How Much Information Kinect Facial Depth Data Can Reveal about Identity, Gender and Ethnicity?

Elhocine Boutellaa[a,b], Messaoud Bengherabi[a], Samy Ait-Aoudia[b], Abdenour Hadid[c]

[a]Centre de Développement des Technologies Avancées (DZ),
[b]Ecole Nationale Supèrieure d'Informatique (DZ),
[c]University of Oulu (FI)

Abstract. Human face images acquired using conventional 2D cameras may have inherent restrictions that hinder the inference of some specific information in the face. The low-cost depth sensors such as Microsoft Kinect introduced in late 2010 allow extracting directly 3D information, together with RGB color images. This provides new opportunities for computer vision and face analysis research. Although more accurate sensors for detailed facial image analysis are expected to be available soon (e.g. Kinect 2), this paper investigates the usefulness of the depth images provided by the current Microsoft Kinect sensors in different face analysis tasks. We conduct an in-depth study comparing the performance of the depth images provided by Microsoft Kinect sensors against RGB counterpart images in three face analysis tasks, namely identity, gender and ethnicity. Four local feature extraction methods are investigated for both face texture and shape description. Moreover, the two modalities (i.e. depth and RGB) are fused to gain insight into their complementarity. The experimental analysis conducted on two publicly available kinect face databases, EurecomKinect and Curtinfaces, yields into interesting results.

1 Introduction

Human face is involved in an impressive variety of different activities. It houses the majority of our sensory apparatus - eyes, ears, mouth, and nose - allowing the bearer to see, hear, taste, and smell. Apart from these biological functions, it also provides a number of signals about our health, emotional state, identity, age, gender etc. Machine analysis of faces (i.e. automatic face analysis) plays also a key role in many emerging applications of computer vision, including biometric recognition systems, human-computer interfaces, smart environments, visual surveillance, and content-based retrieval of images from multimedia databases. Due to its many potential applications, automatic face analysis which includes, e.g., face detection, face recognition, gender classification, age estimation and facial expression recognition, has become one of the most active topics in computer vision research [1].

Face analysis problems have been mainly extensively studied using conventional RGB cameras at visible light. However, this makes some face analysis tasks a challenging problem. Furthermore, face images acquired using such conventional sensors may have inherent restrictions that hinder the inference of some specific information in the face. For instance, illumination changes are still challenges in face recognition while near infrared imaging is shown to be less prone to this problem; face spoofing (e.g. detecting sign of liveness) is a threat in face recognition using RGB images while thermal cameras can easily solve this problem; analysing faces under pose variations from 2D images is a complex task which can be better handled in 3D. So, face sensing using new technologies and beyond the visible light is needed.

The recent introduction of low-cost depth cameras (such as Microsoft Kinect) provides exciting new opportunities for computer vision and face analysis research. Kinect sensors allow extracting directly depth information, together with RGB color images. This is a potential alternative to classical 3D scanners which are usually slow, expensive and large-sized, making them inconvenient for many practical applications. Consequently, low-cost depth sensing has recently attracted a significant attention in the research community [2,3].

Among the major drawbacks of the facial depth images provided by Microsoft Kinect are the low-resolution and noisy nature of the images. This can be due, for instance, to missing data (holes) in some parts of the face, inaccurate depth value computation and limited distance coverage from the sensor (2 to 4 meters). More accurate sensors (e.g. Kinect 2) for more detailed facial image analysis are expected to be available soon. Despite of the aforementioned limitations, Kinect depth images (after efficient pre-processing) have already been successfully used in some facial analysis tasks such as head pose estimation [4] and gender classification [5]. Of significant importance when dealing with Kinect depth images is also the use of effective face descriptions. Local features are usually shown to perform better than global features due to their ability to cope with local changes.

The intriguing question is how much information Kinect depth data can reveal about faces? To answer this question and to gain insights into the usefulness of the depth images in different face analysis tasks, this work provides the first comprehensive analysis comparing the performance of the depth images versus RGB counterparts in three face analysis tasks, namely identity, gender and ethnicity. Four local feature extraction methods are considered for encoding face texture and shape: Local Binary Patterns (LBP) [6], Local Phase Quantization (LPQ) [7], Histogram of Oriented Gradients (HOG) [8] and Binarized Statistical Image Features (BSIF) [9]. Moreover, the complementarity of the two sources of information (i.e. depth and RGB) is also studied through experiments fusing the two modalities. Extensive experiments are conducted on two recent publicly available benchmark databases namely EurecomKinect [5] and Curtinfaces [10] face databases. The obtained results point out interesting findings.

The remainder of this paper is organized as follows. Section 2 reviews some works related to the use of Kinect depth images. Then, Section 3 presents our methodology for studying the usefulness of Kinect depth images in different face analysis tasks. In Section 4, we describe the extensive experiments and discuss the obtained results. Section 5 draws some conclusions and highlights future perspectives.

2 Related work

It is well-known that illumination and pose variations can be better tackled using 3D scans of faces than 2D images. However, 3D scanners are usually expensive, bulky and slow, and this limits their use in practical applications. The recent introduction of low-cost depth cameras (such as Microsoft Kinect) provides exciting new opportunities for computer vision and face analysis research. Kinect sensor allows extracting directly depth information, together with RGB color images of the scene at video rates. The sensor is based on time-of-flight technology and is initially introduced as a peripheral of Microsoft Xbox games console. Since its introduction in late 2010, it is widely adopted by the computer vision research community in various applications [2].

Among the attempts to use Kinect sensors for face analysis is the work of Li et al. [10] who aimed at tackling the problem of face recognition under pose, illumination, expression and disguise using Kinect. The authors proposed a preprocessing chain that generates canonical frontal views for both depth map and texture of the face regardless of its initial position. To this end, Iterative Closest Point (ICP) is used for registering a given face to a reference model. Then, facial symmetry is employed to recover missing face parts, fill holes and smooth the face depth data. Finally, sparse representation classifier (SRC) is used for both depth and texture separately. Experimental results on the CurtinFaces dataset [10] yields in a recognition rate of 88.7% using depth data only and 96.7% when face texture and depth are fused.

Similarly, Goswami et al. [11] used images obtained from Kinect for face recognition. The proposed method computes the HOG descriptor on the entropy of RGB-D faces and the saliency features from a 2D face. The probe RGB-D descriptor is used as input to a random decision forest classifier to establish the identity. Experimental results on a private database comprising 106 subjects with multiple RGB-D images of each subject indicated that the RGB-D information obtained by Kinect can be used to enhance face recognition performance compared to 2D and 3D approaches.

In another work, Min el al. [12] explored the use of Kinect sensor for real-time 3D face identification. Instead of registering a probe to all instances in the database, the authors proposed to only register it with several intermediate references (i.e. canonical faces) randomly selected from the gallery, thus reducing the processing time without significantly affecting the recognition performance. Moreover, ICP was implemented on a GPU. Good identification results were reported on a dataset of 20 subjects with an average speed ranging from 0.04 seconds to 0.38 seconds, depending on the number of canonical faces. It is worth noting, however, that the proposed approach was tested only under limited variations of head pose, expression and illumination.

More recently, Pamplona Segundo et al. [13] addressed continuous face authentication problem using 3D faces acquired with Kinect. Faces are first detected and normalized using ICP. Each face is then registered according to its pose and classified into frontal, left profile or right profile. HOG features are then extracted from the region of interest and matched to the corresponding region. The approach was evaluated on four 40 minutes long videos with variations in facial expression, occlusion and pose. An equal error rate (EER) of 0.8% was reported.

Inspired by 3DLBP [14], (a variant of LBP based on the statistics of range image differences), Huynh et al. [5] proposed a novel descriptor, called Gradient-LBP, and applied it to the problem of gender classification from Kinect depth images. Gradient-LBP encodes the facial depth difference as well as its sign. The depth differences are computed from different orientations yielding in a separate depth difference image per each orientation. Hence, the depth difference at each pixel for all the orientations is encoded. Experiments were carried out on both high quality 3D range images (obtained by a 3D scanner) and images of lower quality obtained from Kinect (EURECOM Kinect Face Dataset). The reported results pointed out the usefulness of facial depth information when used together with RGB images for gender classification.

It appears that most of the few attempts on using Kinect in face analysis are mainly devoted to the face recognition problem hence overlooking and ignoring other face analysis tasks such as gender recognition, age estimation and ethnicity classification. Moreover, most of the proposed works focused on the fusion of Kinect depth information and RGB images but did not explicitly explore how much information Kinect facial depth data alone can reveal about the faces. Some of the results are also reported on size-limited and/or private Kinect databases. Finally, most of the existing works used only basic features with the depth data and RGB images.

To tackle these drawbacks, this present work provides the first comprehensive analysis comparing the performance of the Kinect depth images versus RGB counterparts in three different face analysis tasks (identity, gender and ethnicity) and using four local feature extraction methods (LBP [6], LPQ [7], BSIF [9] and HOG [8]). Extensive experiments are carried out on two recent publicly available Kinect benchmark databases (EurecomKinect [5] and Curtinfaces [10]).

3 Methodology

3.1 Preprocessing

Since depth images provided by Kinect sensor are usually noisy and of low quality, a preprocessing is needed and crucial before further analysis. The noise in the depth images can be originated from the unknown distance between the sensor and the face. The depth maps usually contain many holes that should be filled.

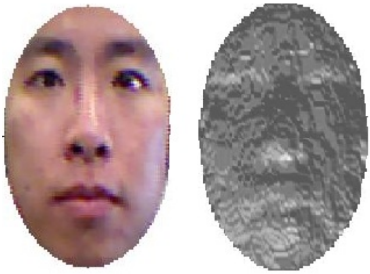


Fig. 1. Examples of 2D (left) and 3D (right) cropped face images obtained with the Microsoft Kinect sensor after preprocessing.

On the other side, 3D information is useful for assisting 2D analysis under severe pose variation by registration to a common face model using ICP algorithm.

Firstly, we transform the depth maps provided by Kinect into real world 3D coordinates. Thus, each pixel is represented by six values: x, y and z coordinates and the three RGB values. We translate the resulting cloud of points so that the nose tip is located at the origin by subtracting the nose coordinates. The nose has indeed been shown to be the most reliable point to crop the face region from the depth images [15]. Thus, we extract the face region using an ellipsoid centered at the nose tip. Therefore, the points located outside the ellipsoid are discarded. Then, we smooth and re-sample the face point cloud to a grid of 128×96. Examples of cropped 2D and 3D face images are shown in Fig. 1. Finally, for the 3D face, we drop the x and y coordinates hence keeping only the z coordinates describing the face shape.

3.2 Feature extraction

After preprocessing, facial descriptors are computed from the depth and RGB images. The function of the descriptors is to convert the pixel-level information into a form, which captures the most important facial properties but is insensitive to irrelevant aspects caused by e.g. blur, noise and illumination changes. In contrast to global face descriptors which compute features directly from the entire face image, local face descriptors representing the features in small local image patches have proved to be more effective in real world conditions. Hence, we adopted four state-of-the-art local descriptors which are briefly described below.

Local Binary Patterns (LBP) [6] is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. The discriminative power, computational simplicity and tolerance against monotonic gray-scale changes are behind the great success of LBP in many computer vision problems. In LBP, a pixel code is computed by thresholding its value with the neighborhood. The signs of the differences are coded as a binary string which is converted to a decimal number representing the pixel code. The occur-

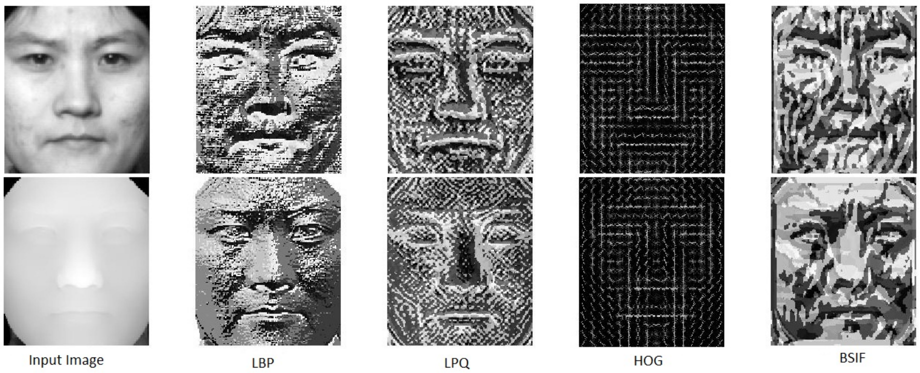


Fig. 2. Applying the four descriptors on a face texture and depth images. From left to right : the original face image (top: texture image and bottom: its corresponding depth image) and the resulting images after the application of LBP, LPQ, HOG and BSIF descriptors, respectively.

rences of the LBP codes in a given face image can be collected into a histogram. The classification can then be performed by computing histogram similarities.

Local Phase Quantization (LPQ) [7] is shown to be robust for texture analysis [16] and face recognition [7] from blurred images. In LPQ, an image is described using the phase information of short-term Fourier transform (STFT) locally computed on a rectangular window at each pixel. The phase information of four Fourier coefficients are coded by examining the signs of the real and imaginary parts of each component. For a given image, each pixel is labeled with a blur invariant LPQ code. Similarly to LBP, the occurrences of the LPQ codes are collected into a histogram for classification.

Histogram of Oriented Gradients (HOG) [8] describes local object appearance and shape within an image by the distribution of intensity gradients or edge directions. The magnitudes of the gradient at each pixel are accumulated into a histogram according to the gradient direction. The image is first divided into small connected regions from which histograms of gradient directions or edge orientations of the pixels are extracted. The combination of these histograms yields in the HOG descriptor. The method was initially developed for human detection but later extended and applied to other computer vision problems including face analysis.

Binarized Statistical Image Features (BSIF) approach [9] was recently proposed for face recognition and texture classification. Inspired by LBP and LPQ, the idea behind BSIF is to automatically learn a fixed set of filters from a small set of natural images, instead of using hand-crafted filters such as in LBP and LPQ. The set of filters are learnt based on statistics of training images [9]. The training images, which are normalized to zero mean and unit variance, are randomly sampled into small patches. The mean of each patch is subtracted and PCA is applied to reduce the dimension and whiten the data. Finally, the filters are estimated as the independent components obtained by ICA algorithm. An image is represented by the quantization of the filters responses.

Figure 2 shows the results of applying the four descriptors on a face texture and depth images of a subject from the FRGC database [17]. In the case of the BSIF descriptor, we extended the method to handle depth images as follows. We learnt the filters using facial depth images from the FRGC database. These filters are then used to compute BSIF features on Kinect depth images. We found this new learning to perform better than the original filters. To the best of our knowledge this is the first work that uses BSIF features for describing depth images.

For each descriptor (LBP, LPQ, HOG and BSIF), the RGB and depth images are first divided into several local regions from which local histograms are extracted and then concatenated into an enhanced feature histogram used for classification.

3.3 Classification

Once the face descriptors are extracted from both RGB and depth face images, we use the well-known support vector machine classifier (SVM) with an RBF kernel for the three face analysis problems. The libsvm[1] implementation is employed in our experiments. The face features are fed to the SVM classifier and the average classification accuracy is reported.

4 Experimental analysis

For extensive experimental evaluation, we analyzed the performance of the four local descriptors (LBP [6], LPQ [7], BSIF [9] and HOG [8]) presented in Section 3.2 on two publicly available Kinect face databases, EurecomKinect [5] and Curtinfaces [10], containing both RGB and depth facial images. We report the results in three different face classification problems: face identification, gender recognition and ethnicity classification. We describe below the experimental data, the setup and the obtained results.

4.1 Experimental Data

The EurecomKinect face database [5] contains both RGB and depth facial images of 52 subjects acquired using Kinect sensor. There are 14 females and 38 males in the database. The people in the database belong to six different ethnicity groups (Asian , Black, Hispanics, Indian, Middle East and White). The data is captured in two sessions separated by two weeks. In each session, the facial images of each person are captured under 9 different facial variations (neutral, smile, open mouth, strong light, eyes occlusion, mouth occlusion, paper occlusion, left profile and right profile. Face image samples from this database are shown in Fig. 3.

[1] `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

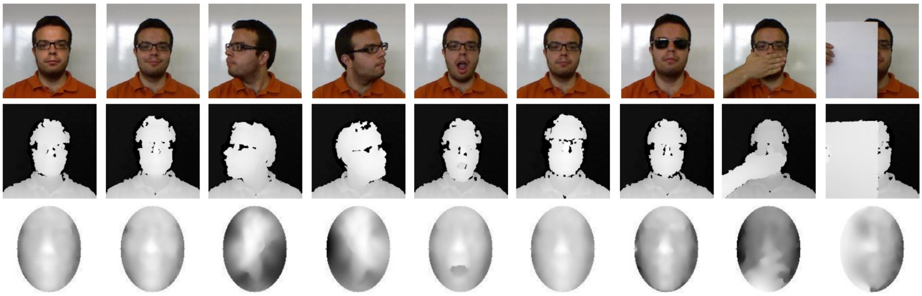


Fig. 3. Face samples from the EurecomKinect database. Top: RGB faces, middle: the corresponding raw depth maps and bottom: depth cropped face.

The CurtinFaces Kinect database [10] contains over 5000 images of 52 subjects in both RGB and depth maps obtained by Kinect sensor. The participants consist of 10 females and 42 males. Three ethnic groups (Caucasians, Chinese and Indians) are included. The facial images have various variations in pose, illumination, facial expression as well as sunglasses and hand disguise. The faces of each subject are provided with many combinations of these challenges. For each subject, there are 49 images under 7 poses and 7 facial expressions, 35 images under 5 illuminations and 7 expressions, and 5 images under disguise (sunglasses and hand). The full set for each person consists of 97 images. Face samples from this database are shown in Fig. 4.

4.2 Experimental Protocol and Setup

We considered experimental scenarios including face images under pose, illumination and expression variations. For fair evaluation, we divided each of the two databases into two subsets: development (*Dev*) and evaluation (*Eval*). In the case of EurecomKinect database, we used the images of the first session for training and those of the second session for tests. For CurtinFaces, we selected 18 and 69 images per person for training and test, respectively, so that both parts include pose, illumination and expression variations. In all the experiments, we tuned the optimal parameters of the methods on the development subset, and utilized these parameters to report the classification accuracy on the evaluation subset.

4.3 Experimental Results

Tables 1 and 2 summarize the classification performance of the four local descriptors on the two Kinect face databases for face identification, gender recognition and ethnicity classification. These results point out several findings:

– As expected, the overall classification rates indicate better performance on the EurecomKinect database (Table 1) compared to the CurtinFaces database

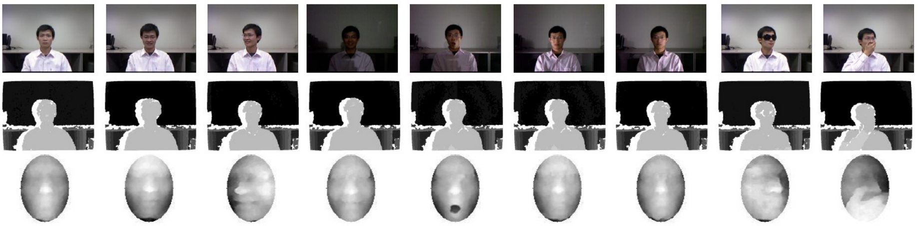


Fig. 4. Samples from the CurtinFaces database face images. Top: RGB faces, middle: their corresponding raw depth maps and bottom: depth cropped face.

(Table 2). CurtinFaces database is indeed more challenging in terms of variations of pose, expression and illumination.
- In overall, the RGB images yield in better performances compared to the depth images. Nevertheless, the results of the depth images alone are still good and actually much better than our expectations based on the human perception. It is indeed quite hard to visually distinguish the subjects using only the depth images.
- Regarding the best performing methods, the four different descriptors perform comparably on the RGB images under controlled conditions. On the depth images, HOG yields in the best classification rates under controlled environments followed by LBP and BSIF while LPQ seems to suffer the most. Under pose, expression and illumination variations, BSIF shows the highest robustness for both RGB and depth images followed by LPQ.
- A close look at the results in Table 1 and 2 indicates that gender classification is the least challenging task compared to face identification and ethnicity classification. This holds for both RGB and depth images and is in concordance with the findings of previous studies.
- The depth images provided by Kinect sensor are usually of low quality and noisy thus requiring a crucial preprocessing before analysis. The outcomes on such images are highly depending on the preprocessing step and hence cannot be easily generalized or compared to previously reported results if a different preprocessing is applied.

In another set of experiments, we analyzed the results of combining the RGB images with the depth information. We considered a simple feature level fusion strategy by concatenating the features extracted from RGB and depth images. As shown in figure 5, the obtained results indicated a slight but a clear performance improvement in all cases when combining the two modalities. This is also in agreement with the results of previous studies pointing out the usefulness of depth data when used together with RGB images [11].

5 Conclusion

We presented the first comprehensive study in the literature exploring the usefulness of the depth information acquired by the low-cost depth sensor, Microsoft

Table 1. Classification rates (%) using texture (RGB), depth for facial identity, gender and ethnicity classification on EurecomKinect database.

Method	Classification Rates (%)					
	Identity		Gender		Ethnicity	
	RGB	Depth	RGB	Depth	RGB	Depth
LBP [6]	100	94.2	94.2	96.1	97.1	81.7
LPQ [7]	99.0	91.3	99.0	88.4	98.0	78.8
HOG [8]	100	95.1	98.0	95.1	97.1	85.5
BSIF [9]	98.0	92.3	96.1	93.2	98.0	81.7

Table 2. Classification rates (%) using texture (RGB) and depth for facial identity, gender and ethnicity classification on CurtinFaces database.

Method	Classification Rates (%)					
	Identity		Gender		Ethnicity	
	RGB	Depth	RGB	Depth	RGB	Depth
LBP [6]	85.6	76.7	93.2	90.0	78.0	76.5
LPQ [7]	89.6	82.6	94.2	92.7	83.5	79.8
HOG [8]	81.3	82.6	92.7	91.5	74.9	76.7
BSIF [9]	93.2	80.8	95.0	92.9	84.9	84.7

Kinect, in different face analysis tasks including face identification, gender recognition and ethnicity classification. We experimented with four state-of-the-art local face descriptors on two publicly available Kinect face databases.

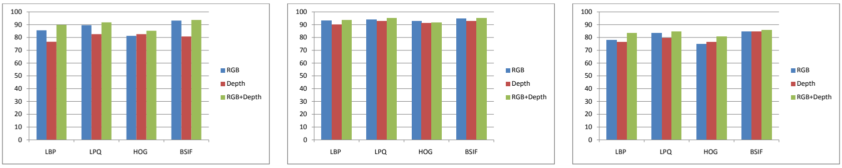
While it is difficult to visually distinguish the subjects using only the depth images, the obtained results showed that the depth information alone provides promising classification results beyond the expectations based on the human perception. This is a very interesting finding. With more accurate low-cost depth sensors for detailed facial image analysis which are expected to be available soon (e.g. Kinect 2), many face analysis problems will be much more feasible to solve.

The experiments also confirmed some findings of previous studies showing that (1) gender classification is the least challenging task compared to face identification and ethnicity classification, (2) combining the RGB images with the depth information does provide performance enhancement and (3) the performance of Kinect depth images highly depends on the preprocessing step which is a very crucial step before further analysis.

Regarding the best performing methods, the introduced BSIF features derived from a new set of filters provide promising results for both RGB and depth images under different variations of pose, expression and illumination. This should be further investigated especially with the expected Kinect 2.

As a future work, it is of interest to extend the work to other face analysis related tasks including age estimation and kinship verification combining RGB and depth facial information.

Finally, it is worth mentioning that the findings of our work should be further confirmed with larger Kinect databases and under more challenging settings in



(a) Identity (b) Gender (c) Ethnicity

Fig. 5. Results (%) for (a) identity , (b) gender and (c) ethnicity classification using LBP, LPQ, HOG and BSIF methods on texture (RGB) and depth images and their feature level fusion (RGB+Depth) on CurtinFaces dataset. The results indicate a slight but a clear performance improvement in most cases when combining the two modalities, hence pointing out the usefulness of depth data when used together with RGB images.

terms of illumination and pose variations. Toward this goal, we plan to record a large Kinect face database and make it publicly available to the research community along with well-defined evaluation protocol and baseline results in order to follow the progress on using low-cost depth data in face analysis.

Acknowledgements: This work is funded by the CDTA-FNR FUSMBIO project. Authors are thankful for DGRSDT and MESRS for their support. We would like also to thank the authors of EurecomKinect [5] and Curtinfaces [10] for making these Kinect face databases publicly available for research purposes.

References

1. Li, S.Z., Jain, A.K., eds.: Handbook of Face Recognition. Springer, 2nd edition, New York (2011)
2. Han, J., Shao, L., Xu, D., Shotton, J.: Enhanced computer vision with Microsoft Kinect sensor: A review. Cybernetics, IEEE Transactions on **43**(5) (Oct 2013) 1318–1334
3. Andersen, M., Jensen, T., Lisouski, P., Hansen, A., Gregersen, T., Ahrendt, P.: Kinect depth sensor evaluation for computer vision applications. Technical report, Department of Engineering, Aarhus University, Denmark (2012)
4. Fanelli, G., Weise, T., Gall, J., Gool, L.: Real time head pose estimation from consumer depth cameras. In: Pattern Recognition. Volume 6835., Springer Berlin Heidelberg (2011) 101–110
5. Huynh, T., Min, R., Dugelay, J.L.: An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In: Computer Vision - ACCV 2012 Workshops. Volume 7728. (2013) 133–145
6. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28**(12) (Dec 2006) 2037–2041
7. Ahonen, T., Rahtu, E., Ojansivu, V., Heikkila, J.: Recognition of blurred faces using local phase quantization. In: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. (Dec 2008) 1–4
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1. (June 2005) 886–893 vol. 1

9. Kannala, J., Rahtu, E.: BSIF: Binarized statistical image features. In: Pattern Recognition (ICPR), 2012 21st International Conference on. (2012) 1363–1366
10. Li, B., Mian, A., Liu, W., Krishna, A.: Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In: Applications of Computer Vision (WACV), 2013 IEEE Workshop on. (Jan 2013) 186–192
11. Goswami, G., Bharadwaj, S., Vatsa, M., Singh, R.: On RGB-D face recognition using kinect. In: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. (Sept 2013) 1–6
12. Min, R., Choi, J., Medioni, G., Dugelay, J.: Real-time 3D face identification from a depth camera. In: Pattern Recognition (ICPR), 2012 21st International Conference on. (Nov 2012) 1739–1742
13. Pamplona Segundo, M., Sarkar, S., Goldgof, D., Silva, L., Bellon, O.: Continuous 3D face authentication using rgb-d cameras. In: IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW). (2013) 64–69
14. Huang, Y., Wang, Y., Tan, T.: Combining statistics of geometrical and correlative features for 3D face recognition. In: British Machine Vision Conference, Proceedings of the. (September 2006) 879–888
15. Mian, A., Bennamoun, M., Owens, R.: An efficient multimodal 2D-3D hybrid approach to automatic face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on **29**(11) (Nov 2007) 1927–1943
16. Ojansivu, V., Heikkil, J.: Blur insensitive texture classification using local phase quantization. In: Image and Signal Processing. Volume 5099. (2008) 236–243
17. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Conference on. Volume 1. (2005) 947–954 vol. 1