An RGB-D Database Using Microsoft's Kinect for Windows for Face Detection

Idskou Høg, Rasmus; Jasek, Petr; Rofidal, Clement; Nasrollahi, Kamal; Moeslund, Thomas B.; Tranchet, Gabrielle

Published in:
IEEE 8th International Conference on Signal Image Technology & Internet Based Systems

DOI (link to publication from Publisher):
10.1109/SITIS.2012.17

Publication date:
2012

Document Version
Accepted author manuscript, peer reviewed version

An RGB-D Database Using Microsoft's Kinect for Windows for Face Detection

R.I. Hg, P. Jasek, C. Rofidal, K. Nasrollahi, T.B. Moeslund, and G. Tranchet
Laboratory of Visual Analysis of People
Aalborg University, Aalborg, Denmark
kn@create.aau.dk

***Abstract*—The very first step in many facial analysis systems is face detection. Though face detection has been studied for many years, there is not still a benchmark public database to be widely accepted among researchers for which both color and depth information are obtained by the same sensor. Most of the available 3d databases have already automatically or manually detected the face images and they are therefore mostly used for face recognition not detection. This paper purposes an RGB-D database containing 1581 images (and their depth counterparts) taken from 31 persons in 17 different poses and facial expressions using a Kinect device. The faces in the images are not extracted neither in the RGB images nor in the depth hereof, therefore they can be used for both detection and recognition. The proposed database has been used in a face detection algorithm which is based on the depth information of the images. The challenges and merits of the database have been highlighted through experimental results.**

***Keywords*-RGB-D face database; face detection; face recognition; Kinect;**

I. Introduction

Facial images are of great importance in many applications, like games (for facial expression analysis), human-computer interactions, access control (for face recognition), surveillance, etc. The first important step of any facial analysis systems is face detection. Though it has now been studied for many years both in 2D and in 3D ([1]-[8], to name just a few) and very good detection rates have been reported [1], face detection is still a challenging task due to problems like rotation, occlusion, illumination, overlapping, and noise. To overcome these problems different sources of information have been used like skin color, geometry of the face, 3D information of the face, etc. Among these, depth information has proven to be very useful for both face detection and also face recognition [9], [10], [11]. It is indeed shown in [10] that combining depth information with any 2D face recognition systems, improves the recognition rates at the end. Though there are many databases that can be used for 3D face recognition [9], there are not that many public databases which can be used for 3D face detection. This paper purposes such a database, for which a Kinect sensor which is capable of providing RGB-D data has been used. Kinect devices have been extremely popular recently, due to their low-cost and availability. The latest version of Kinect, which is known as Kinect for Windows has the possibility of providing depth information from a close range as 500mm to 3000mm. This makes these devices quite suitable for many applications including 2d and 3d facial analysis systems. Such a Kinect device has been used for generating the proposed database.

The rest of this paper is organized as follows: the description of the proposed database is given in the next section. Then, in section III a face detection algorithm which has been tested using this database is explained. Then, section IV presents the experimental results and finally section V concludes the paper.

II. The Proposed RGB-D Face Database

To generate the proposed database **RGB-D Face Database** first of all the following considerations were taken into account:

- The Kinect was set to work in *near mode*.
- Both the colour and depth images were taken exactly at the same time.
- The colour images were taken in resolution 1280x960 pixels.
- The depth images were taken in resolution 640x480 pixels.

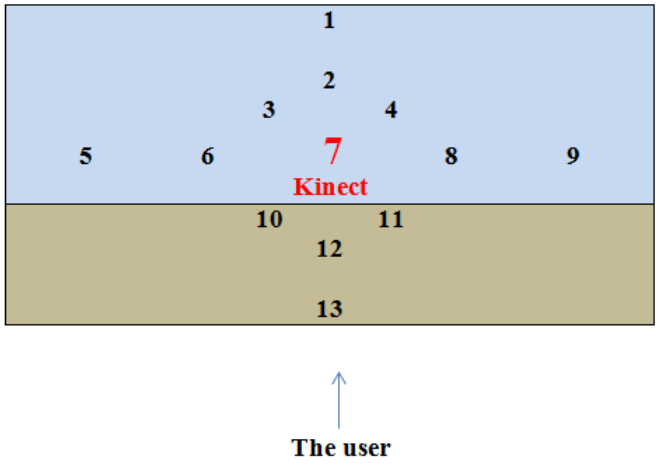

Figure 1. The setup of the system: higher part shows a wall and the lower area shows a table right in front of the user, these two are perpendicular to each other. The Kinect is placed on point number 7.

The scene was set to produce different face positions. Thirteen points on a wall behind the Kinect were chosen and

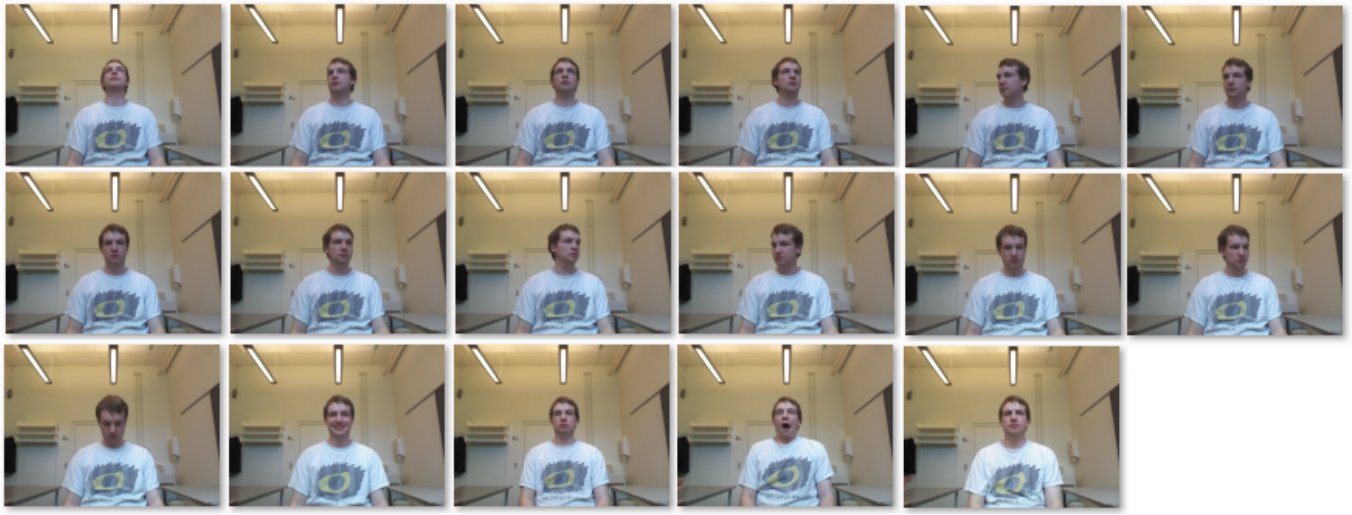

Figure 2. 17 images (as the ones shown above, and their depth counterparts) have been taken for each of the 31 test subjects in the database, and the process has been repeated three times for each test subject.

each person looked at these points sequentially to achieve roughly the same angles for each person. The position of points is shown in Figure 1. The positions 1, 5, 9 and 13 create an outer circle. The head of a person for these points was turned at angle of circa 55 degrees. The positions 2, 3, 4, 6, 8, 10, 11, 12 form an inner circle. For these points the head was turned approximately at angle of 30 degrees. For position 7 the head was facing the Kinect. Furthermore four different facial expressions were chosen. For each facial expression the person was facing the Kinect. The facial expressions were chosen such that they change the appearance of a face. The four expressions were: smile, sad, yawn and angry. Thus, altogether there were 17 face poses. This process has been repeated 3 times for each person resulting in $3\times17\times31$=1581 RGB images and 1581 depth images in the database.

The Kinect was placed on a table approximately in a height of the chest of the user. Differences in heights between persons were reduced by adjusting of the chair's height. The distance of the sitting person from the Kinect was approximately 85 cm. The lightening in the room was uncontrolled. Figure 2 shows the color sequence of one of the test subjects of the database.

A. *Storing of the data*

Every image both in colour and depth was stored as a single file. Color images were stored as standard bitmap files with 32 bit depth and in resolution of 1280x960 pixels. The depth images were stored as a text files where every pixel of a depth image is represented by its depth value. This depth value is in fact a distance from the Kinect in millimeters. Depth images were stored in resolution of 640x480 pixels. Values for the valid depth are in range from 400 to 3000 millimeters. Other values are constants specified by the Kinect:

- -1 for undefined depth
- 0 for too near depth and
- 4095 for too far depth.

The database can be downloaded at the home-page of laboratory of Visual Analysis of People at: http://www.vap.aau.dk/

III. Face Detection

The proposed database has been used for testing a face detection algorithm similar to the one proposed in [3]. Here the depth information is first used to find the closest object to the camera, which is a valid assumption for most facial analysis systems as it is usually the case that a user stands in front of the camera in these systems. Having reduced the search space of the input image, again depth information is used to find some face candidate regions. The depth image data may contain points or even areas with undefined depth. These areas are usually small but they can also be fairly large and can seriously affect the result of the face detection. Therefore, it should be first checked if there is any such holes in the user's face. If so, they need to be filled up. Here a mean filter of size 13×13 has been used for this purpose. Figure 3(b) shows the results of applying such a filter to the face image shown in Figure 3(a).

Having filled up the holes in the image, as mentioned before, the depth information are used to obtain face candidate regions. To do so, a curvature analysis technique similar to [3] has been used. In this method for every depth point two curvature measures are calculated as:

$$H(x,y) = \frac{(1+f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1+f_x^2)f_{yy}}{2(1+f_x^2+f_y^2)^{3/2}} \quad (1)$$

and

$$K(x,y) = \frac{f_{xx}f_{xy} - f_{xy}^2}{2(1+f_x^2+f_y^2)^2} \quad (2)$$

wherein f is the depth image, and f_x, f_y, f_{xy}, f_{x^2}, and f_{y^2} are its corresponding first and second order derivations at location (x, y). Having obtained these two curvature measures for each depth point the HK-classification method [3] is used to determine the type of the surface that this given point belongs to. The conditions given in Table I have been used for this purpose.

The type of surface that a depth belongs to is used to determine if it is a nose or eye. For this purpose, the two previously found curvature measures need to be thresholded. Following [3] a point will be considered to be a nose if $K > T_{K_{nose}}$ and $H > T_{H_{nose}}$ where $T_{K_{nose}} = 0.0325$ and $T_{H_{nose}} = 0.0925$. Similarly a point will be considered to be a mouth if $K > T_{K_{eye}}$ and $H > T_{H_{eye}}$ where $T_{K_{eye}} = 0.000275$ and $T_{H_{eye}} = -0.018$. Examination of the depth image using these thresholded measures reveals the possible points for nose and mouth on the depth image. It should be mentioned that a 4×4 neighborhood relationship will be used to connect pixels which belong to the same type of element: eye or mouth and at the end blocks representing possible eyes and mouths will be considered (Figure 3(c)).

After finding the possible candidates for the two facial elements the next step is to merge them into triangular candidate faces. Obviously such a triangle will include two eyes and one nose as its three vertexes. Though considering all the possible combinations may results in many candidate faces, many of them will be discarded very fast using the following simple rules:

- Both eyes have to be above the nose.
- The triangle should be approximately equilateral.

This results in very few candidates and ideally only one as it is shown in (Figure 3(d)). Regardless of the number of the found candidates, they will be all registered to a frontal face, and then their *faceness* will be measured using their depth information.

A. Depth face image registration

In order to assess the faceness of a possible candidate face image, it first needs to be frontal in such a way that the nose is in the center of image. It is done by applying a translation and two rotations which map the nose to the center of the image and maps the other pixels correspondingly. The first rotation matrix is constructed from unit vectors along the axes of the rotated space. The first vector u is found using the two eye points and is along the x-axis of the rotated space. Taking the cross product of two vectors in the face triangle, the u vector and the vector between the right eye point and the nose point, gives a vector w perpendicular to the face triangle and along the z-axis of the rotated space. The third vector v is found as the cross product of u and w and is along the y-axis of the rotated space. This results in the face triangle being parallel to the x, y plane and the eyes to have the same y position. A further 45 degree rotation around x is needed to make the face look in the direction of z and not downwards. To get an image of the point cloud from the new direction interpolation is used. Finally a mask is applied to the range image so only the rigid eye and nose part of the face is left. The image is now ready for validation. Figure 4 shows 15 examples of faces which have been registered using this method.

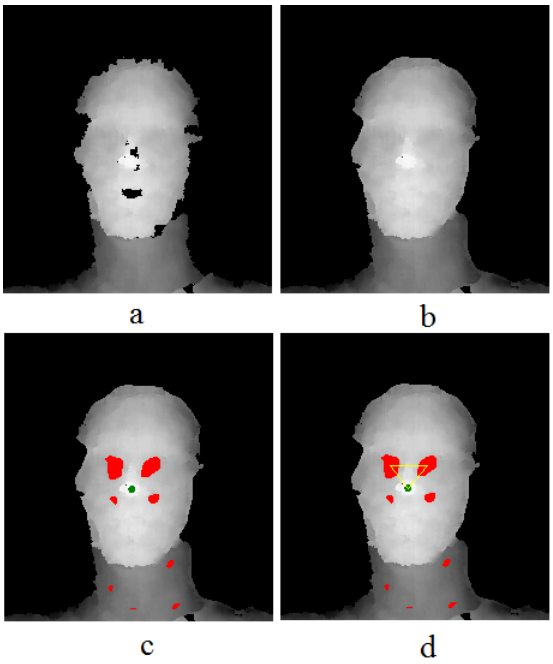

Figure 3. Finding the candidate faces: a) the initial estimate of the face region, b) filling up the holes, c) finding possible eyes and noses, and d) finding the most probable face triangle

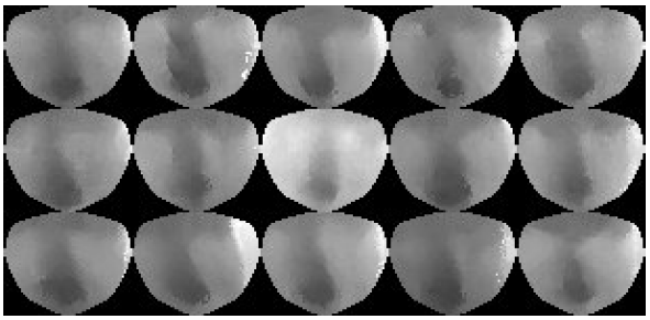

Figure 4. Some examples of registered depth face image

B. Face validation

To check if the registered depth images actually contain a face image, a Principal Component Analysis (PCA)-based face validation technique has been used similar to [3]. Here

	$K < 0$	$K = 0$	$K > 0$
$H < 0$	Hyperbolic concave	Cylindrical concave	Elliptical concave
$H = 0$	Hyperbolic symmetric	Planar	Impossible
$H > 0$	Hyperbolic convex	Cylindrical convex	Elliptical convex

Table I
SURFACE TYPE OF A DEPTH PIXEL BASED ON THE TWO MEASURED CURVATURE FEATURES [3].

in the training step, the depth images are used to build a face space. Then, in the testing step the test images will be projected into this face space and then they will be reprojected back to their original image space. If the reconstruction error is less than a predefined threshold, the depth image will be considered as a face image. It is shown in the experimental results that this method can efficiently detect faceness of depth images.

IV. EXPERIMENTAL RESULTS

The two main steps of the employed face detection algorithm have been tested using the proposed database and the results are given in the following subsections.

A. Triangle detection

For testing the face detection, first the average number of the detected triangles per image per pose has been studied. These are shown in Table II. It can be seen from the third column of this table, that for most of the poses the desired triangle has been found when the user is facing the Kinect. It can also be seen that face is detected more likely when person is looking left than right. This is probably due to the fact that the infrared light emitter is situated on the right side of the Kinect. It can be understood from the last column that there are usually some undesired triangles, which they often can be removed by the rules mentioned in Section III.

B. Testing PCA validation

For testing the second part of the face detection algorithm, PCA validation, three depth images of pose 7 of each person, that is 93 images, resulting in 93 face triangles at correct positions (face image samples) and 86 face triangles at wrong positions are used (non-face image samples). Then cross-validation will be used for testing the PCA validation. To do so, the order of each sample group face and non-face are randomized and then each group is split in half. The first part of each group is for training and the second for testing. First the PCA needs a face space. For that 15 images from the training part of the face group is randomly selected and used for generating the face space. Next the PCA validation is run on each sample of the training data giving the error value for each. The threshold is then decided by minimizing the total error, that is the number of false positives plus the number of false negatives. Using this decided threshold the PCA validation is run on the test data and the results are given in Table III.

a	b	c	d
All	2.181	91.208%	51.739%
Without outer circle	2.199	98.180%	59.576%
Only inner circle	2.180	97.372%	57.228%
Only outer circle	2.121	68.548%	36.028%
All facial expressions	2.419	100.000%	55.108%
#1	1.269	95.699%	83.871%
#2	1.892	89.925%	77.419%
#3	1.548	100.000%	77.419%
#4	1.559	96.774%	70.768%
#5	1.419	21.505%	6.452%
#6	2.161	97.849%	56.989%
#7	1.925	100.000%	59.140%
#8	1.968	100.000%	59.140%
#9	1.774	56.989%	36.559%
#10	2.892	83.871%	31.183%
#11	2.968	100.000%	41.935%
#12	2.271	98.925%	40.860%
#13	4.022	100.000%	17.204%
Smile	1.925	100.000%	62.336%
Sad	2.129	100.000%	53.763%
Yawn	2.677	100.000%	51.613%
Angry	2.237	100.000%	52.688%

Table II
FACE TRIANGLES DETECTION, A) POSES, B) AVERAGE TRIANGLES PER FACE, C) PERCENTAGE WHERE THE DESIRED TRIANGLE WAS FOUND, AND D) PERCENTAGE WHERE ONLY THE CORRECT/DESIRED FACE TRIANGLE WAS FOUND.

	a	b	c	d	e
Face	44	3	47	93.62 %	6.38 %
Non-Face	1	42	43	97.67 %	2.33 %

Table III
PCA POSITION 7 TEST RESULTS: A) DETECTED AS FACE, B) DETECTED AS NON-FACE, C) SAMPLES, D) DETECTION RATE, E) ERROR RATE

The results of the PCA validation for the other positions for the first 15 persons of the database is shown in Table IV.

For pose 7, that is the person looking straight at the Kinect, the correct face triangle is always among the found face triangles, see Table II. For this pose 93.62 % of the correct face triangles pass PCA validation. This means the system detects the face 93.62 % of time for pose 7. However, this detection rate drops for the other poses which makes sense, because the employed method for face detection in this paper is based on finding facial triangles, which is obviously not visible in rotated images where the user does not face Kinect.

	a	b	c	d	e
Position 1			54	74.07 %	25.93 %
Face	33	14	47	70.21 %	29.79 %
Non-Face	0	7	7	100 %	0 %
Position 2			65	49.23 %	50.77 %
Face	12	32	44	27.27 %	72.73 %
Non-Face	1	20	21	95.24 %	4.76 %
Position 3			65	86.15 %	13.85 %
Face	36	8	44	81.82 %	18.18 %
Non-Face	1	20	21	95.24 %	4.76 %
Position 4			78	78.21 %	21.79 %
Face	30	14	44	68.18 %	31.82 %
Non-Face	3	31	34	91.18 %	8.82 %
Position 5			62	83.87 %	16.13 %
Face	0	10	10	0 %	100 %
Non-Face	0	52	52	100 %	0 %
Position 6			109	62.39 %	37.61 %
Face	11	33	44	25 %	75 %
Non-Face	8	57	65	87.69 %	12.31 %
Position 8			104	87.5 %	12.5 %
Face	32	13	45	71.11 %	28.89 %
Non-Face	0	59	59	100 %	0 %
Position 9			104	69.23 %	30.77 %
Face	0	32	32	0 %	100 %
Non-Face	0	72	72	100 %	0 %
Position 10			146	78.77 %	21.23 %
Face	4	28	32	12.5 %	87.5 %
Non-Face	3	111	114	97.37 %	2.63 %
Position 11			157	96.18 %	3.82 %
Face	41	4	45	91.11 %	8.89 %
Non-Face	2	110	112	98.21 %	1.79 %
Position 12			140	76.43 %	23.57 %
Face	11	33	44	25 %	75 %
Non-Face	0	96	96	100 %	0 %
Position 13			196	83.16 %	16.84 %
Face	13	33	46	28.26 %	71.74 %
Non-Face	0	150	150	100 %	0 %
Position 14 (smiling)			102	93.14 %	6.86 %
Face	39	6	45	86.67 %	13.33 %
Non-Face	1	56	57	98.25 %	1.75 %
Position 15 (sad)			102	97.06 %	2.94 %
Face	45	2	47	95.74 %	4.26 %
Non-Face	1	54	55	98.18 %	1.82 %
Position 16 (yawn)			130	88.46 %	11.54 %
Face	36	11	47	76.6 %	23.4 %
Non-Face	4	79	83	95.18 %	4.82 %
Position 17 (angry)			113	92.92 %	7.08 %
Face	42	3	45	93.33 %	6.67 %
Non-Face	5	63	68	92.65 %	7.35 %
Total			1727	82.34 %	17.66 %
Face	385	276	661	58.25 %	41.75 %
Non-Face	29	1037	1066	97.28 %	2.72 %

Table IV
PCA TEST RESULTS FOR EACH POSITION EXCEPT POSITION 7. AT EVERY POSITION THE GENERATED FACE TRIANGLES FOR THE FIRST 15 PEOPLE IS USED FOR TESTING, A) DETECTED AS FACE, B) DETECTED AS NON-FACE, C) NUMBER OF SAMPLES, D) DETECTION RATE, AND E) ERROR RATE

V. CONCLUSION

This paper proposes a public RGB-D database for face detection, it means that the faces are not detected in the database neither in the color images nor in their depth counterparts. The database has been generated by the latest version of Kinect. An algorithm has been developed for face detection which uses the depth information of the images of the database and shows that the faces can be successfully detected (with a detection rate of 93.62%) as long as the users are facing the Kinect.

Handling face detection for rotated cases using color and/or depth information remains as an open question and the database will be freely available for this purpose.

REFERENCES

[1] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511-518, 2001.

[2] F. Tsalakanidou, S. Malassiotis, and M.G. Strintzis, "Face localization and authentication using color and depth images," IEEE Transactions on Image Processing, vol. 14, no. 2, pp. 152-168, 2005.

[3] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," Pattern Recognition, vol. 39, no. 3, 2006.

[4] A. Mian, M. Bennamoun, and R. Owens, "Automatic 3D face detection, normalization and recognition," in Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 735-742, 2006.

[5] R. Niese, A. Al-Hamadi, and B. Michaelis, "A novel method for 3D face detection and normalization," Jouranal of Multimedia, vol. 2, no. 5, pp. 1-12, 2007.

[6] H. Wu, K. Suzuki, T. Wada, and Q. Chen, "Accelerating face detection by using depth information," in Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology, pp. 657-667, 2009.

[7] W. Burgin, C. Pantofaru, W. D. Smart, "Using depth information to improve face detection," in Proceedings of the 6th international conference on Human-robot interaction, USA, 2011.

[8] S. Bodiroza, "Using image depth information for fast face detection," Towards Autonomous Robotic Systems, Lecture Notes in Computer Science, vol. 6856, pp 424-425, 2011.

[9] F. Tsalakanidou, D. Tzovaras, M.G. Strintzis, "Use of depth and colour Eigenfaces for face recognition," Pattern Recognition Letters, vol. 24, no. 910, pp. 1427-1435, 2003.

[10] K. Chang, K. Bowyer, and P. Flynn, "An evaluation of multimodal 2D + 3D face biometrics," IEEE Transactions on Pattern Analysis and Machine Intellegence, vol. 27, no. 4, pp. 619-624, 2004.

[11] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3D face recognition from depth and intensity Gabor features," Pattern Recognition, vol. 42. pp. 1895-1905, 2009.