Local Feature Based Face Recognition
Computational Vision Coursework

Muhammad Asad
Reg No. 110128953
University Of Sheffield
masad1@sheffield.ac.uk

I. Objective

The objective of this project is to develop a face recognition system based on local features. The local features used are Scale Invariant Feature Transform (SIFT) features. The given image dataset is divided into training and testing sets. These sets are used for model construction and classification respectively. The results are presented along with the algorithms used.

II. Introduction

Face recognition has become one of the most successful applications of Computer Vision. This is due to the fact that it is a basic human gesture which makes everyday tasks simple and secure. Due to this attractiveness of human face recognition, it has been one of the most active research areas in the past decade with some major developments. This can be seen from the fact that there are numerous conferences and commercially available applications of face recognition emerging in many fields [1]. In particular there are two main application domains of face recognition: law enforcement and commercial applications, in which face recognition is used for identification and/or verification[2].

Even with some major advancements in the subject area, face recognition is still a challenging problem, as it is very difficult to classify an object if the image is occluded, changed in scale, viewed from different angle, has variations in background and is rotated or deformed. We humans, on the other hand, achieve the same task with high accuracy and our responses are so quick that we do not even realize how complex this task of face recognition is. Numerous attempts have been made to develop face recognition techniques to match the accuracy and speed of human doing the same task. Based on the approaches used, all these developed techniques can be classified into three different categories [1].

1) **Hollistic Matching Methods:**
   These methods use the whole face region image as input which is used to construct a quotient image. The input image is then compared with these quotient images to recognize the face. One of the most widely used hollistics matching methods is the EigenPictures approach [3], based on principal component analysis, which was further researched to form EigenFaces[4] and similar techniques.
2) **Feature-based Matching Methods:**
   In these methods all the local features of the face are collected e.g. nose, mouth, forehead etc. The location and local statistics of these features are then used for facial recognition. Gabor feature classifier [5] approach has been widely used for development of such Feature-Based Matching methods [6] [7].
3) **Hybrid Methods:**
   These methods combine both local features and the hollistics approach to perform face recognition. These methods are similar as that used by humans. There are number of examples for this approach as it is discussed that this approach can use the advantages of both previous approaches to develop a highly accurate facial recognition approach [8]. Many of these techniques are inspired by biological face recognition [9] [10].

In this project I use SIFT Features to do Face Recognition. The SIFT features transforms an image data into scale-invariant coordinates relative to local features. These keypoint features are invariant to scale, rotation, noise, background clutter, occlusion and partially invariant to changes in illumination and viewpoint. This approach as described in [11] is divided into four major steps, with each of these steps being crucial in the extraction of these scale invariant keypoints. These steps are briefly discussed below:

1) **Scale-space extrema detection:**
   In this stage the difference-of-Guassian approach is used for the identification of potential points of interest that are invariant to scale and orientation.
2) **Keypoint Localization:**
   Points from the previous step are used in this step. Their location and scale is determined and they are selected for the next step based on their stability.
3) **Orientation Assignment:**
   Based on the image gradient directions, each keypoint from the output of the previous step is assigned an orientation. This gives an extra feature which is used, along with the features detected in the previous steps, to do matching operations for face recognition.
4) **Keypoint Descriptor:**

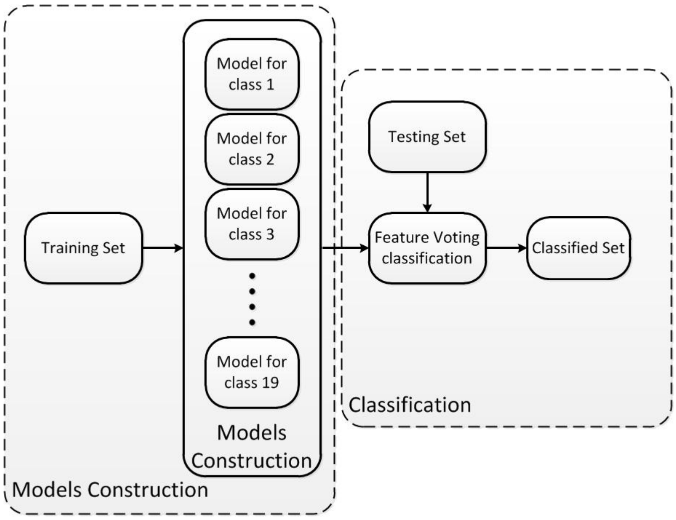


Figure 1: Block diagram of the proposed method

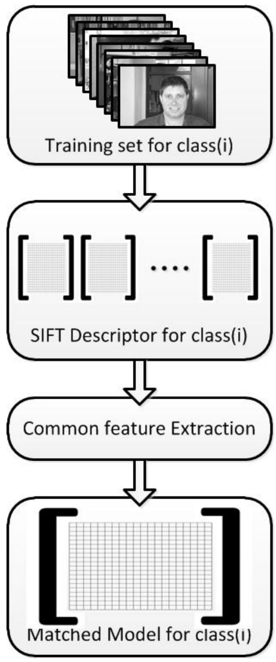


Figure 2: Construction of models

> Local gradients around each keypoint are transformed into a representation which provides feature matching with significant levels of local shape distortion and changes in illumination.

I am using an image dataset containing 408 images of 19 different person with variations in illumination conditions, background clutter, partial face occlusion and slight variations in viewpoints. SIFT features extraction is done using binaries by D.Lowe[11]. These extracted features are then used to construct models for each person. Given an unknown test image, feature voting is done against all the models. The test image is classified to be that person whose model has the highest number of votes in feature voting. Experimental results show that the proposed feature voting technique correctly classified images in the training dataset with an average accuracy of 92.55 percent. It is noted that occlusion and illumination effects the classification method of feature voting. This report first presents the proposed methods for model construction and classification in section III. The experimental results are presented and discussed in section IV and finally section V concludes this report.

III. Methodology

Training dataset consists of 13 images of each person, whereas the remaining images (different number for different person) are used to make a testing dataset. SIFT features for all the images in the dataset are extracted. These features are then used to perform face recognition using the method described below.

The proposed method uses a two step approach for face recognition as shown in Figure 1. The first step constructs models from training dataset and the second step uses these models to perform classification of images in testing dataset. These steps are explained below:

A. *Model Construction*

Main steps involved in model construction for a given person are shown in Figure 2. The extracted SIFT features from training set of a given person are matched with each other. These common matched features are then combined to construct a model for that person. This process is repeated for all the persons, giving 19 distinct models with these common matched SIFT features.

This technique is based on the fact that given an image dataset for the same person, the only thing common in the dataset are the facial features. The cluttered background features are different for different image backgrounds. Therefore these models have large number of facial features and very less background features. This whole process of common feature matching is shown in Figure 3. The red circle represents the facial features of a person in an image, whereas the yellow features represent the SIFT features from the background. Here it is explained that my models have a lot of facial features as these are all common in all those images.

Features matching is done using the approach described

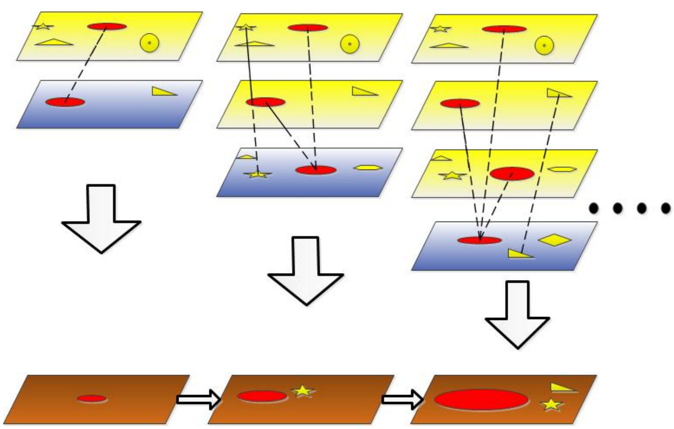


Figure 3: Common feature extraction

in [11]. Each feature keypoint is actually a vector in higher dimensions. This vector is first normalized to make a unit vector using the following formula:

$$\widehat{a} = \frac{a}{\|a\|} \tag{1}$$

These unit vectors are then compared using nearest neighbour cosine distance technique as described in [11]. The formula used for the calculation of cosine distance is given below:

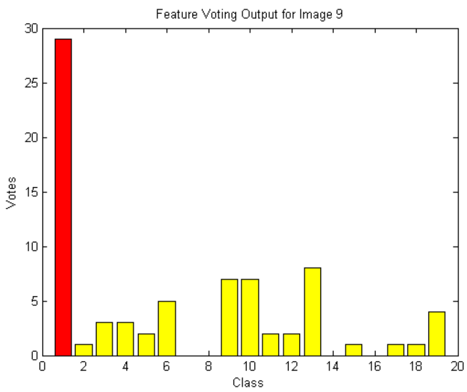
$$\cos\Theta = \frac{a \cdot b}{\|a\|\,\|b\|} \tag{2}$$

In order to remove the matches that arrive due to background clutter, I use the distance ratio of first nearest neighbour to second nearest neighbour. This ratio is called distRatio and a threshold of 0.6 is used [11].

The above process is repeated for all the persons in the database, and individual models for each person are constructed using database of matched keypoints from within each person's own image dataset. This image model is used in next step to classify the images in testing set.

B. *Classification*

The block diagram for classification is given in Figure 1. In a given dataset of different classes, classification is the process of identifying which image contains which class. Humans do classification of objects even though they are occluded, changed in size or shape, looked upon from different angles, changes in background and rotated or deformed.



(a) Feature Voting output for Image 9

(b) Matched features in Image 9

Figure 4: Feature Voting for Person 1

Classification step is used to evaluate the performance of the model using the testing set of images. The algorithm used in this step is based on feature voting technique as shown in Figure 4. Given a test images, each of its SIFT features are then matched with the models of all 19 persons in the database. Number of feature matches for each are recorded. The model with maximum feature votes is used to classify the test image. The dataset used contained 161 test images which were tested one by one.

In the Figure 4(b) red circles represent features which give rise to maximum feature votes. The rest of the matched features are shown as yellow asterisks. Feature votes are shown in Figure 4(a). Using my algorithm, this image is correctly matched with person 1. This can be seen by the huge gap between the maximum number of votes to votes for other models. Observing the location of these matches, most of the correct matches are located on the significant features of the face.

IV. Results and discussion

All of the images in testing dataset are classified, and the results are presented in a Confusion Matrix(Figure 5). The average accuracy of this algorithm for the given dataset is 92.55%. The accuracies for individual classes are shown in Table 6. The Table shows that there are some misclassifications for few classes. These misclassifications arise from a number of different factors, which are discussed below.

		Predicted Class																		
	Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Actual Class	1	6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
	2	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
	13	0	0	0	1	0	0	0	0	0	0	0	2	4	0	0	0	0	0	0
	14	0	0	0	0	1	0	0	0	1	0	1	0	2	11	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	5	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9

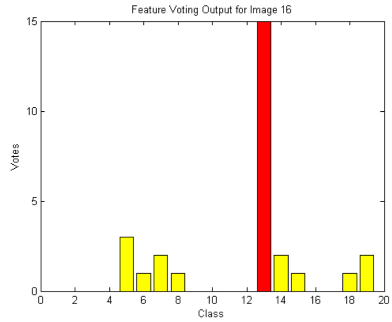
Figure 5: Confusion Matrix

From the confusion matrix in Figure 5 it can be seen that a lot of incorrect matches come from the person 14 testing dataset. This is due to the fact that this person has a lot of images with low illumination and in some of the images the face is partially occluded. There is a lot of cluttered background in almost all of the images for Person 14.

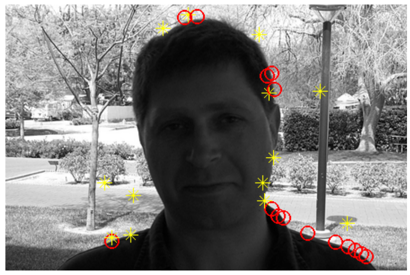
An example of one of the misclassified images for Person 14 is shown in Figure 8. This image has low illumination on face and has a highly cluttered background. It can be observed from Figure 8(c) that SIFT feature's performance is highly affected in the low illumination regions with very few features detected. It is also observed that a lot of

Person Class	Number of Total Images Tested	Number of Correctly Classified Images	Percentage Accuracy
1	8	6	75 %
2	7	7	100 %
3	9	9	100 %
4	8	8	100 %
5	10	10	100 %
6	7	7	100 %
7	8	8	100 %
8	7	7	100 %
9	8	8	100 %
10	12	12	100 %
11	9	9	100 %
12	6	6	100 %
13	7	4	57.14 %
14	16	11	68.75 %
15	7	7	100 %
16	7	5	71.43 %
17	9	9	100 %
18	7	7	100 %
19	9	9	100 %

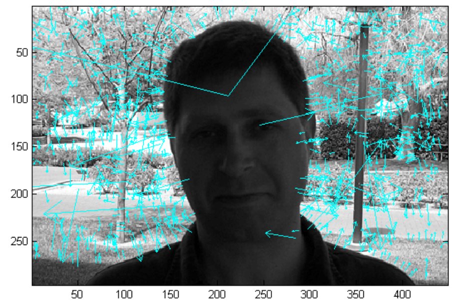
Figure 6: Accuracy Table



(a) Feature Voting output for Image 16



(b) Matched features in Image 16



(c) All of the SIFT Features detected in Image 16

Figure 7: An incorrectly classified image of Person 1

unnecessary features are detected in the highly cluttered background, which increase the amount of incorrect feature matches.

My approach fails with low illumination as in low illumination the SIFT features are not properly detected over the face region. Figure 7(c) shows an incorrectly classified image from testing dataset of person 1. As this image has low illumination on the face of the person therefore there are very less number of features detected and matched in the face region. This image also has a lot of cluttering in the background, which is responsible for incorrect feature matches.

These results also show that SIFT features approach is not invariant to significant illumination changes. Therefore an indoor and outdoor image of same person gives SIFT features which are different and does not match. This gives rise to less number of facial features matched with the model and eventually giving lesser number of feature votes. In case of partial occlusion of the face, again a lesser number of facial features are competing with other feature votes.

V. Conclusion

An algorithm for face recognition is developed using SIFT features. This algorithm produces an average accuracy of 92.55% on the given dataset of images. The SIFT features have a limitation that it is highly sensitive to illumination changes. This was the main reason that this SIFT based face recognition algorithm misclassified certain images with low illumination.

References

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[2] A. Tolba, A. El-Baz, and A. El-Harby, "Face recognition: A literature review," *International Journal of Signal Processing*, vol. 2, no. 2, pp. 88–103, 2005.

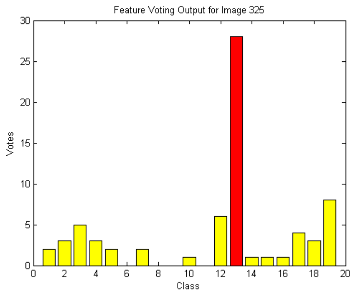
[3] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 103–108, 1990.

[4] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on.* IEEE, 1991, pp. 586–591.

[5] C. Liu and H. Wechsler, "A gabor feature classifier for face recognition," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 270–275.

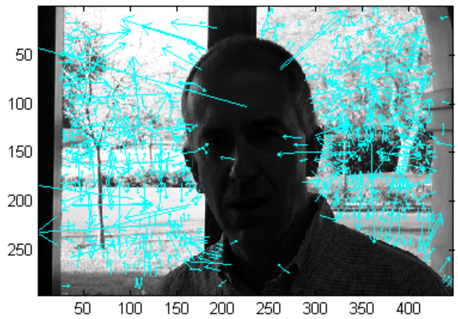
[6] M. Huang, J. Zhou, K. He, S. Xiong, and T. Li, "Local feature based face recognition," in *Computer and Computational Sciences, 2006. IMSCCS'06. First International Multi-Symposiums on*, vol. 1. IEEE, 2006, pp. 274–277.

[7] Z. Qian, P. Su, and D. Xu, "Face recognition based on local feature analysis," in *Computer Science and Computational*



(a) Feature Voting output for Image 325

(b) Matched features in Image 325



(c) All of the SIFT Features detected in Image 325

Figure 8: An incorrectly classified image of Person 14

Technology, 2008. ISCSCT'08. International Symposium on, vol. 2. IEEE, 2008, pp. 264–267.

[8] R. Huang, V. Pavlovic, and D. Metaxas, "A hybrid face recognition method using markov random fields," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3. IEEE, 2004, pp. 157–160.

[9] Y. Wang, P. Anderson, and R. Gaborski, "Face recognition using a hybrid model," in *Applied Imagery Pattern Recognition Workshop (AIPRW), 2009 IEEE*. IEEE, 2009, pp. 1–8.

[10] M. Razzak, M. Khan, and K. Alghathbar, "Bio-inspired hybrid face recognition system for small sample size and large dataset," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*. IEEE, 2010, pp. 384–388.

[11] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.