FIDA: Face Recognition using Descriptive Input Semantics

Nipun Bhatia, Rakshit Kumar, Samir Menon
Department of Computer Science, Stanford University.
{nipunb,rakshit,smenon}@stanford.edu

December 14, 2007

Abstract

Generic face recognition systems identify a subject by comparing the subject's image to images in an existing face database[11]. These systems are very useful in forensics for criminal identification[7] and in security for biometric authentication[8], but are constrained by the availability and quality of subject images[1]. In this paper, we propose a novel system, FIDA, that uses descriptive non-visual human input of facial features to perform face recognition without the need for a reference image for comparison. FIDA maps images in an existing database to a fourteen dimensional descriptive feature space using softmax regression, and compares input feature descriptions to images in feature space. FIDA clusters database images in feature space using feature-weighted K-Means clustering[4] to offer computational speedup while searching feature space for matching images.

1 Introduction

The face recognition problem involves searching an existing face database for a face, given a description of the face as an input. The face identification problem is one of accepting or rejecting a person's claimed identity by searching an existing face database to validate input data[8]. Both are well studied problems in the computer vision community and have been tackled with a variety of approaches[11]. Many databases for face identification and recognition have been built and are now widely used[3]. However, most systems that have been developed in the past are constrained by images being the primary, and often singular, form of input data[11]. In cases where images are not available as sample input, it not possible for such systems to perform face recognition. Our system, FIDA, uses general facial descriptions as input to retrieve images from a database. Users may utilize FIDA to identify images by just entering general descriptions, removing the constraint of input images for face recognition and identification purposes.

FIDA formalizes subjective human descriptions into discrete feature values and associates seven descriptive and seven geometric features to face images. A softmax classifier maps geometric facial features from images in our database to a descriptive variables. The seven discretized geometric features combine with the seven descriptive features to form a composite fourteen dimensional feature set for FIDA. Similar images are clustered in feature space using weighted K-means clustering[4]. User input, in the form of facial descriptions, directly maps to the fourteen dimensional descriptive feature space. Thereafter, the input description is compared to the three closest clusters of images in feature space iteratively, to check for matches. A set of prospective matches is then identified and returned.

Jain et al.[5] suggest that it is questionable whether the geometry of the face itself, without any contextual information, is a sufficient basis for recognizing a person from a large number of identities with an extremely high level of confidence. This is reinforced by Sinha et al.[9] who suggest that humans are good at recognizing degraded images because of their holistic processing of visual input. To ensure consistent results with FIDA, we chose a large holistic facial feature set to model aspects of facial geometry as well as descriptive facial information. Evaluating the consistency of human input for different facial features allowed us to qualitatively compare different features. Later, while clustering this information was used to apply large weights to features humans were consistent at.

Tong et al.[10] use semantic relationships for recognizing facial action units(AU). They tag a list of AUs with their interpretations and use them as training data for a learning mechanism based on Gabor feature representation and AdaBoost classification. Once trained, they use the classifier to assign semantic relationships to their corpus of data. We use a similar approach and label a set of training images with semantic feature descriptions. We then use the labeled images to train a multinomial softmax classifier to translate numeric geometric ratios of facial features to descriptors. Our system then uses descriptive semantic inputs to retrieve images.

We also explored research on the real world application of criminal identification. Most of the controversy in the area of face recognition has focused upon the suggestiveness of the observation of a single subject by an observer. Some researchers have found that single subject identification procedures result in more false identifications than lineups[2]. This suggested to us that it is more practical for us to show a lineup of prospective images in order to achieve better identification rates. FIDA displays the best three matches for any user input.

2 Descriptive Input Semantics

Our approach draws inspiration from the fact that humans describe faces using abstract and often subjective feature measures such as the shape of a face, the color of the skin, hair color etc.[9]. These semantic descriptions, supplied by humans are immune to picture quality and other effects that reduce the efficiency of contemporary face recognition and identification algorithms. We drew upon work that has been done to try to identify possible facial features that may lead to better recognition[7] while coming to our present feature set.

We used the AR Database [6] for our data set. Sample images from the AR Database may be seen in Figure.1. An example of a sampled image may be seen in Figure.2.

Figure 1: Sample faces in the AR Database

Figure 2: Sampling features from an image

We normalize facial geometric features using reference facial geometry obtained from the images in our database. This is required for accurate identification and recognition because the size of features of the face like nose, eyes, lips vary from person to person[8].

FIDA's set of fourteen descriptive features is:

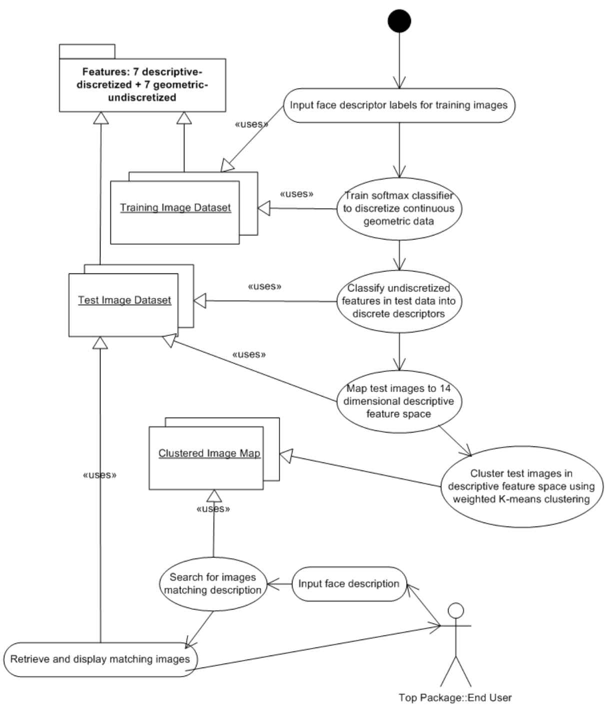
1. Sex: {Male, Female}
2. Ratio of Length of Nose to Length of the Face: {Short , Normal, Long}
3. Ratio of Width of Nose to Width of the Face: {Short, Normal, Wide}
4. Ratio of Length of Forehead to Length of the Face: {Short,Normal, Long}
5. Ratio of Width of Lips to Width of Face: {Narrow, Normal, Wide}
6. Ratio of Distance between eyes to End to end distance of eyes: {Close-set,Normal}
7. Ratio of Height of Eye to Width of Eye: {Narrow, Normal, Wide Open}
8. Ratio of Width of the Face to the Height of Face: {Round, Normal, Long}
9. Eye Colour: {Black, Brown, Blue, Green, Gray}
10. Length of Hair: {Balding/Bald, Close Crop, Normal, Shoulder Length, Long}
11. Colour of Hair: {Black, Brown, Red, Blond, White}

12. Projected Weight: {Thin, Normal, Athletic, Fat}
13. Skin: {White, Tanned White, Yellow, Brown, Black}
14. Jawline: {Round, Pear-shaped, Oval, Angular}

3 The FIDA Algorithm

The FIDA system (Figure.3) has four distinct functional components, data set preparation, discretizing geometric features to obtain descriptive feature values for all features, clustering of descriptive feature vectors of database images, and search for user-input matches in clustered feature space.

Figure 3: The FIDA Algorithm



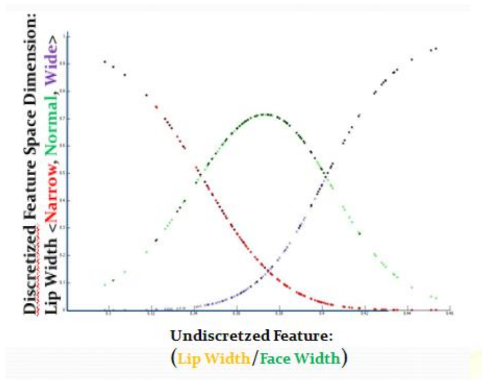
3.1 Preparing Training Data

The training images are labeled with tags such as a long nose, small face, normal ears for classification. We then extract the numeric values of the geometric ratios that define our geometric facial features by specifying control points on the image and automating data extraction.

3.2 Discretization of geometric features with softmax

A softmax classifier is trained using labeled training data to classify continuous geometric feature values into discretized descriptor values. The classifier learns the decision boundaries (Figure.4) for each of the seven undiscretized geometric features independently.

Figure 4: Softmax Classification



We then classify the discretized features of the test image set using the classifier. Merging the seven descretized geometric features with the seven descriptive features we obtain a fourteen-dimensional descriptive feature vector $\{F\}$for each image. Each value of the discretized feature vector of this set would map to descriptions like {long nose, small face, normal ears,.. }.

3.3 Clustering descriptive feature vectors

The overall performance of face recognition and identification systems used in real world applications is assessed in terms of their accuracy, speed, and storage costs[5]. Speed is important for face recognition systems and is critical for face identification systems. Given the possibility of very large image sets in areas of application, such as criminal databases, we cluster the database images in feature space. A form of weighted K-means clustering

Table 1: Results - Absolute Training and Testing Error for different samples

Training:Testing	Training(%)	Testing(%)
120:0	4.78	-
100:20	5.34	3.40
90:30	5.52	2.27
80:40	5.38	5.19
70:50	5.66	5.55
60:60	6.03	3.52
50:70	7.22	11.63
40:80	8.98	5.16
30:90	6.48	6.87
20:100	12.71	3.47

Table 2: Results - Error with unweighted and weighted K-means for different samples

Testing Images	Unweighted(%)	Weighted(%)
120	2.4	2.04
100	2.5	1.47
90	3.36	1.75
80	3.07	1.71
70	1.97	2.22
60	2.90	1.56
50	3.40	1.61
40	2.44	1.67
30	2.26	1.56
20	2.10	1.78

is used in FIDA, with weights applied to features humans are more consistent at recognizing.

3.4 Recognizing descriptive user input

We allow the user to input discretized descriptors for different features which translate to a vector $\{U\}$ of discrete variables in the fourteen dimensional descriptive feature space. The euclidean distance of U with the different cluster centroids of the database is measured and the closest three clusters are chosen for comparison. The clusters are iteratively searched by rank for the best three image matches with U. We assign an decreasing reward to the algorithm as it moves from the top three matches in each cluster, and across the three clusters. The reward starts at 1 and each move from a reported image to the next closest reported image incurs a cost of 1/9. Ie. An image match for U in the second cluster at the third slot implies three jumps in the first cluster and two in the second leading to a reward of 4/9. If the image match for to U does not lie in the three closest clusters or is not amongst the best three matches in one of the three, we report failure.

Finally, we calculate our error and return the three best matches for the given description U.

4 Results

The absolute testing and training errors for FIDA are mentioned in Table.1.

Table 3: Results - Error with one missing feature

Missing Feature	Error(%)
Sex	19.44
Eye color	7.57
Hair Length	3.75
Hair Color	3.33
Weight	4.24
Skin Color	5.00
Jawline	5.14
Nose Size	1.67
Nose Width	1.11
Forehead Length	1.94
Lip Width	1.67
Eye Position	1.46
Eye Opening	1.39
Facial Structure	1.39

The error was obtained by comparing user input across all test images to the results given by the algorithm. A match at the first position was given a reward of 1, at second, of 0.66, and at third of 0.33. We term the appearance of a match at a rank lower than three to be failure. Training error showed a consistent increase with a decrease in training samples. Testing error showed irregular movement, possibly due to the error prone nature of user input.

We compared the performance of our weights using the K-means reward metric described earlier. The results are mentioned in Table.2. These results consistently show less error with our weighted K-means algorithm and justified the feature weights we chose for FIDA.

We also compared the performance of FIDA when users were unable to provide one feature. The results are mentioned in Table.3. These results are a qualitative judge of feature quality and were used to weight the K-means algorithm. Notable is the fact that the geometric features do not affect the error as much as descriptive features do. This is because of the consistent mis-classification of these features by softmax due to highly inconsistent input data labels. This correlates with the finding of Jain et al.[5] that humans do not prefer spatial-geometric descriptors.

5 Discussion and future work

We believe that our approach could be of great use for forensic face recognition and criminal identification systems which require descriptive input semantics, since the available data often consists of witness' descriptions. In addition, our method of searching for data using descriptive semantics could combine with existing automated face recognition systems and augment them.

Adler et al.[1] concluded in 2006 that humans effectively utilize contextual information while recognizing faces, and in general equal or outperform even the best automated systems. Extensions to our work could include the annotation of contextual data to images using the descriptive semantic method. This could help improve our face recognition method by obtaining qualitatively better user input as well as improving our recognition performance

In general, the use of descriptive input features allows for input data to bear different semantics than the data being searched for. We believe that this could yield good results for other data types as well, specially where direct pattern recognition is either infeasible or yields unsatisfactory results.

References

[1] A. Adler and M.E. Schuckers. Comparing human and automatic face recognition performance. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 37(5):1248–1255, Oct. 2007.

[2] Sherrie L. Davey Bruce W. Behrman. Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, 25, Issue - 5:475–491, 2001.

[3] Ralph Gross. Face databases. February 2005.

[4] Wong M. A. Hartigan J. A. A k-means clustering algorithm. applied statistics,. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 28:100–108, 1979.

[5] Anil Jain, Lin Hong, and Sharath Pankanti. Biometric identification. *Commun. ACM*, 43(2):90–98, 2000.

[6] A.M. Martinez and R. Benavente. The ar face database. *CVC Technical Report*, 24, 1998.

[7] Sinjini Mitra, Nicole Lazar, and Yanxi Liu. Understanding the role of facial asymmetry in human face identification. *Statistics and Computing*, 17:57 – 70, January 2007.

[8] Kosuke Sato Motonori Doi, Qian Chen and Kunihiro Chihara. Lock-control system using face identification. *Lecture Notes in Computer Science*, 1206/1997:361 – 368, April 2006.

[9] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, Nov. 2006.

[10] Yan Tong, Wenhui Liao, and Qiang Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1683–1699, Oct. 2007.

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.