *pose* and *expression*.

Table 2. Recognition rates (%) for *poses* $\times$ *expression* variations. Note that the results for D (Depth map) and T (Texture map) are obtained after preprocessing with the help of the 3D data.

	Without Symmetric			With Symmetric		
Pose	D	T	Fusion	D	T	Fusion
Frontal	100	100	100	100	100	100
$\pm 30^o yaw$	49.5	98.1	93.6	88.3	99.8	99.4
$\pm 60^o yaw$	14.9	80.4	55.1	87.0	97.4	98.2
$\pm 90^o yaw$	1.0	**39.4**	14.4	74.0	83.7	**84.6**
$\pm 60^o pitch$	77.2	**91.3**	90.9	81.6	**89.1**	92.8
Average	**46.2**	87.6	77.0	**85.4**	95.0	96.3

In this experiment, we evaluate the proposed system against pose and expression variations. There are 39 test images per subject as shown in Figure 6. Recognition is based on a single RGBD query image per subject. The recognition rate is reported in Table 2 for different poses. The most important result to be emphasized is the 84.6% recognition rate for the profile views which is attributed to the proposed symmetric preprocessing technique. One can see that when the symmetric filling step is excluded, only 39.4% recognition rate is achieved. Moreover, both depth and texture benefit greatly from the symmetric filling except for the texture image with pitch poses (which drops from 91.3% to 89.1%) where missing data can not be estimated by symmetry. In all cases, depth map benefits the most from symmetric filling. Recognition rate using depth data alone increases from 46.2% to 85.4% on the average. This is because, besides correcting the pose, symmetric data also helps in smoothing the noisy facial depth surface. Although fusion of preprocessed depth and texture only improves the performance slightly, the depth information is also contributing towards the pose correction and symmetric filling of the RGB data.

3.3. Illumination and Expression Robust Face Recognition

Figure 7. Sample un-preprocessed test images with simultaneous variation in *illumination* and *expression*.

Table 3. Recognition rates (%) for *illumination* $\times$ *expression* variations. Note that the results for D (Depth map) and T (Texture map) are obtained after preprocessing with the help of the 3D data.

	Without Symmetric			With Symmetric		
Illumination	D	T	Fusion	D	T	Fusion
Front	89.1	96.8	98.4	92.5	97.1	98.9
Back	89.4	96.6	97.6	93.8	96.5	98.6
Low Ambient	87.2	91.0	95.8	**91.3**	91.0	**97.1**
Average	**88.8**	95.6	97.6	**92.8**	95.6	98.4

In this experiment, we evaluate the proposed system against illumination and expression variations. There are 30 test images per subject as shown in Figure 7. The recognition rate is reported in Table 3. The most important result worth noticing is that the depth map benefit from symmetric filling in all cases (increase from 88.8% to 92.8%

on average), which is consistent with the pose experiment in section 3.2. However, all faces in this experiment are frontal. Therefore, the improved performance of depth map must be due to the fact that the Kinect noisy data has been smoothed by the symmetric filling. Moreover, fusing depth and texture data improves the performance significantly in this case especially in the case of limited light source (increases from 91.3% to 97.1%). This is because the texture information degrades dramatically when ambient lighting is too low, while depth information is unaffected.

3.4. Robustness to Disguise

Figure 8. Sample un-preprocessed test images with disguise.

Table 4. Recognition rate (%) under pose/illum. variation and sunglasses. Note that the results for D (Depth map) and T (Texture map) are obtained after preprocessing with the help of the 3D data.

	Without Symmetric			With Symmetric		
Condition	D	T	Fusion	D	T	Fusion
Frontal	90.4	34.6	96.2	**96.2**	34.6	94.2
Illumination	90.4	32.7	92.3	**94.2**	32.7	93.3
Pose	23.1	33.7	40.4	81.7	33.7	85.6
Average	63.5	33.5	72.3	89.6	33.5	**90.4**

In this experiment, we evaluate the proposed system against sunglasses disguise. There are 5 test images per subject as shown in Figure 8. The recognition rate is reported in Table 4. Notice that the depth information performs much better than texture (56.1% better on the average). This is because the surface of sunglasses is very different to the surface of human faces and can be easily identified as outliers. On the other hand, the texture of sunglasses (black in this case), is similar to the human eyes area especially when under strong shading caused by illumination. Therefore, fusing the texture information can decrease the performance. Lastly, one can see that the symmetric filling technique increases the performance of depth map in all cases.

3.5. Timing

Table 5. Recognition time in seconds for the complete test set and a single query image (average time).

	Whole Set	Single Query
Face Cropping	235	0.061
ICP Registration	13310	3.467
Symmetric Filling	3805	0.989
Resampling	1836	0.477
DCS Transform	4	0.001
Sparse Code (Depth)	100	0.026
Sparse Code (Texture)	323	0.084
Fusion	65	0.017
Total	19678	5.114

Table 5 reports the test time for our algorithm using a 64-bit Matlab implementation on an Intel Core2 Quad CPU @ 3GHz and 4GB RAM. No extra effort was made for code optimization except for sparse coding as mentioned in section 2.3. Most of the time is taken by the Matlab implementation of the ICP algorithm. This can be avoided if a C++ implementation is used. The current system can recognize a single query face in about 5 seconds irrespective of the pose, expression and illumination condition.

4. Comparison to Other Techniques

To the best of our knowledge, this is the first work reporting results for face recognition under pose, illumination, expression and disguise using the Kinect sensor. Therefore, performance can not be directly compared. However, in Table 6, we summarize some other reported performances in the literature that are related to non-intrusive face recognition. The proposed framework achieves 88.7% average recognition rate when using only the noisy depth data from Kinect under our challenging experimental setup of CurtinFaces. As we illustrated previously in Figure 2, the Kinect face without texture is too noisy to be recognizable by even a human. Therefore, a recognition rate of 88.7% using the Kinect depth data alone is a significant achievement. This result also suggests the usefulness of such data for face recognition. Our overall system achieves 96.7% recognition rate.

Table 6. Summary of reported performance in existing literature.

Method	Dataset (no. subject)	Conditions	Accuracy
This paper	CurtinFaces (52)	pose illumination expression sunglasses	88.7% (3D) 91.1% (2D) 96.7% (2D+3D)
UR3D-S [13]	UND+UHD (865)	pose expression	83.7% (3D)
PLS [16]	CMU-PIE (68)	pose	90.12% (2D)
Toderici *et al.*[17]	UHDB11 (23)	illumination	~92% (3D aided 2D)
Mian *et al.* [12]	FRGC (466)	expression aging	95.37% (2D+3D)
SRC [19]	AR (100)	illumination sunglasses	97.5% (2D)
Illumination Cone [5]	YaleB (10)	pose illumination	91.3% (2D)

There are also other approaches that tackle similar problems. For example, the UR3D-S [13] address the pose problem using facial symmetry. However, their approach is different from ours and they use high resolution 3D data. They reported an average recognition rate of 83.7% using depth data alone on a larger dataset. PLS [16] uses only 2D data for pose invariant face recognition and achieves 90.12% on a smaller dataset. Toderici *et al.* [17] proposed a 3D bi-directional re-lighting method that achieves about 92% under illumination variations. Mian *et al.* [12] proposed a 2D+3D part based approach that could achieve 95.37% on FRGC [15] with variation in expressions alone. The

SRC [19] method, using only 2D data, achieved 97.5% for sunglasses disguise on a relatively small dataset. The illumination Cone [5] approach achieves 91.3% when there are pose and illumination variations at the same time. Nevertheless, none of these existing techniques are designed to handle poses, expressions, illumination and disguise simultaneously.

5. Conclusion

We proposed a practical solution for robust face recognition using depth and texture information from a consumer level 3D sensor. We found that, facial symmetry can be used to aid face recognition under non-frontal view and it also helps in smoothing out noisy depth data. Although 3D data provided by consumer sensors like Kinect are very noisy, it is still useful for face recognition. Specifically, the 3D information can be used to preprocess the texture, improving face recognition accuracy significantly in situation of extreme pose variations. The preprocessed depth map also help face recognition, especially under low ambient lighting condition and sunglasses disguise. By utilizing RGB-D information, the proposed system performs face recognition with satisfactory accuracy even under simultaneous variations in pose, expression, illumination and disguise. These results suggest that non-intrusive face recognition can be performed well with high-speed low-cost 3D sensors, even though they have low depth resolution.

Acknowledgements

This work was partially supported by the Australian Research Council (ARC) Discovery Grant DP110102399.

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 14(2):239–256, 1992. 2, 3

[2] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101(1):1 – 15, 2006. 2

[3] A. Bronstein, M. Bronstein, and R. Kimmel. Expression-invariant representations of faces. *IEEE Trans. on Image Processing,*, 16(1):188–197, 2007. 2

[4] T. Faltemier, K. Bowyer, and P. Flynn. Using a multi-instance enrollment representation to improve 3d face recognition. In *IEEE Int'l Conf. on Biometrics: Theory, Applications, and Systems*, pages 1 –6, 2007. 2

[5] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 23(6):643–660, 2001. 1, 6, 7

[6] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 26(4):449–465, 2004. 1

[7] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005. 4

[8] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Trans. on Pattern Analysis and Machine Intel.*, 29(4):640 –649, 2007. 2

[9] X. Lu, D. Colbry, and A. Jain. Three-dimensional model based face recognition. In *IEEE Int'l Conf. on Pattern Recognition*, pages 362 – 366, 2004. 2

[10] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11:19–60, 2010. 4

[11] A. Mian. Illumination invariant recognition and 3d reconstruction of faces using desktop optics. *Opt. Express*, 19(8):7491–7506, Apr 2011. 3

[12] A. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 29(11):1927–1943, 2007. 2, 3, 6

[13] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3d face recognition. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 33(10):1938–1951, 2011. 2, 3, 6

[14] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris. Automatic 3d facial region retrieval from multi-pose facial datasets. In *Eurographics Workshop on 3D Object Retrieval*, pages 37–44, 2009. 3

[15] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Int'l Conf. on CVPR*, pages 947–954, 2005. 2, 3, 6

[16] A. Sharma and D. Jacobs. Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In *IEEE Int'l Conf. on CVPR*, pages 593–600, 2011. 1, 6

[17] G. Toderici, G. Passalis, S. Zafeiriou, G. Tzimiropoulos, M. Petrou, T. Theoharis, and I. Kakadiaris. Bidirectional relighting for 3d-aided 2d face recognition. In *IEEE Int'l Conf. on CVPR*, pages 2721–2728, 2010. 2, 6

[18] S.-J. Wang, J. Yang, N. Zhang, and C.-G. Zhou. Tensor discriminant color space for face recognition. *IEEE Trans. on Image Processing*, 20(9):2490 –2501, sept. 2011. 4

[19] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 31(2):210 –227, 2009. 1, 2, 4, 6, 7

[20] J. Yang and C. Liu. Color image discriminant models and algorithms for face recognition. *IEEE Trans. on Neural Networks*, 19(12):2088 –2098, dec. 2008. 4

[21] J. Yang, C. Liu, and J.-Y. Yang. What kind of color spaces is suitable for color face recognition? *Neurocomputing*, 73(10-12):2140–2146, 2010. 4

[22] J. Yang, C. Liu, and L. Zhang. Color space normalization: Enhancing the discriminating power of color spaces for face recognition. *Pattern Recognition*, 43(4):1454–1466, 2010. 4

[23] A. Yip and P. Sinha. Role of color in face recognition. *Journal of Vision*, 2(7), 2002. 4