Decision Tree Approach to Facial Image Retrieval from Databases

A. A. Abayomi-Alli
Sound System and Structures Lab, University of Pittsburgh, USA

O. O. Abayomi-Alli
Dept. of Computer Science, FUNAAB, Abeokuta, Nigeria

O. A. Adedapo
Dept. of CSE, LAUTECH, Ogbomoso, Nigeria

A. A. Sijuade
Dept. of CSE, LAUTECH, Ogbomoso, Nigeria

ABSTRACT
Existing face recognition systems are faced with several image quality variations due to pose, illumination, age, expression, occlusion, etc. These variations have reduced biometric system performance during real time deployment especially in applications of large databases with reduction in recognition accuracy and increased computational cost. This study proposed a decision tree approach for facial image retrieval from large databases. A top-down approach was adopted for the design of the decision tree, namely image selection, distance measurements, clustering, level of impurity and information gain. The SCface surveillance camera database of 4,130 images from 130 subjects was employed and divided into training and testing datasets respectively. The decision tree was designed based on the information gain calculated using the pixel coordinates of the face geometric features and the K-means method was used for cluster analysis. The system was successful implemented using Microsoft C# and appropriate user interfaces were incorporated. The performance of the system in retrieving 879 images from the test dataset shows a confusion matrix with 863 true positives (TP), 3 false positive (FP), 24 true negative (TN), 7 false negative (FN) and over-all accuracy of 0.9889. The system performance was satisfactory and the study concludes with recommendation for future studies.

General Terms
Face Recognition, Machine learning and Decision tree.

Keywords
Algorithm, Clustering, Data-mining, Decision trees, Image retrieval,

1. INTRODUCTION
Facial recognition has been an active area of research in computer vision and psychophysics, over the past decade [1], [2], [3], [4], [5]. Unlike other biometric person-identification methods such as fingerprint analysis, retinal or iris scans, face recognition systems does not rely heavily on the co-operation of the participants [6]. Amongst all these biometrics, face is more attractive as it provides information such as identity, expression, gender and age of an individual. This has led to the demand of a facial recognition system that is resilient to the face lighting conditions and different variations like pose, facial expression or occlusions like spectacles or beard [7].

While the technology for mining text documents in large databases could be said to be relatively mature, the same cannot be said for mining other important data types such as speech, music, images and video. Yet these forms of multimedia data are becoming increasingly prevalent on the internet and intranets as bandwidth rapidly increases due to continuing advances in computing hardware and consumer demand. An emerging major problem is the lack of accurate and efficient tools to query these data directly, so we are usually forced to rely on available metadata such as manual labeling. Currently the most effective way to label data to allow for searching of multimedia archives is for humans to physically review the material. This is already uneconomic or, in an increasing number of application areas, quite impossible because these data are being collected much faster than any group of humans could meaningfully label them and the pace is accelerating, forming a veritable explosion of non-text data [8].

An ideal face recognition system should recognize new images of a known face and be insensitive to nuisance variations in image acquisition. Yet, differences between images of the same face (intraclass variation) due to these nuisance variations in image capture are often greater than those between different faces (interclass variation) making the task extremely challenging [9]. Most systems work well only with images taken under constrained or laboratory conditions where lighting, pose, and camera parameters are strictly controlled. This requirement is much too strict to be useful in many data mining situations when only few sample images are available such as in recognizing people from surveillance videos from a planetary sensor web or searching historic film archives.

The past decade has witnessed a lot of improvements in research for a reliable facial recognition system. Many techniques have been used in the performance modeling of a facial recognition system. However, problems with existing face recognition technology still include the following reduced overall accuracy in the performance of the system, Sensitivity to changes in lighting, subject distance-to-camera and pose, as well as increased computational load of searching or retrieving images.

This paper presents a decision tree data-mining approach for automatic facial image retrieval from databases with the following specific objectives: objectives of this study are to:

1) Present an updated survey of current method for constructing decision trees;

2) Build a database of face images to mimic real life surveillance scenarios;

3) Design a decision tree for facial image retrieval from large databases;

4) Implement and evaluate the performance of the decision tree.

The proposed model will have the capability of matching faces even when they are subject to variations like lighting, pose or expression variations.

2. LITERATURE REVIEW

Automated face recognition is a relatively new concept. Developed in the 1960s, the semi-automated system for face recognition required the administrator to locate features (such as eyes, ears, nose and mouth) on the photograph before it calculated distance and ratios to a common reference point, which were then compared to referenced data. In the 1970s [10] used 21 specific subjective markers such as hair color and lip thickness to automate the recognition. The problem with both of these early solutions was that the measurements and locations were manually computed.

[11] Applied Principal Component Analysis (PCA), a standard linear algebra technique, to the face recognition problem. This was considered somewhat of a milestone as it showed that less than one hundred values were required to accurately code a suitably aligned and normalized face image [11]. [12] Discovered that while using the Eigen faces technique, the residual error could be used to detect face in images. This discovery enabled reliable real time automated face recognition system. Although the approach was somewhat constrained by environmental factors, it nonetheless created significant interest in furthering development of automated face recognition technology [13]. The technology first captured the public's attention from the media reaction to a trial implementation at the January 2001 Super Bowl, which captured surveillance images and compared them to a database of digital mug shots. This demonstration initiated much needed analysis on how to use the technology to support national needs while being considerate of the public's social and privacy concerns.

Today, face recognition technology is being used to combat passport fraud, support law enforcement, identify missing children, and minimize benefit/identify fraud. There are already some work on performance modeling and prediction of biometric systems, such as fingerprint recognition [14], iris recognition [15], and face recognition [16]. The quality of a fingerprint image is defined as the normalized distance between matching and non-matching similarity scores [14]. An 11-dimensional feature vector is extracted from image analysis algorithms to identify the existence of feature points, e.g., minutia, and outliers. Then a Neural Network is trained using the feature vectors to predict the image quality. The experiments show that the images with higher predicted quality will achieve better recognition accuracy. The feature extraction method for fingerprint image quality prediction cannot be directly used to face recognition since most face recognition methods use holistic appearance instead of feature points.

[15] Provide a probabilistic estimation of lower bound of iris recognition algorithms based on analysis of the hamming distance between query and gallery iris images. The distance is assumed to be a single Gaussian distribution under both genuine and imposter hypothesis and the likelihood ratio is used to identify the pattern that best match the query iris. With learned parameters, the method estimates receiver operating characteristics (ROC) curve of iris recognition by applying the Chernoff bound theory and the large deviation theory. However, both of the lower bounds only provide approximate error orders. They cannot be used to predict either an individual recognition result, or the performance of systems which do not use likelihood ratio method for recognition.

[16] Applied statistical tools to analyze how the human face features, such as age, race, gender, skin, glasses, and expression, affect face recognition accuracy. A generalized linear model was built to regress the relationship between the affecting factors and recognition accuracy. The analysis of variance (ANOVA) is conducted to study how significantly each factor affects the recognition accuracy. To model the performance, the statistical model needs to explicitly identify each affecting factor, which is an extremely difficult task in practical implementation. Also, such factors cannot totally model the face recognition performance. It is shown that about 34% of variance cannot be explained by the generalized liner model.

3. DECISION TREE

A Decision Tree is a tree-structured plan of a set of attributes to test in order to predict the output. Decision tree learning is a method commonly used in data mining [17], [18] [19]. A decision tree is a popular classification method that results in flow-chart like structure where each node denotes a test on an attribute value and each branch represents an outcome of the test. The tree leaves represents the classes [20]. The goal is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning [17]. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions.

3.1 Advantages of Decision Trees

Amongst other data mining methods, decision trees have various advantages such as:

1) It's simplicity to understand and interpret. People are able to understand decision tree models after a brief explanation.

2) It requires little data pre-processing. Other techniques often require data or score normalization.

3) It's ability to handle both numerical and categorical data. Other techniques are usually specialized in analyzing datasets that have only one type of variable.

3.2 Decision tree algorithm

Decision tree algorithm is a relatively simple top-down greedy algorithm. The aim of the algorithm is to build a tree that has leaves that are as homogenous as possible [19]. The major step of the algorithm is to continue to divide leaves that are not homogenous into leaves that are homogenous as possible until no further division is possible [20]. [17] Proposed a decision tree analysis in a summary of four points as:

1) Select significant independent variables;

2) Identify category groupings or interval breaks to create groups most different with respect to the dependent variable;

3) Select as the primary independent variable the one identifying groups with the most different values of the dependent variable;

4) Select additional variables to extend each branch if there are further significant differences.

While [21] gives a simple decision tree algorithm as:

1. ***For*** each attribute ***do***
2. Calculate information gain with this attribute
3. ***End*** for
4. Choose attribute with the highest information gain
5. Create a decision node and divide the training collection

Repeat procedure till nodes encounter no more examples.

4. RESEARCH METHODOLOGY

The methodology for building our facial image retrieval system using decision tree will involve different sections that will fulfill the studies main objectives.

4.1 Image Database

[22] SCface Surveillance Camera face database was used for the implementation and computational processes involved in this study. The database contains 4,160 static images of 130 subjects, taken in uncontrolled lighting conditions with different cameras. For the capturing stage, 6 different surveillance cameras with five different qualities were used; also a digital camera of high quality was used for the mug shots images. Two of the cameras were able to operate in infra-red night vision mode. The database was suitable for this study based on the following features:

1) Different quality and resolution cameras were used;

2) Images were taken under uncontrolled illumination conditions;

3) Images were taken from various distances;

4) Head pose in surveillance images is typical for a commercial surveillance system, i.e. the camera is placed slightly above the subject's head, making the recognition even more demanding; Besides, during the surveillance camera recordings the individuals were not looking to a fixed point;

5) Database contains nine different poses images suitable for head pose modeling and/or estimation;

6) Database contains images of 130 subjects, enough to eliminate performance results obtained by pure coincidence (probability of recognition being a coincidence is less than $1/130 \approx 0.8\%$).

The actual system design is divided into six steps namely image selection, distance measurement, clustering, impurity level, information gain and decision tree.

4.2 Image Selection

To build the decision tree, the face database was divided into training and testing dataset of ratio 70:30 percent respectively. Thus 23 images from 91 subjects were used for training while images from the remaining 31 subjects were used for system testing. A program was written in C# to automatically extract the geometric features of the facial images such as the pixel coordinates of the left and right eyes, tip of their nose and the centre of the mouth, presence of beard, glasses, moustache and facial marks.

4.3 Distance measurement

The next step after image selection is to calculate the distance between the face features given above using their coordinates. Distances will be calculated for eye-to-eye, eye-to-nose, nose-to-mouth, and eye-to-mouth. The sample outputs are presented on Table 1.

The coordinates were calculated using the formulas below:

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \quad (1)$$

The midpoint for the left eye and right eye as

$$\left(\frac{x1+x2}{2}, \frac{y1+y2}{2}\right) \quad (2)$$

Table 1. Distance measurement output

Sub	Dis	Cam	eye-to-eye (a)	eye-to-nose (b)	nose-to-mouth (c)	eye-to-mouth (d)
054	F	F	353	166	209	375
054	1	7	10	8	3	11
054	2	2	19	14	9	23
054	3	3	29	22	15	37
105	F	F	322	147	168	315
086	3	2	29	21	10	31
078	1	5	11	8	6	14
039	F	F	353	166	209	375
039	3	1	49	20	14	34

4.4 Clustering

Clustering is a machine learning technique used to place data elements into related groups without advance knowledge of the group definitions [23]. The clustering technique used in this study is the K-means clustering [24]. K-means clustering is a method of cluster analysis which aims to partition N observations into k clusters in which each observation belongs to the cluster with the nearest mean. The basic step of k-means clustering is simple [25]. Here's how the algorithm works:

1) The algorithm arbitrarily selects k points as the initial cluster centers ("means");

2) Each point in the dataset is assigned to the closed cluster, based upon the Euclidean distance between each point and each cluster centre;

3) Each cluster centre is recomputed as the average of the points in that cluster;

4) Repeat Steps 2 and 3 until the clusters converges.

Convergence may be defined differently depending upon the implementation, but it normally means that either no observations change clusters when steps 2 and 3 are repeated or that the changes do not make a material difference in the definition of the clusters. Table 2 shows the output of the clustered data.

Table 2. Output of clustered data

Sub	Dis	Cam	Eye-to-eye (a)	Eye-to-nose (b)	Nose-to-mouth (c)	Eye-to-mouth (d)
054	F		A3	B3	C3	D3
054	1	Cam3	A1	B1	C1	D1
054	2	Cam1	A1	B2	C1	D2
054	3	Cam1	A2	B2	C1	D2
105	F		A3	B3	C3	D3
086	3	Cam1	A2	B2	C1	D2
078	1	Cam2	A1	B1	C1	D1
039	F		A3	B3	C3	D3
039	3	Cam1	A2	B2	C2	D2

4.5 Impurity Level

The impurity level was calculated using classification error.

Classification error = 1 – {P (max)} (3)

Each subject has 23 images each and there are 91 subjects in the training dataset.

Classification error = 1– (23/(23*91)) = 0.9809.

Therefore, Impurity level is $\cong$ 0.9.

4.6 Information Gain

The decision tree was built based on the information gain of all the attributes in the table above. The measure to compare the difference of impurity degrees is called information gain. Information gain is computed as impurity degrees of the parent table and weighted summation of impurity degrees of the subset table. The weight is based on the number of records for each attribute values.

Information gain (i) = Entropy of parent table - Sum (n k /n * Entropy of each value k of subset table Si).

Table 3 shows the information gain for the entire attribute.

Table 3. Information gain

Attribute	Information gain
Beard	0.102
Moustache	0.102
Glasses	0.1
Eye-to-eye	0.05
Eye-to-nose	0.05
Eye-to-mouth	0.06
Nose-to-mouth	0.02
Distance	0
Camera	0.016
Gender	0.1

4.7 Decision Trees Chart

The decision tree that was drawn based on the information gain in the table above. The information gain table is to compare the measure of impurity between the attributes. Without the information gain table, there will be no decision tree. Figure 1 shows the designed decision tree with its nodes.

5. TESTING AND RESULTS OBTAINED

The facial image retrieval system from large databases was developed based on the decision tree model in figure 1. When an image is selected for as testing image, the system starts by searching bearded/non-bearded images in the database before proceeding to images with glasses/those without glasses, after these 2 operations, some images will be filtered based on these features. Then the system considers subjects coordinates of eye-to-mouth with error limit of ±3. The system filters some images or there will be no match. If there is a/are match(es), it proceeds to subjects coordinates with eye-to-nose and error limit of ±3 then to nose-to-mouth, after these processes it consider camera and filter subjects within the range of the cameras used then it check distances. Since distance is the least on the information gain table. The final search is done and the system output the recognized subjects. Figure 2 shows the query interface for image retrieval from the database while Figure 3 shows the result of the output interface showing recognized subjects. After entering the image query data the recognize button is clicked and result is shown on Figure 3. The choice of the best match depends on the user. A confusion matrix was developed to evaluate the performance of the system in retrieving the 879 images of 39 subjects in the testing dataset. The aggregate confusion matrix as shown in table 4 presents the performance of the decision tree algorithm for facial image retrieval in term of the correctly recognized faces and the confused faces (faces recognized as a subject different from the actual person) as obtained from the testing experiments. The aggregate confusion matrix is obtained by finding the average of all the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) is presented on table 4.

Table 4: Aggregate confusion matrix showing the performance of the decision tree algorithm for facial image retrieval

TP=863	FN=7
FP=3	TN=24

Accuracy can be defined as:

$$acc = TP + TN/TP + TN + FP + FN \quad (4)$$

Thus from Table 5,

$$acc = 863 + 24/863 + 24 + 3 + 7 = 0.9889$$

$$acc \cong 0.99$$

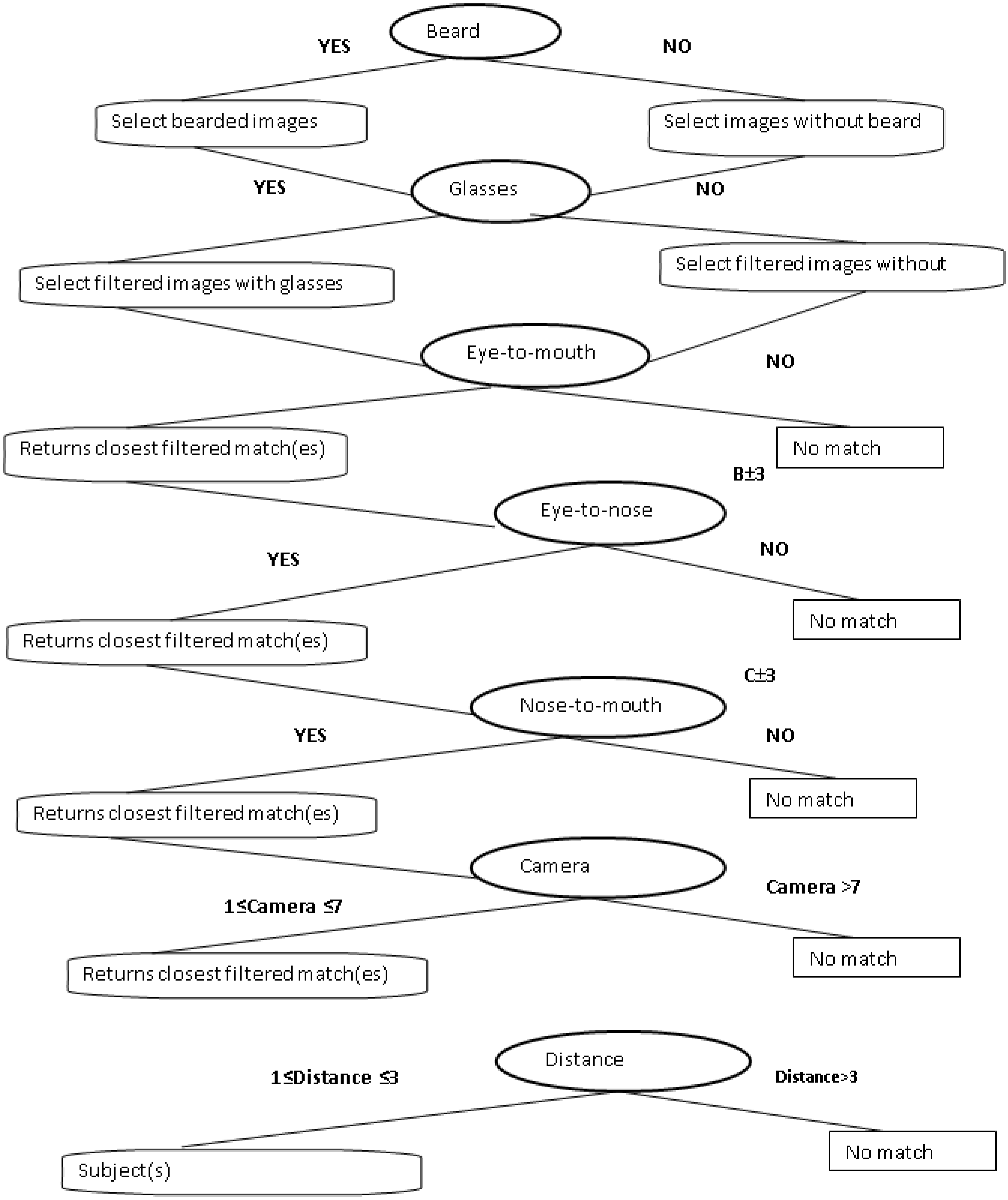

Fig. 1: Proposed decision tree with its node

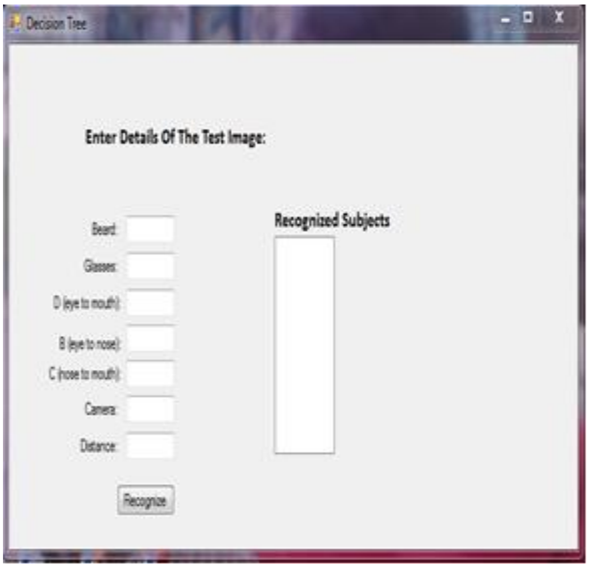

Fig. 2: Query interface for facial image retrieval

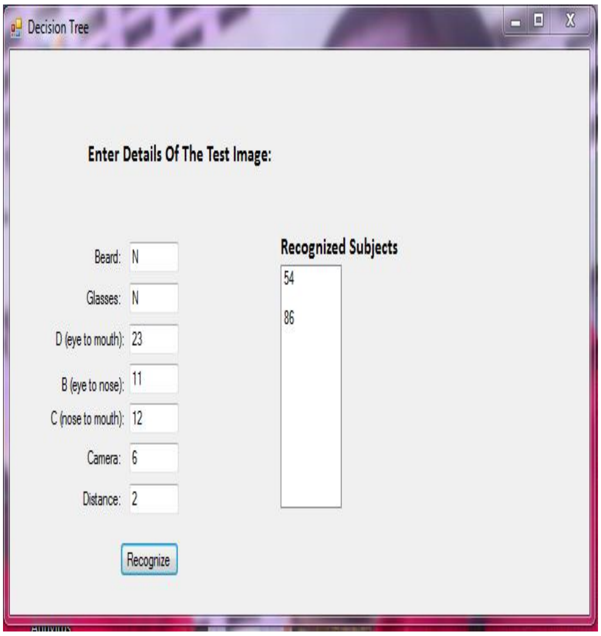

Fig. 3: Output interface showing recognized subjects

6. CONCLUSION

This paper has proposed a decision tree based facial image retrieval system from large databases. The system made use the geometric features of the face. The decision tree was constructed using the coordinates and features of the images in the database; the tree was constructed based on the information gain generated. The system was implemented using C# programming language. Results from the performance experiments shows that our proposed model is able to perform facial image retrieval from large databases with an accuracy of 0.99 when used to retrieve 897 images of 31 subjects in the testing dataset. Although the recognition performance of the system is satisfactory, it can further be improved with some modifications such as (1) modelling the recognition accuracy of retrieved images (2) use larger datasets could be used for the training and testing with different factors affecting performance of face (3) the output of the recognition phase should reveal images and not subject number.

7. ACKNOWLEDGMENTS

This research work is sponsored in part by the Federal Government of Nigeria through the TETFUND academic staff training and development (ASTD) fellowship of the Federal University of Agriculture, Abeokuta, Nigeria.

The authors wish to thank Professors Mislav Grgic, Kresimir Delac and Sonja Grgic for the release of the SCFace-surveillance cameras face database.

8. REFERENCES

[1] Bolme, D., Beveridge, R., Teixeira, M. and Draper, B. 2003. “The CSU Face Identification Evaluation System: It’s Purpose, Features and Structures”, International Conference on Visions Systems, Springer-Verlag, Graz, Austria, pp. 304-311.

[2] Wang, P. and Ji, Q. 2006. “Performance modeling and prediction of face recognition system”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York City, pp. 1-3.

[3] Woodward, J. D., Horn, C., Gatune, J. and Thomas, A. 2003. “Biometrics a look at facial recognition”, RAND Public Safety and Justice for the Virgina State Crime Commission, Santa Monica, pp. 1-7.

[4] Introna, L. D. and Nissenbaum, H. 2006. “Face Recognition Technology”, Department of Media, Culture and Communication, New York University, pp. 10-14.

[5] Nagi, J. 2007. “Design of an Efficient High-speed Face Recognition System”, Department of Electrical and Electronics Engineering, Universiti Tenaga Nasional, March 2007.

[6] Ramanathan, N., Chellapa, R. and Roy-Chowdhury, A. K. 2003. “Facial similarity across age, disguise, illumination and pose,” Centre for Automation research, Dept. of Electrical and Computer Engineering, University of Maryland, pp. 1.

[7] Bhattacharya, J., Lakshmi-Prabha, N. S., Majumder, S. and Sanyal, G. 2012. “A novel face recognition approach using a multimodal feature vector”, Central Mechanical Engineering Research Institute, India, 3(1), pp. 477-486.

[8] Lovell, B. C. and Chen, S. 2003. “Robust face recognition for data mining”, EMI, ITEE, University of Queensland, Australia, pp. 1-3.

[9] Adeniji, Y., Moses, Y. and Ullman, S. 1997. “face recognition: The problem of compensation for changes in illumination direction”, IEEE PAMI, 19 (4), pp. 721-732.

[10] Goldstein, A. J., Harmon, L. D. and Lesk, A. B. 1971. “Identification of Human Faces”, Proc. IEEE, Vol.59 (5), pp. 748-760.

[11] Sirovich, L. and Kirby, M. 1987. “A low dimensional procedure for the characterization of human face”, J. Optical Society of America, Vol. 4(3), pp. 519-524.

[12] Turk, M. A. and Portland, A. 1991. “Facial recognition using eigenfaces”, Proc. IEEE, pp. 586-591.

[13] Lu, J., Plataniotis, K. N. and Venetsanopoulos, A. N. 2003. “Regularized Discriminant Analysis for the Small Sample Size Problem in Face Recognition”, Pattern Recognition Letters, Vol. 24 (16), pp. 3079-3087.

[14] Tabassi, E., Wilson, C. L. and Watson, C. L. 2004. “Fingerprint image quality”, Technical Report NISTIR 7151, National Institute of Standards and Technology.

[15] Schmid, N. A., Cukic, B., Ketkar, M. and Singh, H. 2005. “Performance analysis of iris based identification system at the matching score level”, ICASSP2, pp. 93-96.

[16] Givens, G., Beveridge, J. R., Draper, B. A., Grother, P. and Phillips, P. J. 2004. “How features of the human face affect recognition: a statistical comparison of three face recognition algorithms”, CVPR, Vol. 2.

[17] Thomas, E. 2004. “Data Mining: Definitions and Decision Tree Examples”, State University of New York. pp. 7-15.

[18] Fayyad U., Piatetsky-Shapiro, G. and Smyth, P. 1996. “From Data Mining to Knowledge Discovery in Database”, American Association for Artificial Intelligence. Fall, 1996. pp. 37-54.

[19] Gupta, G. K. 2006. “Introduction to Data Mining and Case Studies”, Prentice Hall of India Pvt. Ltd.

[20] Mitra, S., Parua, S., Das, A. and Mazundar, D. 2007. “A novel data mining approach for performance improvement of EBGM based face recognition engine to handle large database”, CDAC, Kolkata, India, pp. 4-6.

[21] Ghosal, V. 2009. “Efficient face recognition system using random forests”, India Institute of Technology, Kanpur, pp. 36-38.

[22] Grgic, M., Delac, K. and Grgic, S. 2011. “SCface – Surveillance cameras face database”, Multimed Tools Appl., Vol. 51, pp. 863–879. www.scface.org

[23] Ramageri, B. M. 2006. “Data mining techniques and applications”, Indian Journal of Computer Science and Engineering. 1(4), pp. 301-305.

[24] Bezdec, J. C. 1981. “Pattern recognition with fuzzy objective function algorithms”, Plenum Press, New York, USA.

[25] Chapple. M. “K-means clustering-data mining with the k-means algorithm”, New York Times Company, [Online] http://www.about.com, Accessed December 19, 2013.