June 2016

Implementation of face recognition in a mobile robot with RGBD images

Treball final del Grau en Enginyeria Electrònica Industrial i Automàtica

Universitat de Lleida

Maria Alba Ribó Palenzuela

Tutor: Jordi Palacín

1. Index

1.1. Index of Contents

1.2. Figure Index

1.3. Table Index

2. Objectives

The objective of this work is to contribute to developing a system for person identification through the analysis of an RGBD image of the face of the person in front of an Assistant Personal Robot (APR) (Figure 1).

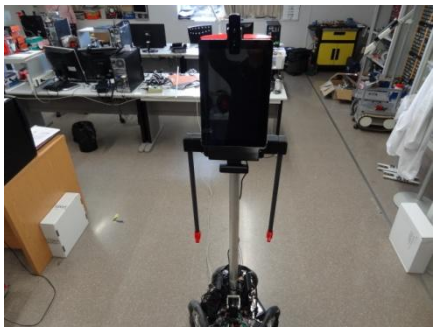



Figure 1. a) APR. b) Interaction with the APR.

To accomplish the purpose, the state-of-the-art works related to this subject will be studied, focusing on subject recognition but also on gender and age identification. Those works will be evaluated in advantages and disadvantages and finally an approach will be chosen to be implemented in this work.

The camera used in this work is the Creative Senz3D Camera which provides RGB, infrared and depth simultaneous information; the RGB image without background is also processed and obtained, furthermore the camera provides the world 3D points. In the final planned application, the APR will capture face images trough the 3D camera and they will be used to identify the person in front of the robot.

3. Related work

This section presents the related works on the topic of subject recognition. This works include facial recognition studies and facial databases.

3.1. Facial Recognition Studies

Here the state-of-the-art works on the topic of subject recognition found in the scientific literature are presented. The works are ordered chronologically.

3.1.1. A New Fast and Efficient HMM-Based Face Recognition System Using a 7-State HMM Along With SVD Coefficients

In [1] Miar-Naimi and Davari on 2008 proposed a HMM-based face recognition system using a 7-state HMM along with Singular Value Decomposition (SVD) coefficients. The seven face regions which were extracted using SVD and assigned to a state in HMM are: Hair, forehead, eyebrows, eyes, nose, mouth and chin as shown in Figure 2.

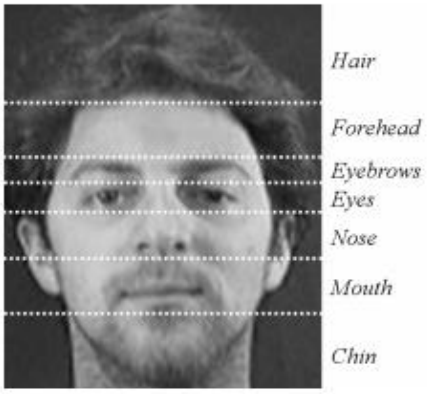


Figure 2. Seven Regions of Face.

The Olivetti Research Laboratory (ORL) Database [9] was the one they used to test their system.

3.1.1.1. Process

First the image is filtered using a 3x3 statistic filter, which replaced the centred element of the window with its minimum element as shown in Figure 3. This step reduces the information of the image.

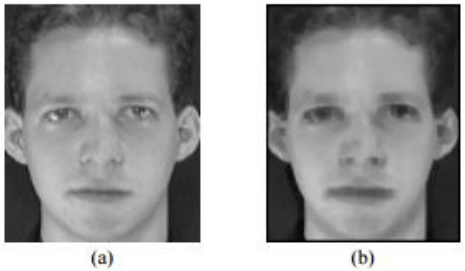


Figure 3. Example of operation of the order static filter. a) Before filtering. b) After filtering.

The original Black and White images from the ORL Database of resolution 112x92 are reduced to 64x64. This dataset contains 10 facial images of 40 different subjects.

The two-dimensional image resulting is reduced into a one-dimensional sequence of blocks of width 64 and height 5. The features are extracted computing SVD coefficients in each block. The features describing each block are the most significant output values of SVD: Σ_{11}, Σ_{22}, U_{11}. These values are later quantized so that they can be modelled by discrete HMM. This process is done by rounding, truncating, or similar irreversible, nonlinear process that lead to information loss. Once each block of images is mapped to a sequence of integers (considered the observation vectors), they are modelled by a seven-state HMM shown in Figure 4.

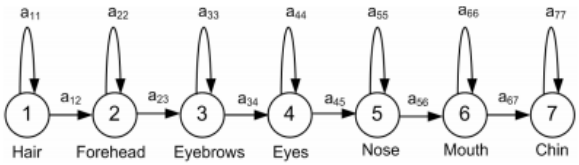


Figure 4. A one dimensional HMM model with seven states for a face image with seven regions.

From each class (person) 5 images are used for training and the remaining 5 for testing. The HMM is trained for each person in the database using the Baum-Welch algorithm. The test image to classify is first represented by its own observation vector to calculate the provability of it given each HMM face model.

3.1.1.2. Results

The recognition rate obtained by this system was 99% testing ORL Database resized and divided in training and testing datasets as explained above. The training time per image was 0.63 seconds and the recognition time per image 0.28s. If the image size was reduced to 32x32 the recognition time diminished to 0.15s but the recognition rate lowered also to 92.5%.

3.1.1.3. Advantages

- The blocks in which the images were divided were represented not with the grey values of each pixel but with just three features (SVD values). However those values come from the grey pixel values.
- The recognition time is 0.28s per image (at a resolution of 64x64 pixels).

3.1.1.4. Disadvantages

- Face detection, rotation or lighting adjustments were not implemented, as all where face images taken at the same distance and with few lighting variations. These processes would increase recognition time.
- The efficiency would lower if larger and more complicated databases were tested. Feature extraction or the use of 2-D HMM may improve the system performance.

3.1.2. Face Component Extraction Using Segmentation Method on Face Recognition System

In [2] Agushinta et al. determined the face components' regions and their characteristics as well as some of the distances between them to be used in face recognition. They used their own dataset of 150 frontal face RGB images of resolution 300 x 300 pixels in JPG format. The face components extraction process works as follows.

3.1.2.1. Process

First the face is detected, based on skin colour model. Luminance level is decreased converting it to YCbCr or chromatic colour, a Gaussian distribution is then applied and finally a binary image is obtained after a threshold process. Then the image is cropped obtaining the face part. These steps are shown in Figure 5.

Figure 5. Processing Steps: a) Original Image, b) YCbCr conversion, c) Result of Gaussian distribution, d) Binary image result of threshold process, e) Face detected, f) Face cropped.

Secondly the face region image is processed to obtain de face components and distances between them. This process is done on YCbCr image and is conducted by dividing the face as shown in Figure 6; extracting the eyes and mouth is done as shown in Figure 7 and Figure 8 respectively; finally the nose is obtained by computing the distance between eyes (Figure 9a) and mapping the lower region of the calculated whole nose region (Figure 9b).

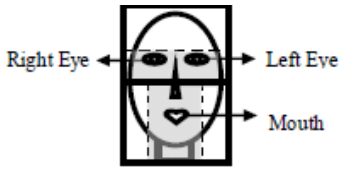


Figure 6. Face divisions

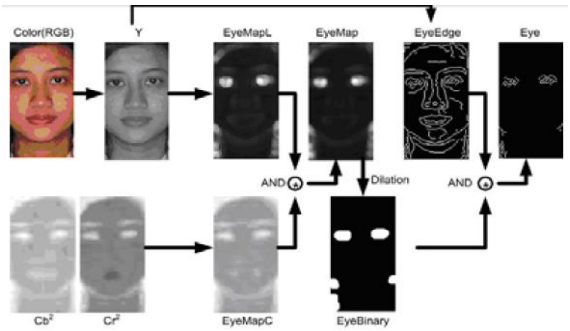


Figure 7. Eye map formation

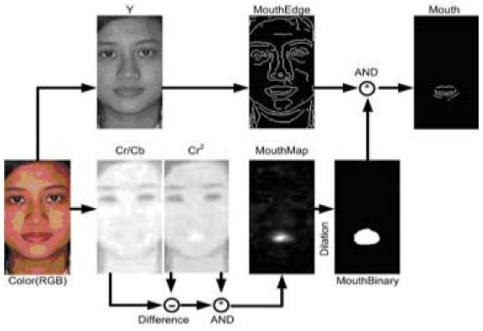


Figure 8. Mouth map formation

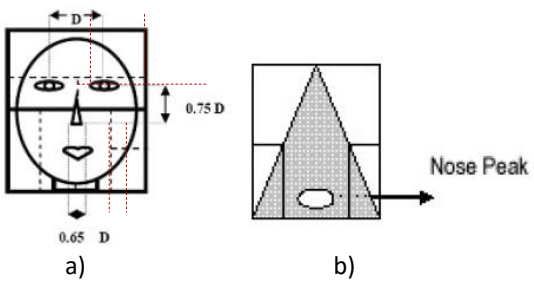


Figure 9. Nose component geometry

Finally the distances between face components are calculated between these points:

- The centre point of right eye box
- The centre point of left eye box
- The centre point of mouth box
- Nose peak
- End points of nose width

The distances calculated are the following, shown in Figure **¡Error! No se encuentra el origen de la referencia.**:

- J1: Right eye – Left eye
- J2: Right eye – Mouth
- J3: Left eye – Mouth
- J4: Right eye – Nose
- J5: Left eye – Nose
- J6: Nose – Mouth
- J7: Nose height
- J8: Nose width

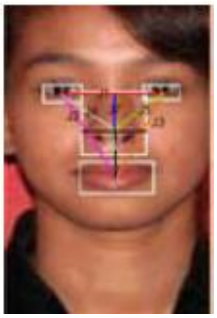


Figure 10. Determination of distances between face components.

3.1.2.2. Results

A study of the combination of those eight distances (J1-J8) that gave uniqueness was performed offering the results presented in Table 1.

Table 1. Face uniqueness level

Nº	Component Distance Combination	Unique Number	Unique (%)
1	J1-J2-J3-J4-J5-J6-J7-J8	150	100
2	J1-J2-J3-J4-J5-J6-J7	150	100
3	J1-J2-J3-J4-J5-J6	150	100
4	J1-J2-J3-J4-J5	148	98.67
5	J1-J2-J3-J4	148	98.67
6	J1-J2-J3-J5	140	93.33
7	J1-J2-J3-J6	150	100
8	J4-J5-J6	126	84
9	J1-J2-J3	140	93.33
10	J7-J8	42	28

Considering the 150 images of the database tested uniqueness was achieved with a minimum of four distances considered (case nº 7). The minimum number of distances should increase along with the number of images considered.

The uniqueness was tested using eigenspace method. The eigenvalues obtained indicated that J1 was the most important component. It also showed that faces are not always symmetrically, for what it is important to consider both J2 and J3 (one doesn't represent the other) and the same happens with J4 and J5.

3.1.2.3. Advantages

- The features used were not the colour or grey level of pictures (which may have substantial variation in changing lighting) but the distance between face parts.

3.1.2.4. Disadvantages

- Images of low resolution may lead to lower differences in joint distances and therefore finding uniqueness harder.
- Using skin detection for face identification could lead to errors when having images of not just the upper body.

3.1.2.5. Singularities

- Method different from Viola & Jones' to detect de faces was proposed:
  - Face detection using skin detection.
  - Face part detection form mapping YCbCr images and dividing face.

3.1.3. Optimizing Face Recognition using PCA

In [3] Abdullah et al. on 2012 improved the (Principal Component Analysis) PCA by decreasing the time of computation keeping the same performance. This implementation's purpose was to perform identification in a security system, recognizing or not the test subject as one in the training database. Two recognition mistakes may occur: false reject (FR) when the system rejects a known person and false acceptance (FA) when it accepts an unknown person.

In PCA face images are decomposed into small sets of characteristic features called eigenfaces, which will represent both train and test database. First the mean face is calculated and subtracted from all the faces calculating the deviation. From that the covariance matrix and the eigenfaces are obtained. In order to faster the process the most representative eigenfaces are calculated and will be the ones used for recognition, which will be done calculating the Euclidean distance between two face key vectors. The images used were part of the Faces94 database [13].

3.1.3.1. Process

First the minimum number of images per person (class) which offered the best recognition percentage was studied. 19 persons were put in the test set with one image for each in the

training set, increasing one person in the training dataset by step of one. 100% recognition was reached when using 6 images in the training database.

Secondly the threshold value of acceptance was studied, setting it to $2.50 \cdot 10^{16}$ (Maximum number of the Euclidean distance between the test and training image if the image is to be accepted). In this case de FA is nearly zero (Figure 10). The recognition was performed using PCA including 230 eigenvectors. Each recognition process took between 1.9s to 2s.

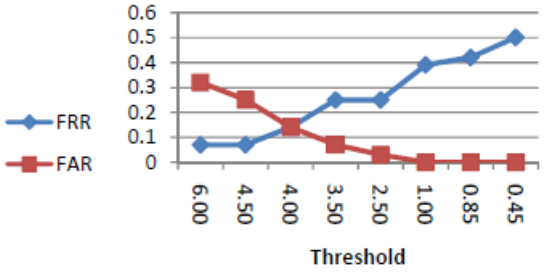


Figure 10. FA and FR rates for PCA

Thirdly the number of eigenvectors considered for recognition was diminished from 230 to 188 and the second experiment repeated. The time taken for each recognition was in this case between 1.4s to 1.5s, while keeping the same threshold value of $4.5 \cdot 10^{5}$ at zero FA rate (Figure 11).

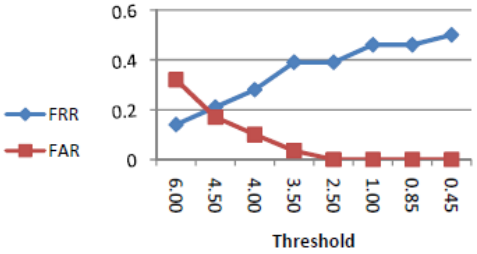


Figure 11. FA and FR rates for optimized PCA

3.1.3.2. Results

Decreasing the eigenvectors considered of the PCA algorithm from 230 to 188 led to a substantial decrease in computational time (from 2s to 1.5s) while keeping the same performance: FR rate below 0.2. Recognition speed of 1.5 seconds per image (at a resolution of 180 x 200 pixels).

3.1.3.3. Advantages

- Threshold value for recognition was calculated

3.1.3.4. Disadvantages

- Persons were not matched, but just considered to be from the database or not.
- The PCA used the grey pixels values as input information. (Which may have substantial variation in changing lighting)

3.1.4. An efficient 3D face recognition approach based on the fusion of novel local low-level features

In [4] Lei et al. on 2013 presented a novel 3D face recognition approach based on low-level geometric features collected from the eyes-forehead and nose regions. The particularity of those regions is that they are less influenced by the deformations caused by facial expressions. Two datasets were used: FRGC v2.0 (Face Recognition Grand Challenge database [10]) and BU-3DFE (Binghamton University 3D Facial Expression database [11]).

3.1.4.1. Process

Pre-processing:

First of all the FRGC v2.0 database is pre-processed, as some of the scans come with spikes and holes. To remove spikes, the values of all three coordinates (*x*, *y* and *z*) of the outlier vertices are smoothed according to the neighbouring vertices information. Then the whole surface is smoothed using a mean value filter. To fill the holes a bi-cubic interpolation is applied along the three coordinate matrices. Then the 3D facial scan is converted into 2D, where the *z* value is represented as the pixel value. The distances between the pixels along x and y directions are of 1mm.

Later, the pose is corrected to frontal view and the face is cropped using a mask which locates the nose tip, then PCA is applied on the cropped points to find their principal directions.

Feature extraction:

To represent the face three regions are cropped using a mask, see Figure 12:

- Nose (rigid region)
- Eyes-forehead (semi-rigid region)
- Mouth (non-rigid region)

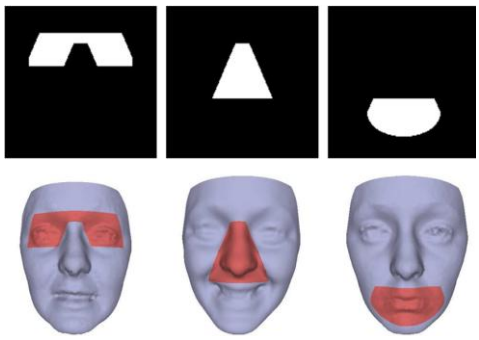


Figure 12. Region based 3D facial representation. 1st row: binary masks which are used to detect the semi-rigid, rigid and non-rigid regions of a face respectively. The 2nd row illustrates some extracted regions taken from the BU-3DFE dataset.

To diminish the information the cropped regions are sampled at uniform (x, y) intervals of 2mm. And only the seed points are kept. The Mouth region is discarded due to the high variance in facial expressions.

Next, the 3D geometric features are extracted. To do so the 2D images from the regions are converted into 3D point clouds. Each of the two facial regions is represented using multiple spatial triangles, where one vertex is the nose tip and the other two are randomly picked from the corresponding surface region (Figure 13a). From the triangles four features are proposed (Figure 13b):

- A: Corresponds to the angle between the two lines determined by the two random vertices and the nose tip.
- C: is defined as the radius of the circumscribed circle to the triangle determined by the two random vertices and the nose tip
- D: is defined as the distance of the line between the two random vertices
- N: is defined as the angle between the line defined by the two random vertices and the z-axis.

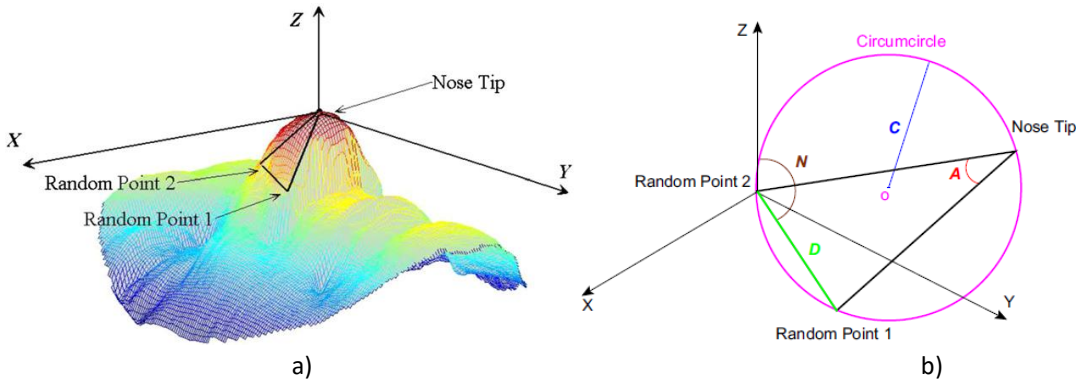


Figure 13. a) Example of one of the triangles used for the extraction of the low-level geometric features. b) Representation of the four low-level geometric features.

When the feature values are calculated from all the possible triangles of a certain region, they are located into the histogram representing that particular feature and region. This process will reduce the amount of data to analyse, as instead of having the information for all the triangles, this it is reduced to the information of the m bins of the histogram. The number of bins is chosen to be 180 for all the histograms, as it is found to be the optimal one which is sensitive to the subject change but not sensitive to noise or facial expression change. Finally the histograms of the same region are concatenated. In consequence, a face is represented by a rigid region histogram descriptor and a semi-rigid region histogram descriptor, each of dimensions 180x4.

In Figure 14 a representation of the histograms of two individuals performing four different facial expressions are shown. In each graph eight different histograms are plot. The subjects are differenced by colour. It clearly demonstrates that the histograms which correspond to the same individual cluster together.

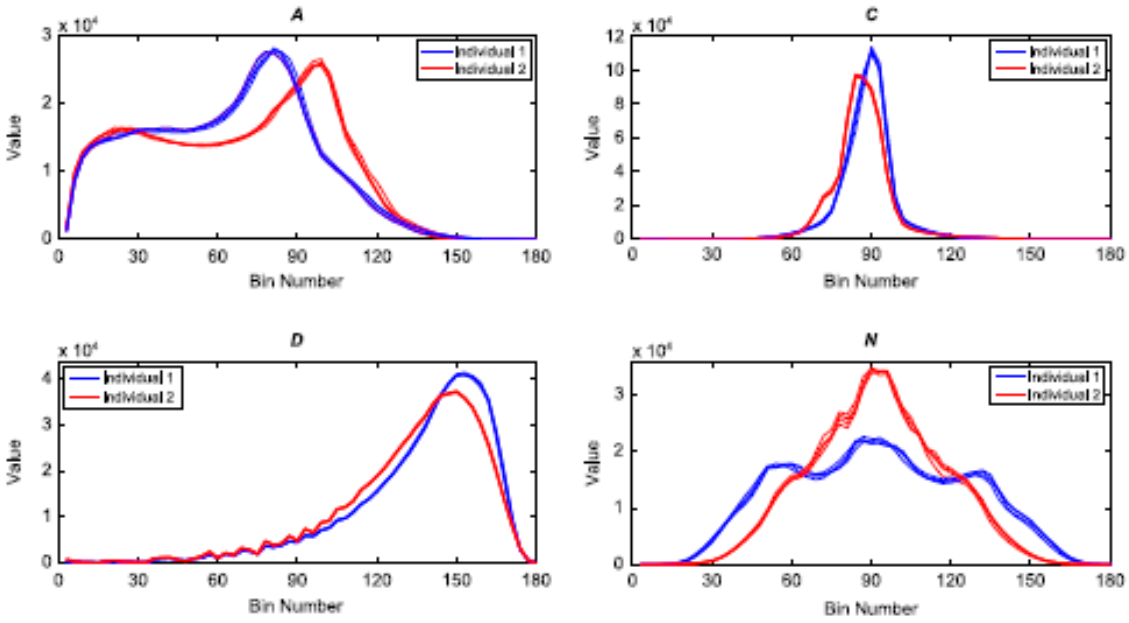


Figure 14. Histogram representation of the four features: A, C, D and N.

Classification:

In this work, SVM is adopted for classification for its excellent discriminating performance. The basic idea of SVM is to map the feature vectors into a higher-dimensional space and then to find an optimal hyper-plane to separate one cluster from another by calculating the maximal margin.

SVM is a maximal margin classifier which performs classification by finding an optimal hyper-plane that maximizes the distance to the closest points in a higher-dimensional space. It was first designed to solve binary classification problems. To use it in this multiclass classification problem the one-vs-all method is used, in which *k* SVMs are trained to classify *k* classes, and each SVM is responsible for the separation of the samples of one class from all the other samples of the other classes of the training set. In this work non-linear Gaussian radial basis function (RBF) kernel is the one used to map each data into a higher-dimensional space.

To perform the classification two histograms from each subject are available: the semi-rigid region histogram and the rigid region histogram. To considerate both histograms two methods are studied:

- Feature-level fusion: Concatenate both histograms, having one set of SVM
- Score-level fusion: Perform the classifications of the two sets of SVM (for both histogram) and fuse the results to obtain the final classification.

The feature level fusion is found to be the most effective one, as it gives the maximum classification rate.

3.1.4.2. Results

The optimal regions to consider are found to be the semi-rigid (eyes-forehead) and rigid (nose).

The optimal number of bins of each histogram is found to be 180.

The highest recognition rate is obtained considering the four features. D is found to be the most discriminative, followed closely by N. C is the least discriminative.

The best method to fuse both semi-rigid and rigid histograms is found to be the feature-level fusion, concatenating both histograms and using one set of SVM.

The classification rates on the two datasets tested (FRGC v2.0 containing 466 individuals and BU-3DFE, containing 100 individuals) range from 95.6% to 98.7% using the feature-level fusion method.

3.1.4.3. Advantages

- The input to the SVM is not the pixels colour information (which may have substantial variation in changing lighting) but the histograms representing the features' values obtained from depth data.
- Large databases tested with high recognition rates
- Robust to expression change.

3.1.4.4. Disadvantages

- No information about the time required for training or testing

3.1.4.5. Singularities

- Local features obtained from depth images are used.

3.1.5. Age Group Estimation using Face Features

In [5] Jana et al. on 2013 estimated the age group of humans in images of their own database using face features.

3.1.5.1. Process

First the images are pre-processed (Figure 15). They are cropped at the face area using the MATLAB inbuilt function of Viola & Jones, the eye-pair, mouth, nose and chin are also detected.

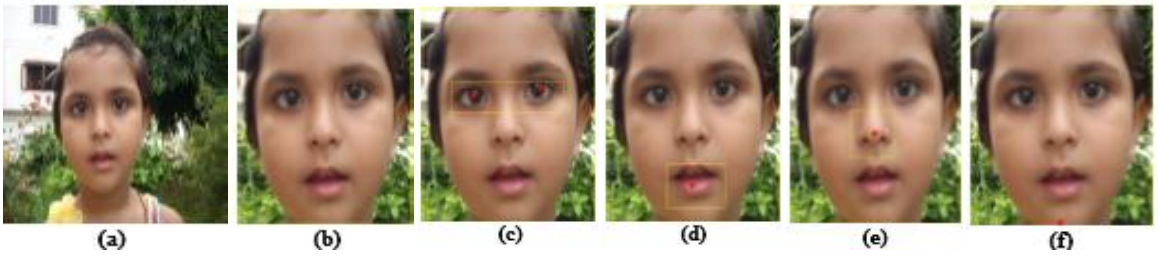


Figure 15. Image pre-processing: a) Face Image. b) Cropped face image. c) Eye pair detected image. d) Mouth detected image. e) Nose detected image. f) Chin detected image.

Secondly a series of features are extracted. The global features measured are the following:

- F1 = (distance from left to right eye ball) / (distance from eye to nose)
- F2 = (distance from left to right eye ball) / (distance from eye to lip)
- F3 = (distance from eye to nose) / (distance form eye to chin)
- F4 = (distance from eye to nose) / (distance from eye to lip)
- F6 is the angle between right eyeball, mouth point, and left eye ball in face image

Using the Grid features of face image feature F5 is calculated based on the wrinkle geography. It includes the regions shown in Figure 16a. To calculate F5 the colour image is converted into grey scale image, then canny edge detection technique is applied on that image, giving a binary face image with wrinkle edge (Figure 16b). The white pixels of the wrinkle region give wrinkle information in the face image. The white pixels are represented as 1 and black pixels as 0. The sum of the binary pixels in wrinkle region will be more when more wrinkles are present.

- F5 = (sum of pixels in forehead region / number of pixels in forehead region) + (sum of pixels in left eyelid region / number of pixels in left eyelid region) + (sum of pixels in right eyelid region / number of pixels in right eyelid region) + (sum of pixels in left eye corner region / number of pixels in left eye corner region) + (sum of pixels in right eye corner region / number of pixels in right eye corner region)

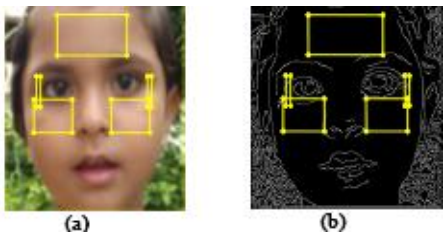


Figure 16. a) Grid features region of face image. b) Canny edges of face image.

Thirdly the classification is done based on the above six features using K-Means clustering algorithm into 2, 3, and 4 age range groups. Then the same classification is done excluding F6 and finally just F5 feature is considered.

3.1.5.2. Results

The results show that as more range groups to classify, lower is the correct classification percentage. They also prove that only wrinkle features is more important for age group classification. The results of the classification using only feature F5 are shown in Table 2.

Table 2. Classification results based on wrinkle feature F5.

No. of group	Group No.	Age Range In years	No. of faces actually in this group	No. of faces falling in this group	Correct percentage
2	1	1-40	34	32	96%
	2	41-80	16	16	
3	1	1-30	29	27	84%
	2	31-45	10	5	
	3	46-80	11	10	

3.1.5.3. Advantages

- The features used for classification are not the colour or grey level (which may vary in different lighting conditions) but the distances between face parts or the density of edges which could possibly be wrinkles.
- The face is detected and cropped.
- The classification rate reaches 96% accuracy for two age group cluster.

3.1.5.4. Disadvantages

- No information provided about changing lighting conditions or rotation of the images in the dataset. No information provided about their resolution either.
- Low number of classification groups.
- For proper eye and eyeball detection the image should be without spectacle. The image should be of a single straight frontal face.

3.1.6. Facial Age Group Classification

In [6] Prajapati et al. on 2014 improved the previous approach by increasing the number of age groups to five, see Table 3:

Table 3. Age group name with its age range.

Age Group Name	Age Ranges
Babies	0-2
Children	3-16
Young Adults	17-30
Middle-aged Adults	31-45
Old Adults	Above 45

3.1.6.1. Process

First the image is pre-processed; this consists in converting the RGB image to grey scale. Then the image is cropped to the face region and resized to 256 pixels height x 192 pixels width.

Later the face is detected using MATLAB function implementing Viola & Jones algorithm. To detect the face *FrontalFaceCART* Model is used. The eye pair, nose and mouth are also detected.

Then, various distances between the eye pair, nose, mouth and chin are calculated. These distances will differentiate between baby and child, as they are closely related to the craniofacial aging, see Figure 17.

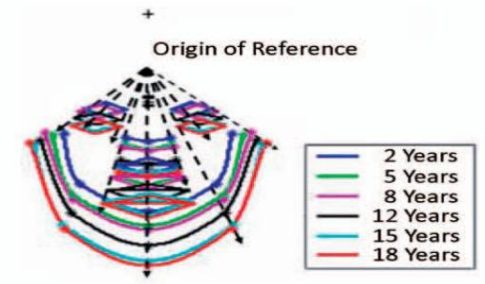


Figure 17. Craniofacial growth model for age progression.

To differentiate young, middle-aged and old wrinkle features are extracted, the regions considered can be found in Figure 18. To detect the wrinkles gradient operations are used. Finally the wrinkle density (wrinkle pixels / total pixels) is calculated. To correctly classify the threshold value must be adjusted appropriately.

Figure 18. Wrinkle's regions to be considered for detection.

This method was tested in the FG-Net Database

3.1.6.2. Advantages

- Local facial features are used for classification: feature distances and wrinkles.
- Classification is done between 5 age groups, including babies.

3.1.6.3. Disadvantages

- No information about the correctly classified rate.

3.1.7. Gender Classification using Geometric Facial Features

In [7] Kalm and Guttikonda on 2014 measured facial distances for gender classification. Mathematical operations on frontal pose face images were done using MATLAB.

3.1.7.1. Process

First the input image is pre-processed. The RGB colour images are converted into a two dimensional grey scale and resized into 256x256. Then, noise reduction is applied. Finally an edge detection method is used for detecting and extracting the boundary features. More precisely, canny edge detector is used. The result of this method can be seen in Figure 19.

a) b)

Figure 19. a) Input image. b) Edge detection of Input image.

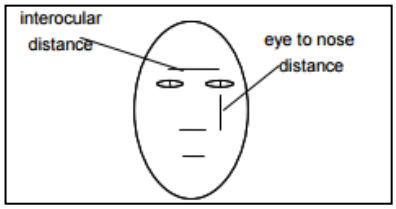
After the pre-processing, the features are extracted. There are two types of features:

- Geometric-based features (local features) use high level information. Distances between nose, eyes, eyebrow thickness or face length are examples of this kind.
- Appearance-based features (global features) use low level information about face image areas based on pixel values. Some popular global features are: Texture features as LBP, histogram of gradients and wavelet transformation of image coefficient such as Gabor wavelet.

The features used in this paper are local features. To extract those features, the eyes, nose, lips and chin have to be located. To do so, the images are divided into four equal parts; the centre of each part is calculated using the region properties. The distances between the centroids are calculated. The eyes are located at upper right and upper left part and then the lips. To locate chin and nose, values of +45 and -30 were added respectively to the centroid value of lips.

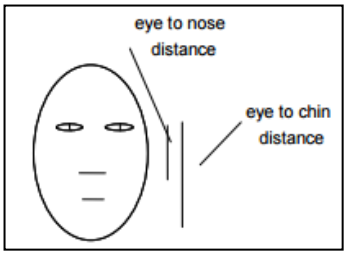
The facial features to be considered are the following (see Figure 20):

$$ratio_1 = \frac{LeftEye\ to\ RightEye\ distance}{Eye\ to\ Nose\ distance}$$



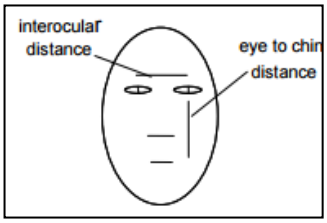
a)

$$ratio_2 = \frac{Eye\ to\ Nose\ distance}{Eye\ to\ Chin\ distance}$$



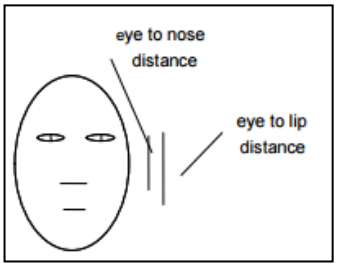
b)

$$ratio_3 = \frac{LeftEye\ to\ RightEye\ distance}{Eye\ to\ Chin\ distance}$$



c)

$$ratio_4 = \frac{Eye\ to\ Nose\ distance}{Eye\ to\ Lip\ distance}$$



d)

Figure 20. Facial features: a) Ratio 1, b) Ratio 2, c) Ratio 3, d) Ratio 4.

The threshold values for female have been decided as:

$$ratio_1 \geq 1.1 \;\&\&\; ratio_2 \geq 0.745 \;||\; ratio_3 \leq 1.3714 \;\&\&\; ratio_4 \geq 0.6404$$

The threshold values for male have been decided as:

$$ratio_1 \leq 1.09 \;\&\&\; ratio_2 \leq 0.744 \;||\; ratio_3 \geq 1.3714 \;\&\&\; ratio_4 \leq 0.64$$

If all these values satisfy the above conditions then the given image can be a male or a female.

3.1.7.2. Results:

125 images from mostly the ORL Database of both male and female were trained; ratios were also calculated and compared with the threshold (see Figure 21 for an example of the ratios values in male and female). The testing of images is performed using Indian face database. The gender recognition rate is 95.6%.

a) b)

Figure 21. Example of a) Male image and ratios values, b) Female image and ratios values.

3.1.7.3. Advantages

- Parametric Method for classification. It is not the grey value of pixels what is compared but the distances of certain facial features
- High gender recognition rate: 95.6% (for images of resolution 256x256)

3.1.7.4. Disadvantages

- Vague description of the method to locate the eyes, nose, lips and chin.

3.1.8. Face Recognition and Using Ratios of Face Features in Gender Identification

In [8] Bao et al. on 2015 developed a system of algorithms for human face identification and for gender classification.

3.1.8.1. Process

The identification of the face location is done by applying the Distance Regularized Level Set Evolution (DRLSE) method to a low resolution image of 50x50.

The gender identification is done by analysing and processing the left eye and eyebrow features, which have been recognized using Gabor wavelets. An example of the output image using 2-D Gabor filter is given in Figure 22.

Figure 22. Magnitude output of Gabor filter.

From that image three ratio parameters from the left eye region only are extracted. First the left eye zone is located dividing the face into 3x3 square sub-regions as in Figure 23a. Left eye is typically located in the first region, to ensure all eye and eyebrow are included that region is extended, see Figure 23b.

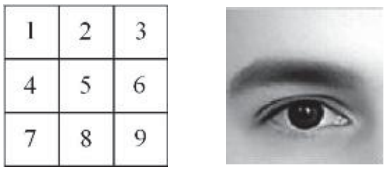


Figure 23. a) 3x3 squares in which the face is divided. b) Extended image of left eye area.

The selected area is then divided into eyebrow and eye. The adaptive threshold algorithm by Niblack is used, which finds threshold values within a local window by the calculation of pixel wise threshold using the local mean and sthe tandard deviation for a pixel.

The three parameters used for classification are the following:

$$ratio_1 = \frac{EyebrowLenght}{EyebrowHeight}$$

$$ratio_2 = \frac{EyeballLenght}{EyeballHeight}$$

$$ratio_3 = \frac{EyebrowHeight}{EyeballHeight}$$

The length and height of the left eyeball are measured as the size of the rectangle containing the extracted visible eye area (Figure 24c and Figure 25d). The eyebrow height is the height of the middle point location of the eyebrow. The length of the eyebrow is approximated as the sum of the two line segments of the eyebrow (Figure 24a and Figure 25b).

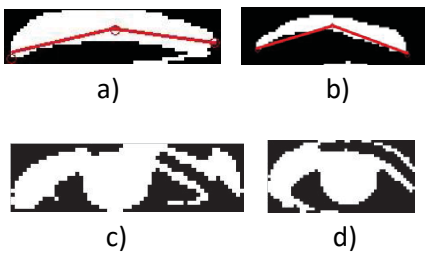


Figure 24. Extracted binary images of: a) Male eyebrow, b) Female eyebrow, c) Male eye, d) Female eyebrow.

Those three parameters are calculated for all images of their custom training set (21 female and 21 male images). Fisher's Linear Discriminant (FLD) method is then used to separate the two classes (genders) building a criterion with the equation of the discrimination surface:

$$y_0 = c_1 \cdot x + c_2 \cdot y + c_3 \cdot z$$

The function of the discrimination surface results in:

$$-0.0457x - 0.0069y + 0.2510z + 0.2691 = 0$$

The points calculated from the training set and the discrimination surface can be found in Figure 25.

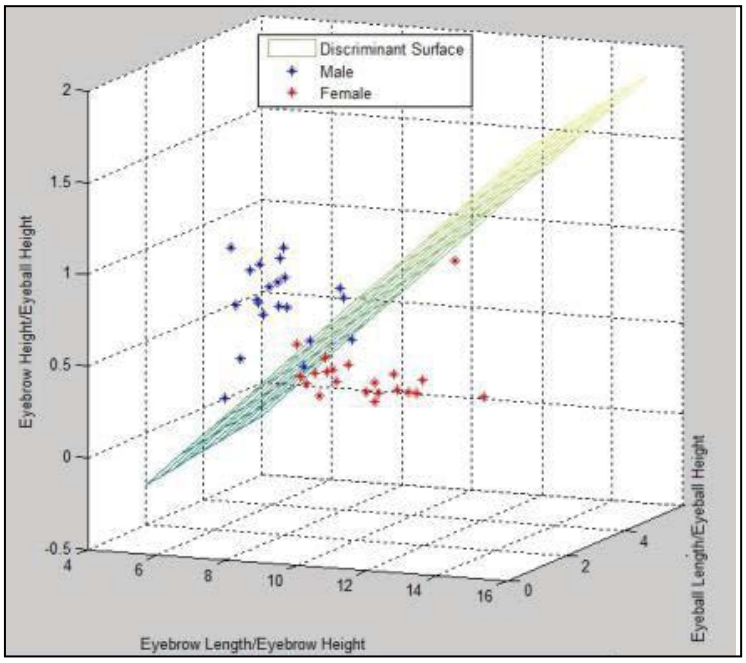


Figure 25. Discriminative surface and data used to build the discrimination function.

3.1.8.2. Results

This criterion is applied to a new set of 20 men and 20 woman's frontal face images for gender recognition. The correctly recognized men were 19, being the 95%. The correctly recognized women were 18 (90%).

Previous studies extracted features of just the eyebrow, reaching a maximum recognition rate of 73.3% for male and 84% for female. The method described here proved that both eyeball and eyebrow information provides more significant gender information, as the classification rates are more accurate.

3.1.8.3. Advantages

- Parametric Method for classification. It is not the grey value of pixels what is compared but the distances of certain facial features
- Relatively low resolution image (50x50) when recognizing the face region.

3.1.8.4. Disadvantages

- Precise calculation of the eye and eyebrow position is needed
- No information provided about the dataset: lighting conditions, face rotation, background…
- No information provided about the resolution when recognizing and calculating eye and eyebrow features.

3.2. Face Databases

Here there is a collection of Databases used in literature and available online.

3.2.1. Olivetti Research Laboratory (ORL) Database

This database from the AT&T Laboratories Cambridge [9] contains a set of face images taken between April 1992 and April 1994.

There are ten different images of each of 40 distinct subjects (4 female and 36 male). For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The files are in PGM format. The size of each image is 92x112 pixels, with 256 grey levels per pixel. The images are organized in 40 directories (one for each subject), which have names of the form sX, where X indicates the subject number (between 1 and 40). In each of these directories, there are ten different images of that subject, which have names of the form Y.pgm, where Y is the image number for that subject (between 1 and 10).

A preview of the images of four subjects can be found in Figure 26:

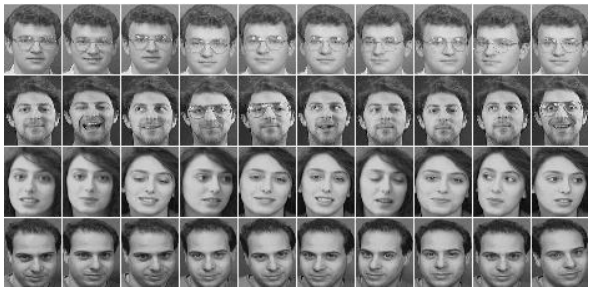


Figure 26. Preview of ORL Database

3.2.2. Yale Database

This database [12] contains 165 grey face images taken in 19. There are 11 different images of each 15 distinct subjects. The 11 images illustrate different facial or lighting conditions: normal, without glasses, with glasses, left-light, right-light, centre-light, happy, sad, wink, sleepy and surprised.

A preview of the images of four subjects can be found in Figure 27.

Figure 27. Preview of Yale Database

3.2.3. Faces94 database

This database is part of a group of databases [13] constructed by Dr Libor Spacek. Faces94 database contains images of 153 subjects (20 female and 133 male). The images resolution is 180 x 200 pixels in 24 bit RGB JPG format. The subjects (mostly between 18-20 years old) were asked to speak while a series of pictures were taken. The speech was used to introduce facial expressions variations. There is no variation in lighting. The background image of the images is plain green.

A preview of three of the images of eight subjects the can be found in Figure 28.

Figure 28. Preview of Faces94 Database

3.2.4. Faces95 database

This database is part of a group of databases [13] constructed by Dr Libor Spacek. Faces95 database contains 20 images for the 72 subjects (both male and female). The images resolution is 180 x 200 pixels in 24 bit RGB JPG format. The subjects (mostly between 18-20 years old) were asked to take a step forward while the series of photos were taken. This movement was used to introduce significant head (scale and position) variations. As the subject goes forward significant lighting changes occur on faces due to the artificial lighting arrangement. The background consists of a red curtain; this colour may vary due to the shadows caused when the subject moves forward. There is also some facial expressions variation.

A preview of ten of the images of one subject the can be found in Figure 29.

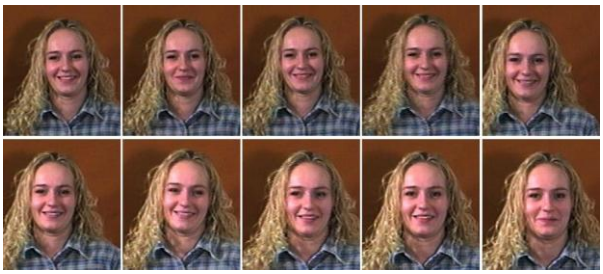


Figure 29. Preview of Faces95 Database

3.2.5. Faces96 database

This database is part of a group of databases [13] constructed by Dr Libor Spacek. Faces96 database contains 20 images for the 152 subjects (both male and female). The images resolution is 196 x 196 pixels in 24 bit RGB JPG format. The subjects (mostly between 18-20 years old) were asked to take a step forward while the series of photos were taken. This movement was used to introduce significant head (scale and position) variations. As the subject goes forward significant lighting changes occur on faces due to the artificial lighting arrangement. Complex and different backgrounds are introduced (glossy posters). There may be also some facial expressions variation.

A preview of ten of the images of one subject the can be found in Figure 30.

Figure 30. Preview of Faces96 Database

3.2.6. Grimace database

This database is part of a group of databases [13] constructed by Dr Libor Spacek. Grimace database contains 20 images for the 18 subjects (both male and female). The images resolution is 180 x 200 pixels in 24 bit RGB JPG format. The subjects (mostly between 18-20 years old) made grimaces while the series of photos were taken. This movement was used to introduce significant head rotation and tilt variations as well as major facial expressions variations. The background is plain.

A preview of five of the images of one subject the can be found in Figure 31.

Figure 31. Preview of Grimace Database

4. Materials

Here there is a reference to the camera used and the data obtained from it, then a description of the datasets obtained from the camera to test the experiment proposed.

4.1. Camera: Creative Senz3D

The camera used is the Creative Senz3D (Figure 32). This camera offers RGB (Figure 33a), Depth (Figure 33c) and IR or Confidence (Figure 33d) images plus the *x,z,y* World 3D Point coordinates. The second coordinate is named *z* as it contains the depth information. The RGB image without background (RGBD) is also processes meshing the Depth and RGB image (Figure 33b). The camera specifications are the following:

- RGB video resolution HD 720p (1280 x 720) or VGA (640 x 480)
- IR and Depth resolution QVGA (320 x 240)
- Frame rate up to 30fps
- FOV (Field-of-View) 74°.
- Range 0.5ft ~ 3.25ft. (0,15m ~ 1m)

Figure 32. Camera Creative Senz3D

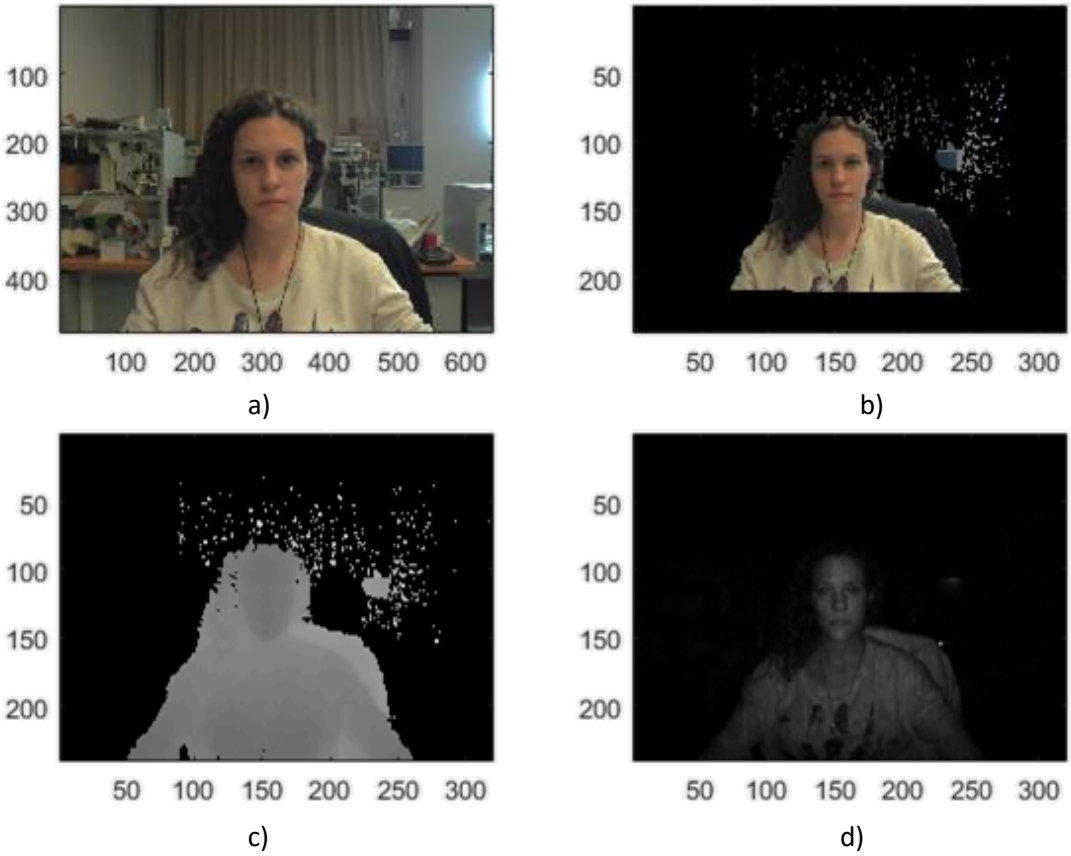


Figure 33. Images obtained by the camera: a) RGB, b) RGBD, c) Depth, d) IR.

To represent the World 3D Point coordinates, first the *pointCloud* MATLAB function is used to directly store the World 3D Point coordinates in a point cloud variable. Then the colour of the point cloud is defined as the one given by the RGBD image. Finally to point the point cloud the *pcshow* function is used (Figure 34).

$$ptCloud = pointCloud(WorldPoint);$$
$$ptCloud.Color = uint8(RGBD);$$
$$pcshow(ptCloud);$$
$$xlabel('X')$$
$$ylabel('Z')$$
$$zlabel('Y')$$

The process of obtaining the data from the camera and updating the point cloud can be done in real time, taking around 0.1s to do all the process (Using MATLAB R2015b, on a PC with 2.5GHz processor, 4GB RAM and Windows 10 64 bits) .

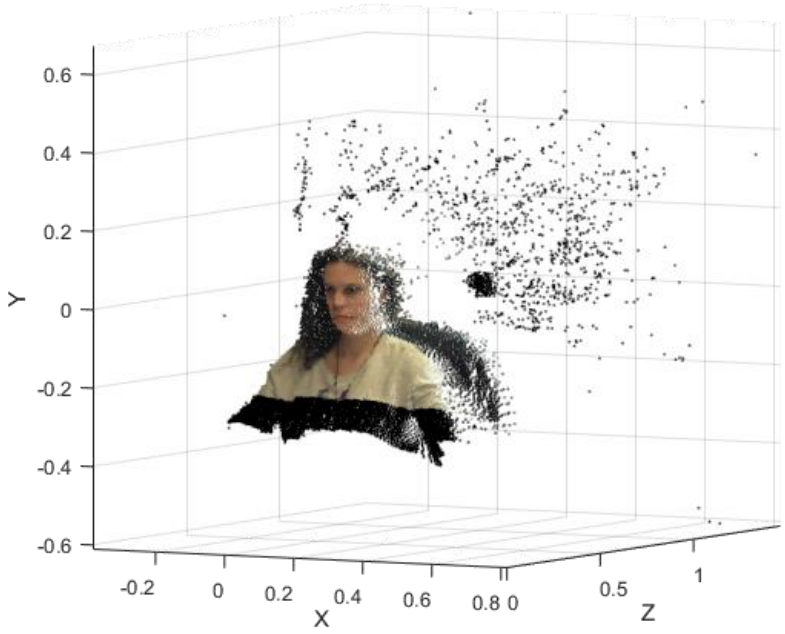


Figure 34. Representation of the original *x,z,y* World 3D Point Coordinates colored with RGBD image.

The images obtained by the camera can also be processed to erase noise and unwanted points. This process is done taking into consideration the depth image and consists on deleting all the regions with less than 50 pixels connected, then, to avoid losing hair or other information in the border of the regions, they are dilated two pixels. Finally the biggest region is the one selected. This final region is applied for the RGBD, Depth and IR Image as well. (Figure 35). Notice that RGB image doesn't suffer any change. In case of the IR image, the change is not noticeable due to the fact that the points deleted are nearly black and they merge with the background.

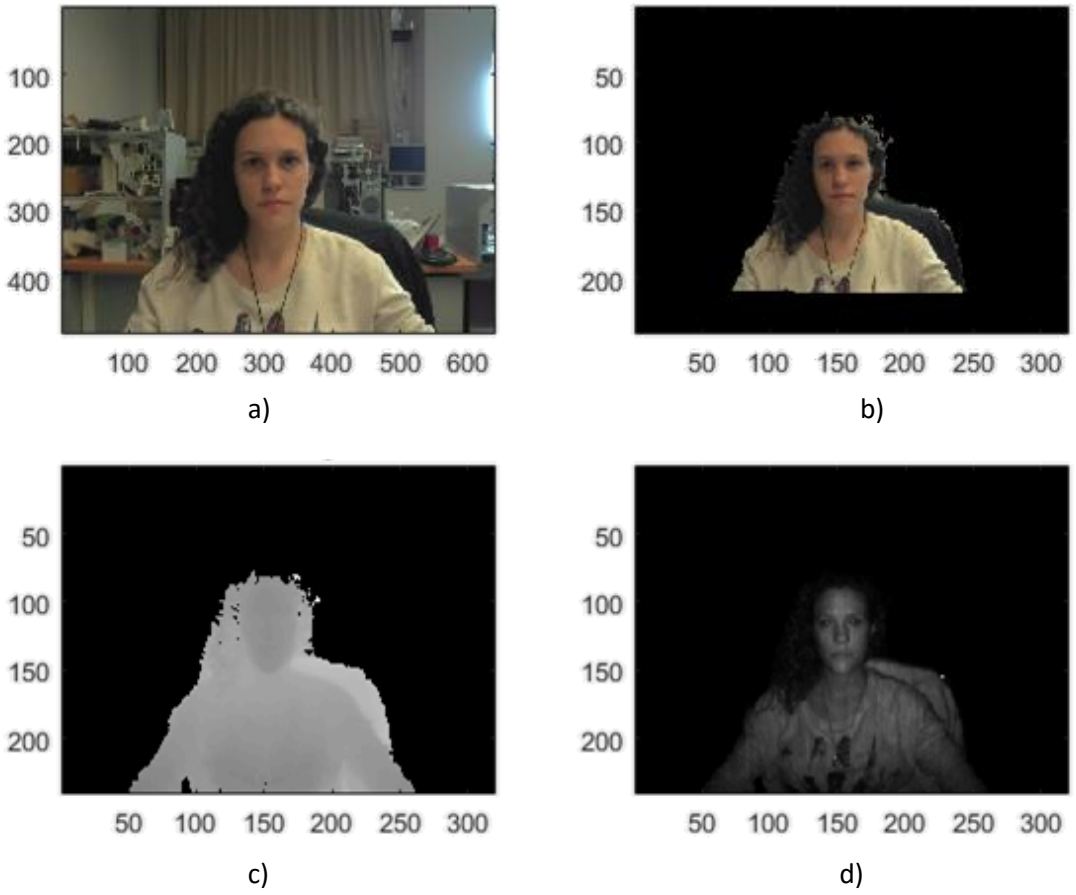


Figure 35. Processed images. a) RGB, b) RGBD, c) Depth, d) IR.

The point cloud can also be processed on its own using the *denoise* MATLAB function. This will remove the outlier points, defined as the ones which have an average distance to its K nearest neighbors above a threshold. The number of neighbors is set to 4 and the threshold to 0.05 (Figure 36):

$$ptCloud = pcdenoise(ptCloud, 'Threshold', 0.05, 'NumNeighbors', 4);$$

Applying this *denoise* function the time spent to capture the image and plot the point cloud raises to 0.785s.

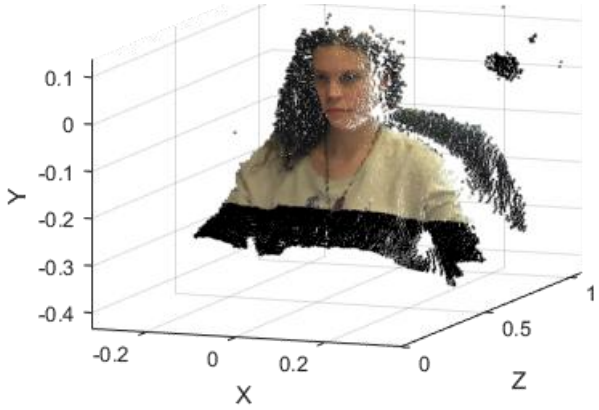


Figure 36. Processed World 3D Points.

4.2. Datasets

To test our proposal, three different datasets of seven data units have been created. Each data unit includes RGBD images, depth images (240x320 pixels) and world 3D points obtained by the camera. The three datasets contain different types of subjects:

- Dataset 1: Same subject with different facial expressions sitting at the same distance from the camera. Figure 37 shows 4 example RGBD images of this dataset.
- Dataset 2: Same subject sitting at different distances from the camera. Figure 38 shows 4 example RGBD images of this dataset.
- Dataset 3: Different subjects. Figure 39 shows 4 example RGBD images of this datasets.

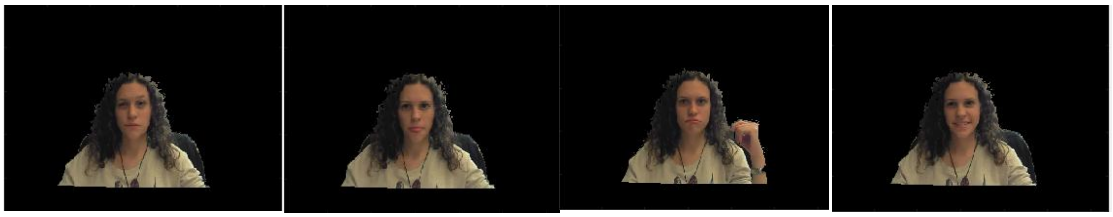


Figure 37. Sample RGBD Images of Dataset 1.

Figure 38. Sample RGBD Images of Dataset 2.

Figure 39. Sample RGBD Images of Dataset 3.

5. Face recognition based on nose identification

The face recognition process is based on the fusion of novel local low-level features from the proposal described in section 3.1.4. [4]. In this work the nose shape will be described by four histograms representing four different features, those histograms will be compared for different subjects with the objective of accomplishing subject recognition. The procedure is based on face detection, nose region determination, 3D data acquisition, nose tip detection, feature acquisition and feature comparison with correlation and subtraction methods.

5.1. Face detection

The face region is detected applying Viola & Jones on the RGBD images. To perform face detection first the cascade object detector is set up using the MATLAB constructor *vision.CascadeObjectDetector*. The classification model is chosen as the *FrontalFaceCART* since it detects faces that are upright and forward facing.

Once the model is built the cascade object is obtained and the *step* method is called with the input image and the cascade object, obtaining the different regions where a face is detected. Even though this method can detect more than one face region, the central face is the one chosen (the one with its central column closer to the central column of the image).

The face region (Figure 40, green square) is described as:

$$region\_face = [first\ column, first\ row, width, height];$$

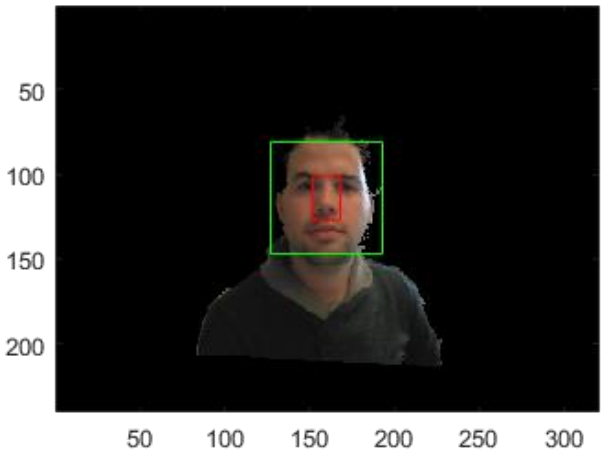


Figure 40. Face Region (green) and Nose Region (red).

The whole face detection process is implemented as:

```
% Build detector
detector.Face = vision.CascadeObjectDetector('FrontalFaceCART');
% Search for faces
bbox_faces = step(detector.Face,Image);
% Number of faces detected
faces_detected = size(bbox_faces,1);
% Initilise output region_face
region_face = [];
```

```
% Check if some faces where detected
if faces_detected >= 1
        % Check if more than one face is detected and restrict to central face
        if (faces_detected > 1)
                % More than one face tected
                % Initial distance to centre (reference distance)
                d_c = 1000000;
                % Analyse all the faces detected
                for i = 1:1:faces_detected
                        % Search the centre of the face's box (x axis)
                        x_c = bbox_faces(i,1) + bbox_faces(i,3)/2;
                        % Check if distance between centre of the face's box (x_c) and the
                        % centre of the image (n_col/2) is less than the smallest found (d_c).
                        if (abs(x_c - n_col/2) < d_c)
                                % Update the reference distance
                                d_c = abs(x_c - n_col/2);
                                % Change the new reference face's box
                                region_face = bbox_faces(i,:);
                        end
                end
        else
                % Just one face detected
                region_face = bbox_faces;
        end
end
```

5.2. Nose region determination

From the face region, the nose region (Figure 40, red rectangle) is determined by a trial and error procedure as:

$$region\_nose(1) = uint8(region\_face(1) + region\_face(3) * 0.38);$$
$$region\_nose(2) = uint8(region\_face(2) + region\_face(4) * 0.3);$$
$$region\_nose(3) = uint8(region\_face(3) * 0.26);$$
$$region\_nose(4) = uint8(region\_face(4) * 0.4);$$

The camera obtains approximately 350 points (or pixels) in that region, this value may vary depending on the distance from the subject to the camera. This region is extracted from the Depth Image and the World 3D points.

5.3. 3D data selection

The nose regions from the depth image and the world 3D points obtained by the camera are processed to obtain four kinds of 3D data from which the features will be obtained:

- World 3D Points

  This data consists of the *x, z, y* world 3D point coordinates obtained by the camera. The second coordinate is the one representing the depth information.

  The *x* and *y* coordinates, $W1\_nas(:,:,1)$ and $W1\_nas(:,:,3)$ respectively, obtained by the Senz3D camera in the nose region are normalized to suppress the differences occasioned by the possible differences in distance between the subject and the camera. The series of operations performed to normalize these distances are the following:

```
offset_x = min(min(W1_nas(:,:,1)));
W1_nas(:,:,1) = W1_nas(:,:,1) − offset_x;

offset_y = min(min(W1_nas(:,:,3)));
W1_nas(:,:,3) = W1_nas(:,:,3) − offset_y;

div = max(max(W1_nas(:,:,1)))/0.035;
W1_nas(:,:,1) = W1_nas(:,:,1)/div;
W1_nas(:,:,3) = W1_nas(:,:,3)/div;
```

As a result the *x* coordinates range from 0 to 0.035. The *y* coordinates are ranged accordingly. The *z* coordinate (depth) is made to have a zero value at the point further from the camera and increase as getting closer to the camera.

Figure 41a shows the world 3D points of the nose colored with the RGBD image.

- Interpolated world 3D points.

This data consists of the interpolated world 3D points descried before.

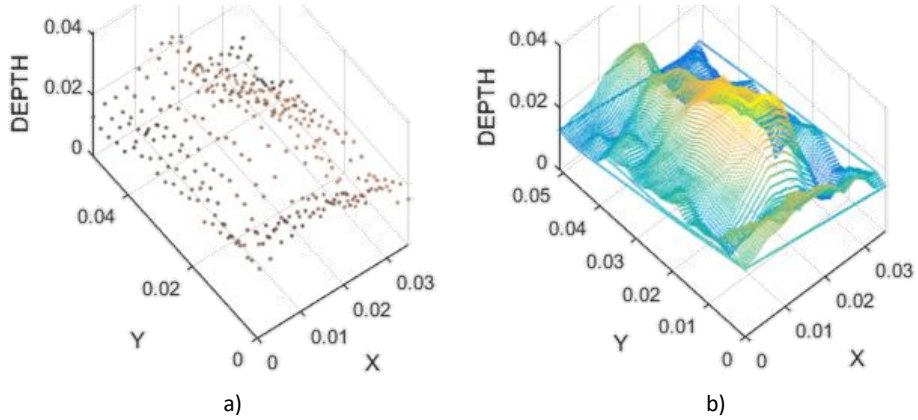
To obtain a better precision the world 3D Points they are interpolated by using the *scatteredInterpolant* function of MATLAB with the *natural neighbour interpolation method*. Finally a total amount of 10,000 new points inside the original nose region is obtained using that interpolation method (Figure 41b):

```
% Original points x,z,y
x_vals(:) = W1_nas(:,:,1);
y_vals(:) = W1_nas(:,:,3);
z_vals(:) = W1_nas(:,:,2);;

% Create scatter interpolant
F = scatteredInterpolant(x_vals', y_vals', z_vals');
F.Method = 'natural';

% Find x and y limits
x_min = min(x_vals);
x_max = max(x_vals);
y_min = min(y_vals);
y_max = max(y_vals);

% Create matrix x,y with 10,000 points inside the limits
Xq_orig = linspace(x_min, x_max);
Yq_orig = linspace(y_min, y_max);
Xq = [];
Yq = [];
for i = 1:100
    for j = 1:100
        Xq = [Xq Xq_orig(i)];
        Yq = [Yq Yq_orig(j)];
    end
end
% Interpolate
Zq = F(Xq, Yq);
```



a) b)

Figure 41. a) World 3D point cloud. b) Interpolated world 3D point cloud.

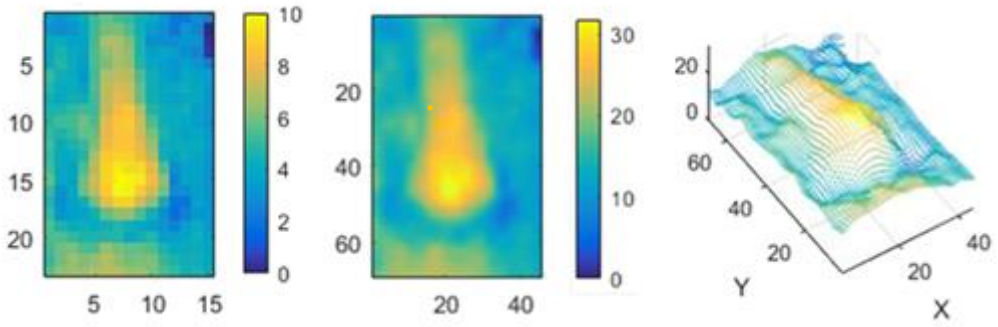
- Depth image

  This data consists of the original depth image of the nose region resized to meet a standard size of 15 x 22 pixels.

  This image is converted to 3D points: The pixel value is the depth information normalized from 0 to 10 and reversed, being 10 the point closer to the camera and 0 the further. This normalization is done to obtain an optimal relation between the pixel value (depth distance) and the distances between pixels (considered as the *x, y* spatial coordinates). Applying this range, the visualization of the nose is improved. (Figure 42a).

- Interpolated depth image

  This data consists of the interpolated depth image.
  To obtain a better precision the depth image was resized as nine times its original size, the pixel value is normalized from 0 to 30 for the same reason described above. (Figure 42b and 42c)



a) b) c)

Figure 42. a) Depth Image and b) Interpolated Depth Image with the nose tip as a red dot. c) Point Cloud representing the interpolated depth image.

5.4. Nose tip detection

From the four types of data containing the nose region, the nose tip (Figure 43, red dot) is chosen to be the point closer to the camera (the point with maximum value in z axis for World 3D point data, or the pixel with maximum depth value in the Depth image).

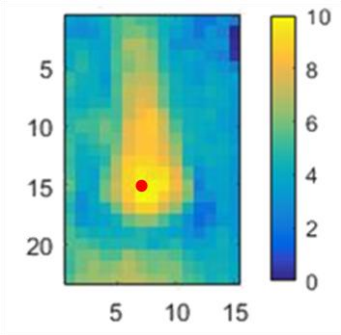


Figure 43. Nose region and nose tip (red dot).

5.5. Feature acquisition

The four features studied between the nose tip (*n*) and two random points within the nose region (*r1* and *r2*) are the following (Figure 13b):

- A: Corresponds to the angle between the two lines determined by the two random vertices and the nose tip:

$$v1 = r1 - n;$$
$$v2 = r2 - n;$$
$$A = acosd\ (\ sum(v1.* v2)/(\ norm(v1) * norm(v2))\ );$$

- C: Defined as the radius of the circumscribed circle to the triangle determined by the two random points and the nose tip.

  C is obtained with the MATLAB function provided in [14] by Johannes Korsawein. This function computes the centre and radius of a circle that interpolate a given triple of points in 3D space. The information useful for this experiment is the radius.

- D: Defined as the distance of the line between the two random vertices:

$$D(i) = norm(r1 - r2);$$

- N: Defined as the angle between the line defined by the two random vertices and the z-axis:

$$v1 = [0, 0, 1];$$
$$v2 = r2 - r1;$$
$$N(i) = acosd\ (\ sum(v1.* v2)/(\ norm(v1) * norm(v2))\ );$$

The features are calculated through 10,000 iterations of pairs of random points and the nose tip. Then the information obtained is summarized in one histogram per each feature:

- HA: Histogram representing the A feature.
- HC: Histogram representing the C feature.
- HD: Histogram representing the D feature.
- HN: Histogram representing the N feature.

All histograms have 180 bins as specified by Lie et al [4]. To obtain the histograms, the value of the centre of each bin is assigned. The two histograms which describe angle features (HA and HN) range from 0 to 180. The histograms which describe distance features (HC and HD) will range depending on the type of 3D data used. Values which are out of boundaries (and subsequently considered wrong) will be stored in the last bin of each histogram.

An example of histograms obtained with an interpolated depth image is shown in Figure 44.

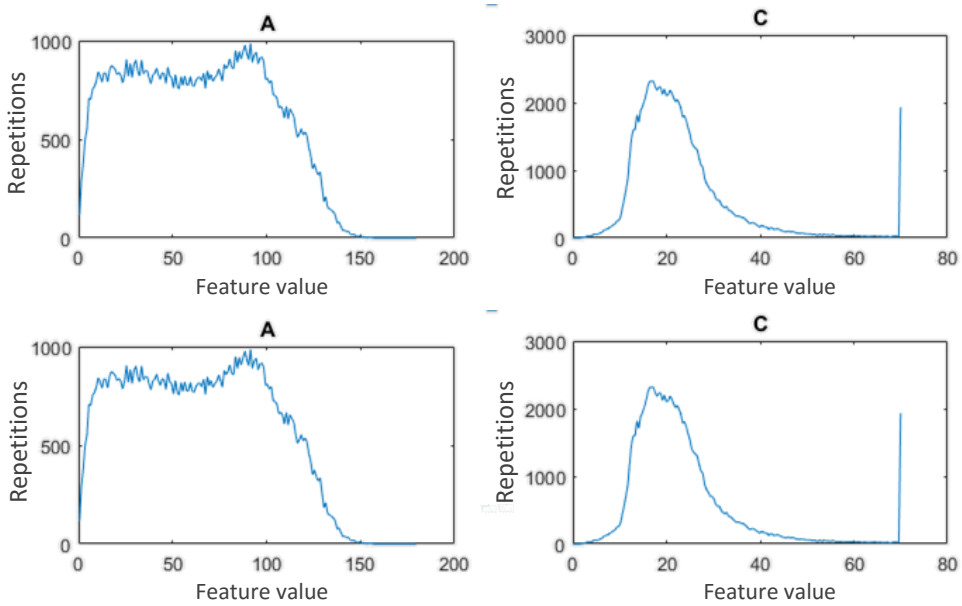


Figure 44. Histogram example for features A, C, D and N.

5.6. Feature comparison procedure

To compare the features, the coefficient of determination (from the correlation coefficient) and the linear differences (subtraction) are calculated between every pair of histograms of the same feature obtained from the same data type and same dataset (Among the ones described in section 4.2).

The *corrcoef* MATLAB function is used to compute the correlation coefficient. To enhance the differences, it is squared, obtaining the coefficient of determination (R^2).

To compute the subtraction, the vectors forming the histograms are normalized having an area below the curve of 1, then the subtraction method is computed as the cumulative addition of the differences between two normalized histograms.

The values obtained as de coefficient of determination and subtraction are stored in a matrix in its (*i, j)* position, where *i* and *j* refer to the subjects (in case of dataset 3) or subjects' images (in case of dataset 1 and 2) from which the histograms are compared.

If the procedure can perform subject recognition, the matrices summarizing the results of dataset 3 should contain more differences and lower correlation factors than the ones of the other two databases. Moreover, dataset 1 and 2 should offer approximately the same result, as the distance to the camera should not make a difference. To obtain visual results the matrices will be represented as colour images.

The different types of variables studied are summarised in Table 4. The aim is to find which combination of feature, data type and comparison would offer the best recognition results, which will be evaluated with our three custom datasets.

Table 4. Variables Studied

Feature Histogram	Data Type	Comparison
A C D N	World 3DPoints Interpolated World 3D Points Depth Image Interpolated Depth Image	Correlation Subtraction
Testing Datasets		
Dataset 1: Same subject, same distance form camera but different facial expression Dataset 2: Same subject, different distance from camera Dataset 3: Different subjects		

5.7. Results based on correlation

The different matrices obtained with the correlation comparison of the histograms are shown in Figures 45 to 48. The matrices present the correlation histogram comparison of the features obtained from: World 3D points (Figure 45), interpolated world 3D points (Figure 46), depth image (Figure 47) and interpolated depth image (Figure 48).

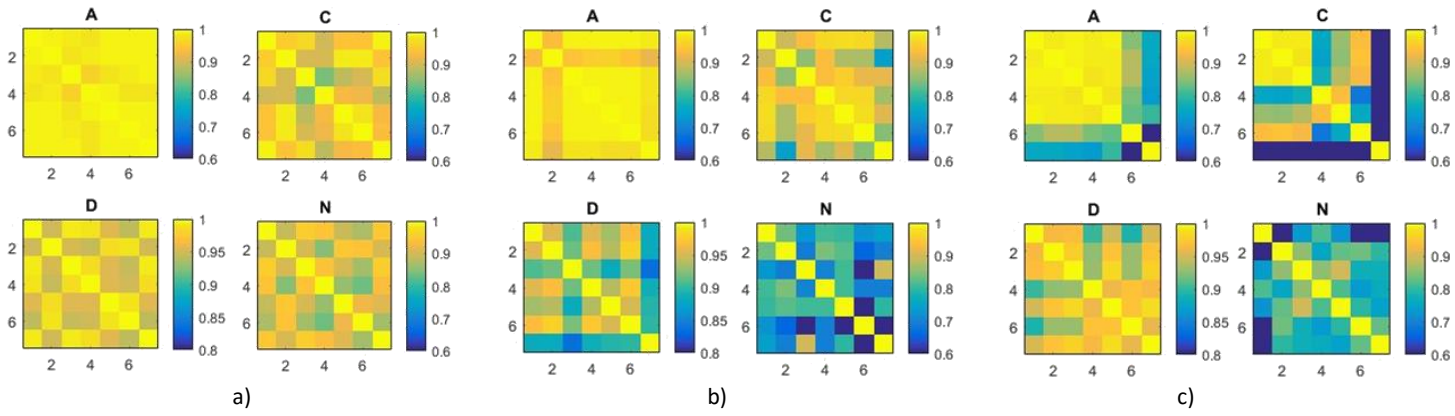


Figure 45. Correlation comparison matrices for the A,C,D and N feature histograms obtained from world 3D points. a) Dataset 1. b) Dataset 2. c) Dataset 3.

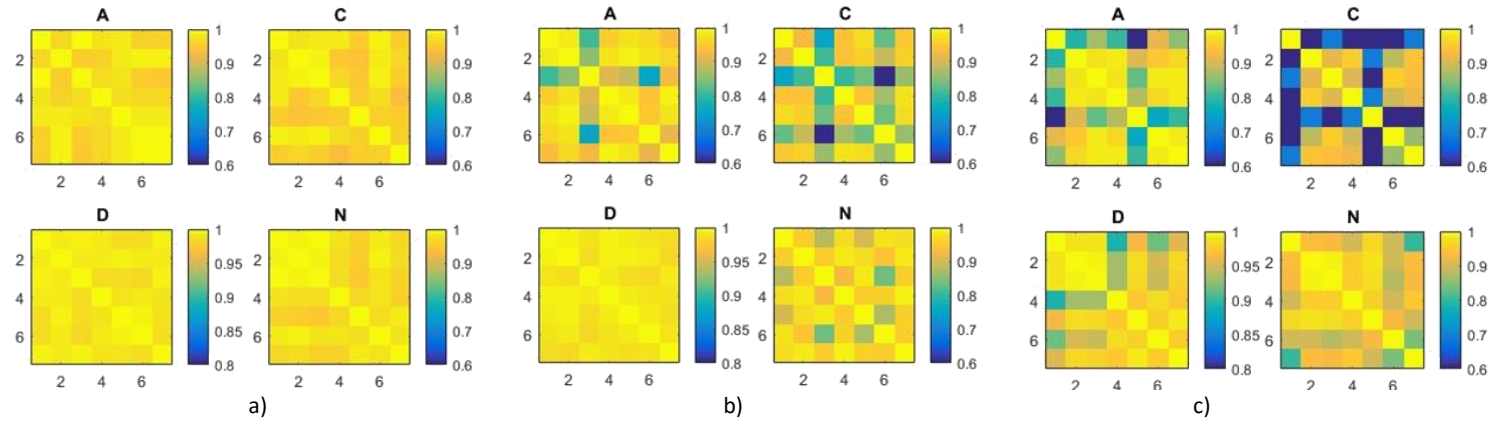


Figure 46. Correlation comparison matrices for the A,C,D and N feature histograms obtained from interpolated world 3D points. a) Dataset 1. b) Dataset 2. c) Dataset 3.

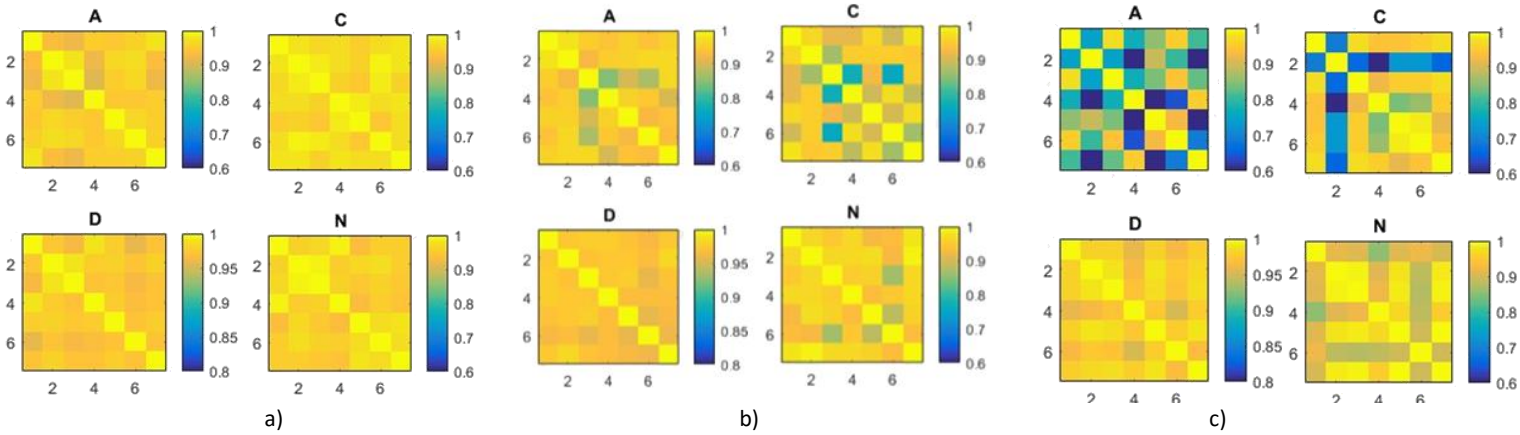


Figure 47. Correlation comparison matrices for the A,C,D and N feature histograms obtained from depth images. a) Dataset 1. b) Dataset 2. c) Dataset 3.

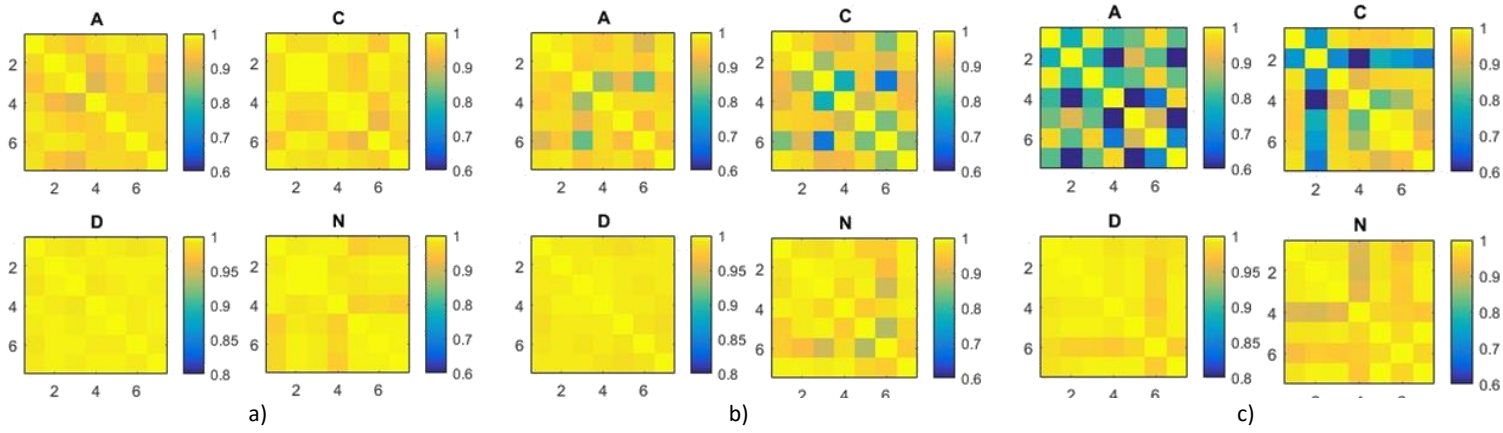


Figure 48. Correlation comparison matrices for the A,C,D and N feature histograms obtained from interpolated depth images. a) Dataset 1. b) Dataset 2. c) Dataset 3.

First of all, Figures 45, 46, 47, and 48 show that the positions (*i, i*) in every matrix have an R^2 of 1, as they compare exactly the same image.

Regarding the matrices obtained with world 3D points, they offer high coefficient of determination factors for dataset 1 (mostly above 0.85). However, in case of dataset 2, the correlation factors diminish. In this case, the factor with lower R^2 coefficients is N, being around 0.8 and presenting a decrease around 10% from the values of dataset 1.

Comparing dataset 2 and 3, the R^2 of A and C factors are sometimes greater in dataset 3 than in dataset 2 and for D and N factor, the R^2 are generally greater for dataset 3 than for dataset 2.

The above analysis leads to the conclusion that despite the original point cloud is processed to erase the differences in x and y coordinates when sitting at different distances from the camera, it appears that the depth coordinate should also be adjusted, since lower R^2 appear when the distance from the subject to the camera changes. As a result, there are no substantial differences between the coefficients of determination obtained from dataset 2 and 3. In case of the interpolated world 3D points, the matrices are neither conclusive, and there are not significant differences between the correlation factors for dataset 2 and dataset 3.

Regarding the results obtained with depth image, in general, the interpolated depth image offer higher R^2 as the noise which appears in the histograms will be diminished by the appearance of higher number of pixels in the image. For both interpolated and not interpolated depth image, dataset 1 and dataset 2 present high coefficients of determination (mostly above 0.9 for A, C and N features and above 0.95 for D features), being only slightly higher when the subject is always sitting at the same distance from the camera.

The depth histograms of dataset 3 of D and N features offer the same range of R^2 (above 0.85) as for dataset 2. The R^2 of C feature for dataset 3 differs only in one subject from the values obtained from dataset 2 (being around 0.7 for the differential subject). The A feature is the most discriminative as it offers higher differences between the R^2 of dataset 2 and dataset 3, being approximately 0.1 points lower in dataset 3 than for dataset 2. However, the correlation between the A feature of two different subjects can also offer R^2 above 0.9.

To discern if a pair of histograms are from the same or different subjects, it is applied a threshold of 0.85 on the R^2 of A feature (the one found as more discriminative) for the depth image (interpolated and not). This would lead to the following results in our reduced database: The false reject (FR) rate would be of 0% in dataset 1, recognising all the subjects as the same one. In case of dataset 2, there would be a pair of images classified as different subjects, leading to a FR of 4.76% in dataset 2. The false acceptance (FA) in dataset 3 would be of 28.6% for the interpolated depth image and of 33.33% for the depth image without interpolation, recognising six and seven pairs of images respectively as being from the same subject.

Visual results are shown in Figure 49 for depth images without interpolation and Figure 50 for interpolated depth images. Yellow colour indicate that the pair of subjects established by the row and column indexes are classified as being from the same subject, dark blue indicates they

are classified as different subjects. In relation to the FA and FR, dark blue indicates FR for dataset 1 and 2, and yellow indicates FA for dataset 3. The comparison of the identical same images is painted with cyan colour as they are not counted for the FA and FR rates:

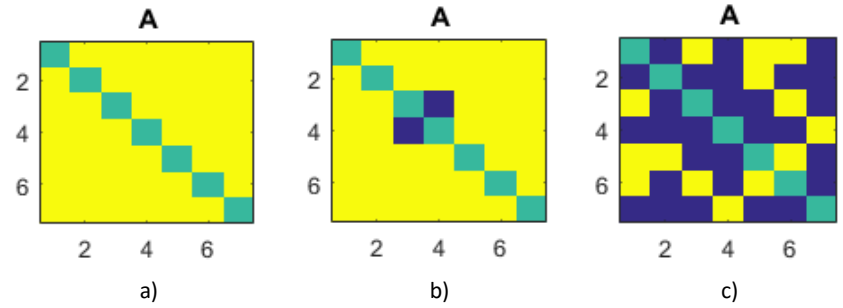


Figure 49. FA and FR for Feature A obtained from Depth Images with Correlation Comparison. Dataset 1 (a), Dataset 2 (b), Dataset 3 (c).

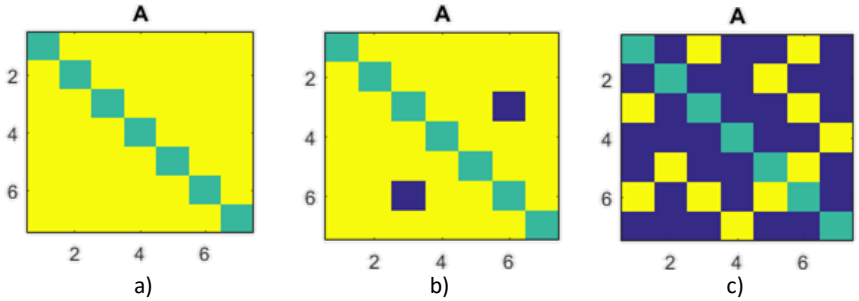


Figure 50. FA and FR Matrices for Feature A obtained from Interpolated Depth Images with Correlation Comparison. Dataset 1 (a), Dataset 2 (b), Dataset 3 (c).

5.8. Results based on subtraction

The different matrices obtained with the subtraction comparison of the histograms are shown in Figures 51 to 54. The matrices present the subtraction histogram comparison of the features obtained from: World 3D points (Figure 51), interpolated world 3D points (Figure 52), depth image (Figure 53) and interpolated depth image (Figure 54).

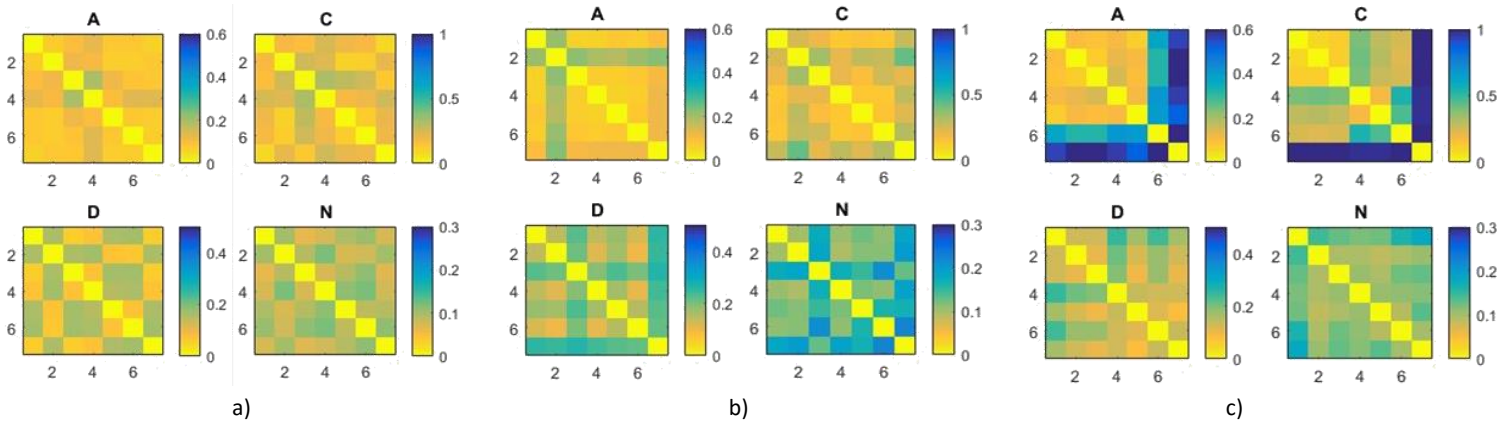


Figure 51. Subtraction comparison matrices for the A,C,D and N feature histograms obtained from world 3D points. a) Dataset 1. b) Dataset 2. c) Dataset 3.

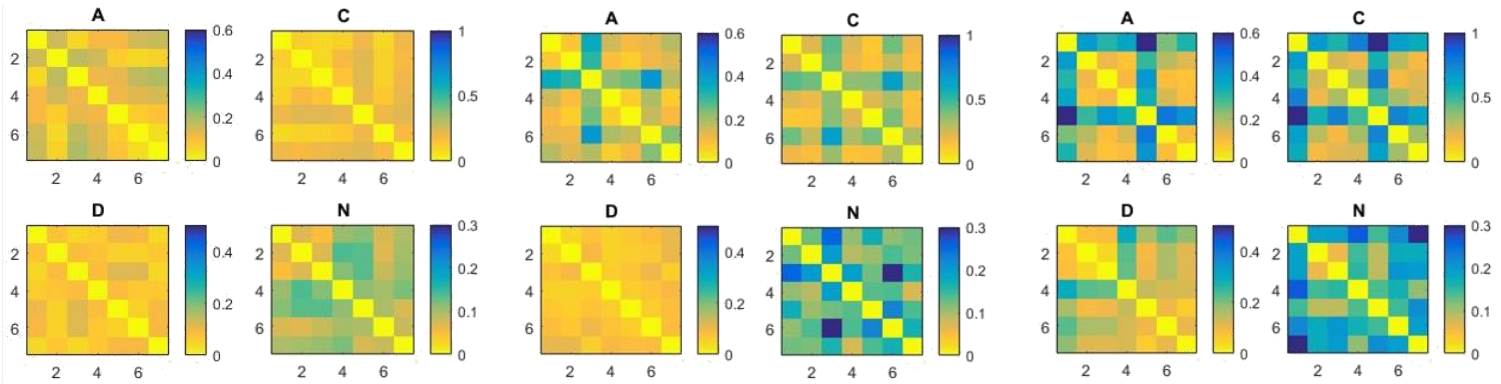


Figure 52. Subtraction comparison matrices for the A,C,D and N feature histograms obtained from interpolated world 3D points. a) Dataset 1. b) Dataset 2. c) Dataset 3.

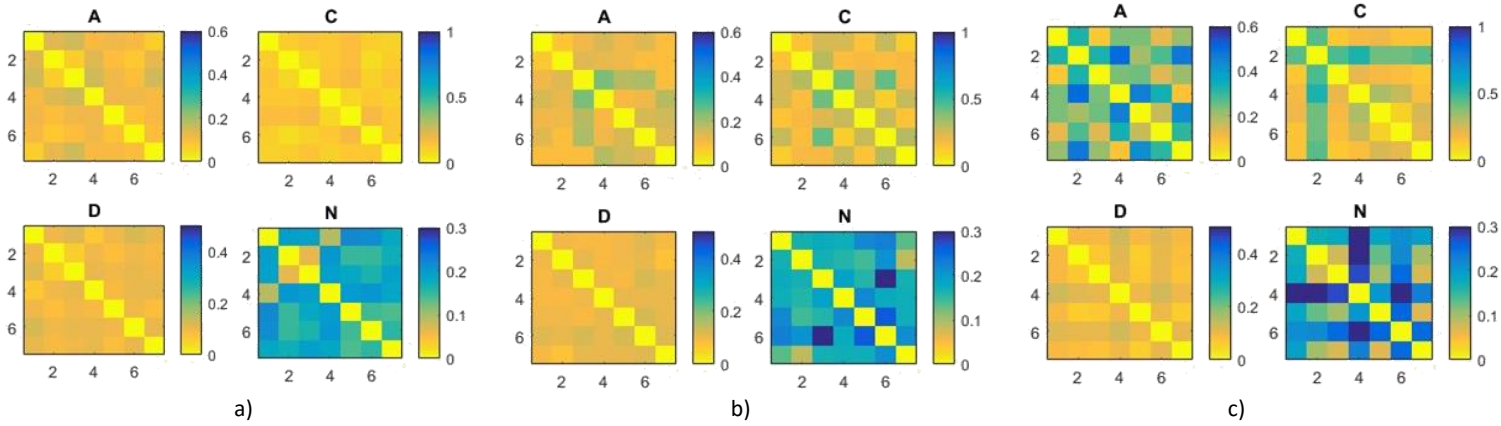


Figure 53. Subtraction comparison matrices for the A,C,D and N feature histograms obtained from depth images. a) Dataset 1. b) Dataset 2. c) Dataset 3.

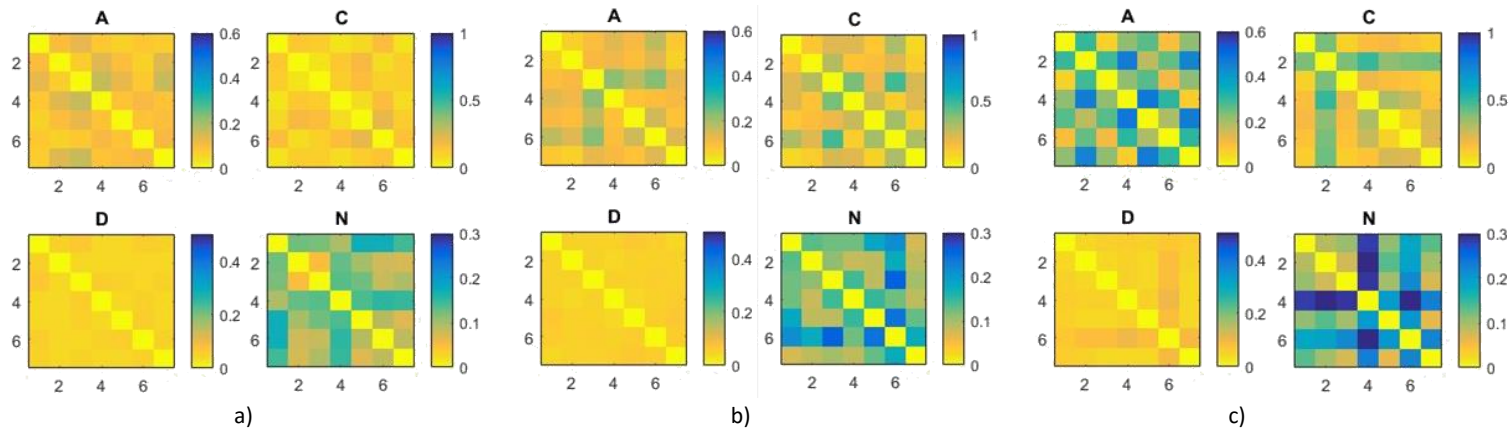


Figure 54. Subtraction comparison matrices for the A,C,D and N feature histograms obtained from interpolated depth images. a) Dataset 1. b) Dataset 2. c) Dataset 3.

In this case, Figures 51, 52, 54 and 53 show that the matrix's positions of the comparison between the same subject (dataset 3) or same subject's image (dataset 1 and 2) have all subtraction 0, as they are exactly the same images and present no differences.

Referring to the matrices obtained with world 3D points (interpolated or not) they do not offer substantial information to perform subject classification.

In case of depth images, the only useful parameter would be A, the same as in correlation comparison, since it is the only with relevant differences between dataset 2 and dataset 3. For this feature, in case of dataset 1 and 2 the sum of differences are usually less than 0.2, whereas in dataset 3 the average would be of 0.3.

The FA and FR rates are calculated with the subtraction comparison for feature A. The threshold is set to 0.19. The results for Depth Images (Figure 55) show that the FR is 0% in dataset 1 and 4.76% in dataset 2; the FA rate is 23.81% in dataset 3. The results for Interpolated Depth Images (Figure 56) show that the FR is 0% in dataset 1 and 9.52% in dataset 2; the FA rate is 23.81% in dataset 3.

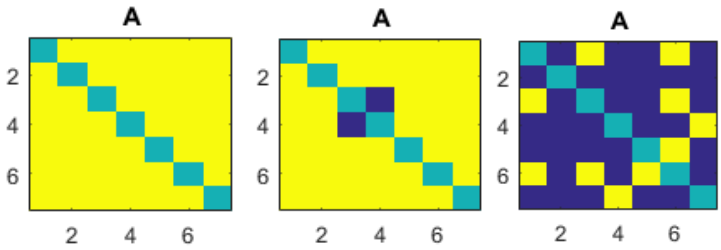


Figure 55. FA and FR Matrices for Feature A obtained from Depth Images with Subtraction Comparison. Dataset 1 (a), Dataset 2 (b), Dataset 3 (c).

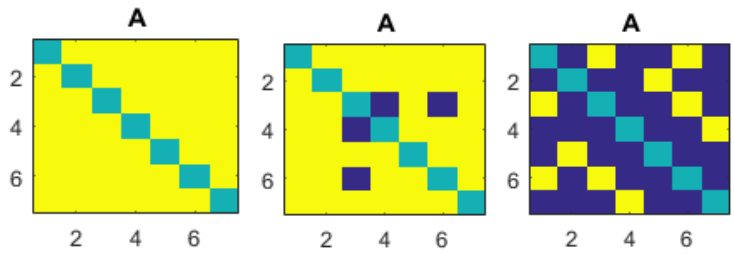


Figure 56. FA and FR Matrices for Feature A obtained from Interpolated Depth Images with Subtraction Comparison. Dataset 1 (a), Dataset 2 (b), Dataset 3 (c).

6. Conclusions

This work proposes the identification of a person based on the images acquired by a senz3D RGBD camera. This work starts with a revision of the state of the art in this subject where the following proposals have been analysed: Face recognition using Hidden Markov Model (HMM) and Singular Value Decomposition (SVD) Coefficients [1], Face Component Extraction using Segmentation Method on Face Recognition System[2], Optimizing Face Recognition using PCA [3], An efficient 3D face recognition approach based on the fusion of novel local low-level features [4], Age Group Estimation using Face Features [5], Facial Age Group Classification [6], Gender Classification using Geometric Facial Features [7], Face Recognition and Using Ratios of Face Features in Gender Identification [8]. This work also describes the face databases publicly available. In the future, this work will have application in the Assistant Personal Robot developed at the Robotics Laboratory of the University of Lleida.

The face identification procedure tested in this work is based on the 3D face recognition approach using the fusion of novel local low-level features [4]. In this proposal the nose shape is described by four different features compared as histograms. The procedure involves face detection, nose region determination, nose tip detection, feature acquisition and feature comparison with correlation (obtaining the coefficient of determination) and subtraction.

The RGBD face images have been analysed by considering world 3D points and depth image. The best results have been obtained in the cases of depth image. The comparison of world 3D points highlights the similarities and differences related to distance from the subject to the camera (despite having normalised the x-*width* and *y-height* distances) but not the differences between the subjects itself. This result is due to the fact that the *z-depth* coordinate should also be normalised as it is nonlinear, which means that an object of depth *Z*, will be captured by the camera as having different depth values depending on the overall distance to the camera.

The comparison has been made between every pair of subjects in three different datasets. Each dataset contain a different type of subject image. Dataset 1 contains seven images of one same subject with different facial expressions and same distance to the camera, dataset 2 contains seven images of one same subject in different distances to the camera and dataset 3 contains seven different subjects. From dataset 1 and 2 the false reject (FR) rate is obtained, from dataset 3 the false acceptance (FA) rate.

The identification results have been computed for each of the four features acquired by direct correlation and by the cumulative addition of the differences.

The comparison of the features obtained from the depth images of dataset 1 (same subject with different facial expressions siting at the same distance from the camera), have demonstrated that the process is invariant to expression change, as every pair of subjects present coefficients of determination above 0.95 for feature D and above 0.9 for the rest of features. This values lead to a FR rate of 0% for A feature. This dataset is the one which offers higher coefficients of determination.

The comparison of the A feature (angle between the nose tip and two random points in the nose region) is the one which offers higher similarities between the values obtained with dataset 1 and dataset 2, along with higher differences between dataset 2 and dataset 3. For this reason it is the feature which highlights the difference between different subjects and is not affected by the change in the distance from the subject to the camera. This is the reason why A feature is the one with better possibilities to perform subject recognition.

The FA and FR rates have been studied, the FA rate has values between 24% and 33% depending on the data type (depth image interpolated and not interpolated) and the comparison procedure (correlation or subtraction). However, the FR rate is null for dataset 1 and (in the best cases) 4.75% for dataset 2.

The case which offers better results (lower FA and FR rates) is the one based on the depth images without interpolation and comparing A feature histograms by subtraction. The results are a FA rate of 4.76% and a FR rate of 23.81%.

The histograms of the A feature obtained from the depth image from the dataset 2 and dataset 3 can be seen in Figure 57 and Figure 58 respectively, different images or subjects are represented with different colours. Those figures show that for dataset 2 the values of A feature from 10 to 90 have all approximately 900 repetitions, the repetitions decrees until reaching zero between the values of 130 and 140. In case of dataset 3, there are higher differences in the histogram values. Nevertheless, there are also histograms of dataset 3 which coincide.

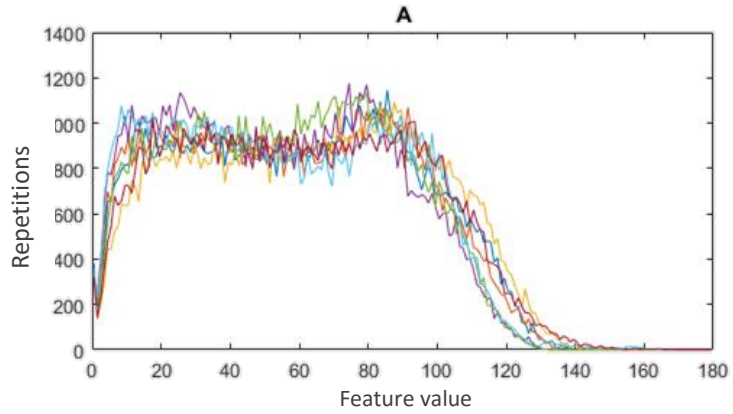


Figure 57. Histograms Obtained from Depth Images of the Dataset 2 representing feature A.

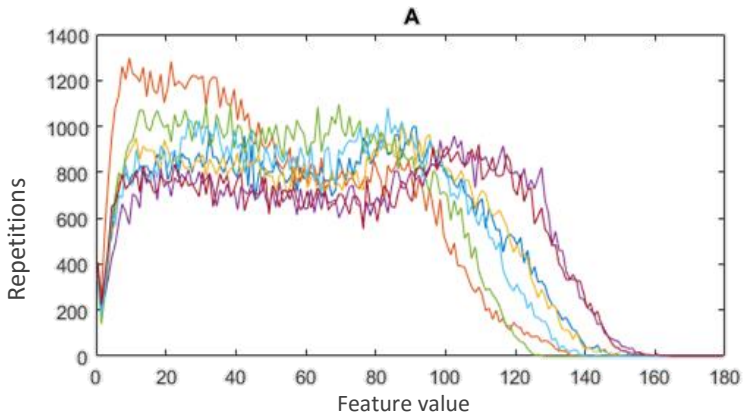


Figure 58. Histograms Obtained from Depth Images of the Dataset 3 representing feature A.

In conclusion, the combination which offered the best results is the subtraction comparison (accumulative sum of differences) of A feature histograms obtained from the depth image. The probability to identify two different subjects as the same subject (nearly 24%) can be considered too high to perform subject recognition. Nevertheless, the FR rate below 5% would enable to perform a preliminary process to subject recognition with the purpose of discarding two different subjects.

7. References

[1] H. Miar-Naimi, P. Davari "A New Fast and Efficient HMM-Based Face Recognition System Using a 7-State HMM Along With SVD Coefficients" *Iranian Journal of Electrical & Electronic Engineering*, Vol. 4, Nos. 46 1 & 2, Jan. 2008.

[2] D. Agushinta, A. Suhendra, S. Madenda, S. H "Face Component Extraction Using Segmentation Method on Face Recognition System" *Journal of Emerging Trends in Computing and Information Sciences* Vol. 2, No. 2, pp.67-72, 2010-11.

[3] M. Abdullah, M. Wazzan, S. Bo-saeed "Optimizing Face Recognition Using PCA", *International Journal of Artificial Intelligence & Applications* (IJAIA), Vol.3, No.2, pp. 23-31, March 2012.

[4] Y. Lei, M. Bennamoun, A. A. El-Sallam "An efficient 3D face recognition approach based on the fusion of novel local low-level features" *Elsiver, Pattern Recognition,* Vol.46, Issue 1, pp. 24-37, January 2013.

[5] R. Jana, D. Datta, R. Saha " Age Group Estimation using Face Features" *International Journal of Engineering and Innovative Technology (IJEIT)* Vol. 3, Issue 2, pp. 130-134, August 2013.

[6] J. Prajapati, A. Patel, P. Raninga, "Facial Age Group Classification" *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE),* Vol. 9, Issue 1, Ver. II, PP 33-39, January 2014.

[7] S. Kalm, G. Guttikonda, "Gender Classification using Geometric Facial Features", *International Journal of Computer Applications (0975 – 8887)* Vol. 85 – No. 7, pp.32-37, January 2014

[8] Y. Bao, Y. Ying, L.Musa, "Face Recognition and Using Ratios of Face Features in Gender Identification" *Int'l Conf. IP, Comp. Vision, and Pattern Recognition,* IPCV'15 pp. 17-23.

[9] *http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html* - ORL Database

[10] *https://sites.google.com/a/nd.edu/public-cvrl/data-sets* - FRGC v2.0 dataset

[11] *http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html* - BU-3DFE dataset.

[12] *http://vision.ucsd.edu/content/yale-face-database* or alternatively *http://vismod.media.mit.edu/vismod/classes/mas622-00/datasets* - Yale Database

[13] *http://cswww.essex.ac.uk/mv/allfaces/* - Dr Libor Spacek Databases: Faces94, Faces 95, Faces 96, Grimace database.

[14] *http://www.mathworks.com/matlabcentral/fileexchange/34792-circlefit3d-fit-circle-to-three-points-in-3d-space* - Fit 3D circle to three points in 3D space