Unsupervised Sub-Graph Selection and its Application in Face Recognition Techniques

Ahmed ElSayed, Ausif Mahmood, Tarek Sobh

Department of Computer Science and Engineering, University of Bridgeport, 221 University Avenue, Bridgeport, CT, 06604

Abstract. One of the limitations of the existing face recognition algorithms is that the recognition rate significantly decreases with the increase in dataset size.In order to eliminate this shortcoming, this paper presents a new training dataset partitioning methodology to improve face recognition for large datasets. This methodology is then applied to the Eigenface algorithm as one of the algorithms that suffer from this problem. The algorithm represents the training face images as a fully connected graph. This graph is then divided into simpler sub-graphs to enhance the overall recognition rate. The sub-graphs are generated dynamically, and a comparison between different sub-graph selection techniques including minimizing edge weight sums, random selection, and maximizing sum of edge weights inside the sub-graph are provided. It is concluded that the optimized hierarchical dynamic technique increased the recognition rate by more than 40 percent in a large benchmark image dataset compared to the original single large graph method. Furthermore, the developed technique is compatible with several other unsupervised face recognition techniques such as ICA, KPCA, RBM, SIFT, and LBP... etc, and other datasets, specially if the number of images per person in the training data are low.

Keywords: Sub-Graph Selection, Graph Theory, Hierarchical Recognition, Face Recognition

1 Introduction

Identity detection is one of the important problems in the fields of security and intelligence. Face recognition is one of the computer vision fields that is charged with this task. Several face recognition algorithms have been proposed and developed in the last decades including Direct Correlation, Principal Component Analysis (PCA) [14,9], Linear Discriminant Analysis (LDA) [16,8], Independent Component Analysis (ICA) [9,3,2], Kernel methods (i.e KPCA and SVM) [15,11], and other high dimension features methods such as LBP, SIFT, or 3D methods. . . etc. Some of these methods are supervised (e.g., LDA and SVM) where the given data is divided into training, testing and validation datasets. The training dataset is then divided into labeled groups (i.e., classes) with each class containing images of one person. The other methods for face recognition are unsupervised and use extracted features from faces (e.g., PCA, KPCA, ICA, LBP and SIFT) where the given data is separated into training and testing datasets. In this case, the training dataset is unlabeled and the algorithm handles class separation process. Various research and experiments have proven that simple recognition algorithms like Eigenface produce good recognition rates with small sized (typically less than 100 images) datasets with accurately clipped face images. However, when the number of images in the dataset slightly exceeds hundred, the recognition accuracy of these algorithms reduce significantly. One of the ways to handle this problem is to use indexing [13].. However, indexing is only applicable to specific features and techniques, in addition to being very sensitive to image normalization, orientation and features calculation. The hierarchical partitioning technique can be used to improve the recognition rate when large datasets are required. This technique divides the given training dataset into smaller subgroups. This way the input image is compared with the stored images located in relatively smaller sets. The best matches are then selected and fed to the groups that are in the subsequent levels until a single small group remains. Finally, the best result from this final group determines a match or mismatch. Such hierarchical grouping principle on the training dataset has been used in [7,6]. These studies utilize a supervised grouping where the training dataset is divided according to the image class. In other words, the images of the same person are grouped together resulting in a clear separation between groups. One major drawback of this approach is that it implies the group size to be reduced to one when multiple images of the same person are not available. Furthermore, the number of groups increases significantly as the number of different individuals increase, especially when there are few images for each person in the database. The proposed unsupervised grouping technique can solve these problems since it will group images without considering identity. This work distinguishes itself from its counterparts and contributes to the related literature by:

1. Introducing a new unsupervised grouping technique for large training datasets,
2. Applying different grouping criteria in the proposed method,
3. Demonstrating the efficiency of the proposed method by providing a comparative study using multiple databases.

The remaining of the paper is divided into five sections. The following section, Section 2 explains the sub-graph selection process. This is followed by a comprehensive description of the proposed hierarchical algorithm (Section 3). Section 4 demonstrates how to further improve the recognition rate by optimizing the grouping process. Section 5 depicts the results of the proposed technique. Conclusions and future work are discussed in Section 6.

2 Sub-Graph Selection Process

The sup-graph selection process requires selecting a sub-graph k_o from a graph G that has a specific criterion. The algorithm assumes that all training face images are a fully connected graph (G) with number of nodes (L) and the edge between every two nodes w_{ij} is the sum of the Euclidean distances between the features of these the nodes i and j. The goal is to obtain the best sub-graphs set $S = \{k_1, k_2, k_3, ..., k_N\}$ where each sup-graph has a number of nodes (l), where k_o is the sub-graph number o, and N is the total number of reconstructed sub-graphs that will be used in the hierarchical technique. Different strategies for this sub-graph selection process are investigated including 1) minimizing the weight of the sum of edges within the entire sub-graph; 2) randomly choosing nodes for the sub-graph and, 3) maximizing the weight of the sum of edges within the entire sub-graph. After sub-graphs are created, regular face recognition technique (Eigenface, in this case) is applied to each fully connected sub-graph to select the top best matches from each group. These sets of matches from the first sub-graphs level form the subsequent level of sub-graphs. This process is repeated until a single small full connected graph of (l) nodes remain. This hierarchical grouping algorithm is presented and different variations are explored by testing it on benchmark datasets to prove the possible improvement in the recognition rate over full connected images graph. Figure 1 shows 2D example for different strategies for the sub-graph selection process.

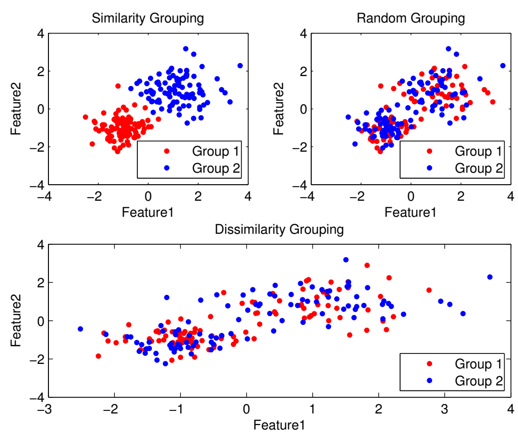

Fig. 1. 2D example for two sub-graphs selection

3 Hierarchical Recognition technique

Testing various face recognition algorithms proved that recognition rate in unsupervised algorithms such as standard Eigenface technique drops down if the number of images in the dataset is approximately above hundred. In order to understand the impact of smaller subsets generated from the entire training set, the paper proposes the following. As also detailed in the previous sections, assuming that the face images are a fully connected graph(G) with number of nodes (L) with a goal to select the best sub-graphs set $S = \{k_1, k_2, k_3, ..., k_N\}$each having a number of nodes ($l \leq 100$), where N is the number of reconstructed sub-graphs to improve the recognition rate over the hierarchical technique. With this goal, applying recognition algorithm over each of these sub-graphs (groups), a few top matched nodes from each sub-graph (group) (2 to 5) are selected. Then new groups are generated from these top matches. Depending on the number of images in the dataset, a number of hierarchical levels are created. Recognition algorithm (e.g. Eigenface) is then applied on each level group. As the final step, the top matched images from the final subgroups are collected, and recognition algorithm is re-applied on this final group to select the best-matched image. Figure 3 shows the block diagram of the proposed hierarchical technique with the Eigenface as the recognition method. The main challenge of this technique is to determine the best sub-graphs selection strategy to improve the overall face recognition rate. There are three possible grouping strategies: i) *Similarity* Grouping by minimizing the sum of weights in the entire sub-graph where similar images are added to the same group (the similarity measurement is the distance between faces features, eg. pixels gray level), this can be achieved by using regular clustering techniques, ii) *Random* Grouping by assigning the images to the groups (sub-graphs) randomly, iii) *Dissimilar* Grouping by maximizing sum of weights in the entire sub-graph, in other words, maximizing the standard deviation within the same group where the grouping process based on dissimilarity (Maximizing metric distance between faces features in the same group). An additional challenge is to obtain the suitable number of levels in the hierarchical system along with the number of matched images to be selected from each level to feed into the next level in the hierarchy. In order to achieve these, these three possibilities have been tested on a large dataset (Extended Yale B+) [1] having different positioning and illumination levels to determine the best approach for the hierarchical face recognition technique. Figure 2 shows examples of datasets images for the proposed sub-graph selection algorithm.

Fig. 2. Examples of the dataset images used

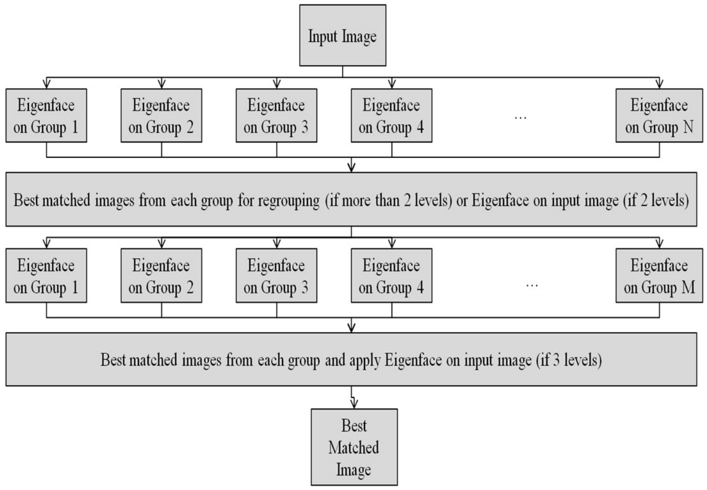

Fig. 3. The Proposed Hierarchical System for Rank 1 Recognition

4 Optimized Dissimilarity Sub-Graph Selection Technique

As also detailed in the results section, the simple dissimilarity measurement (maximizing distances between in group images by taking the mean image as a reference) performed superior compared to the other two grouping techniques (similarity and random selection). However, this method is not without some drawbacks since these criteria will not guarantee the exact dissimilarity between each group's images. To explain further, consider a 2-D set of (x,y) points. If the training dataset includes {(-2,3),(2,3),(-2,-3),(2,-3)} and is required to group these values into two groups based on dissimilarity, then the mean point will be (0,0) and the Euclidean distance between each one of these points and the total mean will be similar for all four points. This will results in poor grouping. It can easily be observed that the best dissimilarity grouping for this case would be {(-2,3), (2,-3)} as one group, and {(2,3), (-2,-3)} as the second group. Mathematically, stated as the variance between all the sub-graph (group) nodes over all basis should be maximized. Therefore applying this method to a face image dataset sub-graphs selection leads to equation (1):

$$\sigma_{total} = \sum_{l=1}^{N} \sum_{k=1}^{m \times n} \sigma_{lk} \tag{1}$$

where m and n are the number of rows and columns of the face image respectively (assuming that the pixels gray level are the image features), N is the number of extracted sub-graphs. σ_{lk} is the standard deviation of image dimension k in the sub-graph l. Equation(1) will be valid if the number of hierarchical grouping levels is 2. If dataset is very large however, a regrouping is required again to the third or higher levels. To ensure this, an additional term guaranteeing that the variance of the next grouping stage is also be maximized is included in Equation(1). This term deals with the inter-sup-graphs mean (the difference between means of different groups), forcing groups far from each other to have the maximum variance between its group members:

$$\mu_{diff} = \sum_{j=1}^{N} \sum_{i \neq j}^{N} d\left(\mu_i, \mu_j\right) \tag{2}$$

where $d\left(\mu_i, \mu_j\right)$ is the Euclidean distance between the mean of sub-graph i and the mean of sub-graph j. Equation (3) is the required objective function to be maximized:

$$\max_{I_{ij}} g\left(I_{ij}\right) = \max_{I_{ij}} \left(\sigma_{total} + \mu_{diff}\right) \tag{3}$$

where I is the face image vector. Equation (3) can be expressed in terms of minimization as given in equation (4).

$$\min_{I_{ij}} g\left(I_{ij}\right) = \min_{I_{ij}} \left(-\sigma_{total} - \mu_{diff}\right) \tag{4}$$

It has been reported in that L1 (absolute difference) metric works better than L2 (Euclidean distance) to measure the distance between two projected images in the Eigen space. Therefore, the effect of both metrics in the grouping process are tested to obtain the best recognition rate. The optimized group generation using equation (4) can be done as a separate process through the utilization of meta heuristics such as Simulated Annealing.

5 Results

The results section is divided into two parts for two different datasets. The first section is dedicated to ORL AT&T whereas the rest of the results involve Extended B+ Yale dataset. Each part provides comparisons between the different grouping techniques as well as the difference when using optimum dissimilarity metric. The proposed techniques have been implemented using MATLAB on Ubuntu 12.04 OS.

ORL AT&T dataset

Total number of images in the dataset is 400. These images have been divided into two sets. A set of 200 images for training, and another set of 200 images for testing. Since the number of images in the dataset is not too large, two level hierarchies have been implemented. The group size is considered as 50 images for the first grouping level. The following table shows the results obtained for rank 1 best match:

Method	Recognition Rate
Original Eigenface	92.5%
Similarity Grouping	94.5%
Random Grouping	94.0%
Mean Dissimilarity Grouping	93.5%
Optimum Dissimilarity L2 Metric	95.0%
Optimum Dissimilarity L1 Metric	94.0%

Extended B+ Yale dataset

The number of images in the dataset is 14,800. Similar to the ORL AT&T dataset the images are divided into two sets. One set of 7,400 images for training, and another set of 7,400 images is for testing. Due to the large number of images, a three level hierarchy has been used (the two levels recognition did not provide significant improvement in recognition rate). The training images have been divided into 140 groups, each with approximately 50 images. Three different grouping strategies have been tested. Following results are obtained for rank 1 best match:

- The original Eigenface algorithm recognition rate is 55%.

- Similarity Grouping: The recognition rate improves to approximately 77% - 83.5% depending on the number of best images selected from each group in the first level, and the number of groups in the regrouping step in the second level. A recognition rate of 83.5% is achieved when the best 5 matching images are selected from each first level group. A group constructed from these images is regrouped into the next grouping level. Then best 5 matches are selected from each subgroup, and a final Eigenface step is applied on these to obtain the best match. Results are shown in Figure 4. The main disadvantage of this grouping method is that, the execution time increases when the number of groups in the second level increases. The algorithm used for grouping is the K-means clustering algorithm.

- Random Grouping: The recognition rate improves to a range of 88% - 88.65% depending on the number of best matching images selected from the first level groups(from 2 to 5), and the number of groups in the regrouping step in the second level. The recognition rates are noted to be less dependent on the number of best images selected from each group. Further, the execution time is almost independent of the number of best images selected from a group, and the number of regroups.

- Dissimilar Grouping (based on the L2 distance from the mean image on the training set): The recognition rate improves to the range of 89% - 90.15% depending on the number of best images selected from each group in the first level, and the number of groups in the regrouping step in the second level. The recognition rates are less dependent on the number of best images selected from first level groups. Further, the execution time is almost similar for any number of best matches selected from a group, and number of regroups.

• Optimum Dissimilarity (based on L2 metric): Images are grouped based on the stated objective function in three level hierarchy with L2 metric. This method improved rank 1 rate to 91.5%.

• Optimum Dissimilarity (based on L1 metric): Images are grouped based on stated objective function in three level hierarchy with L1 metric. This method improved rank 1 rate to 93.6%. Also, for this dataset, the probability that the correct person appears in the best top 10 images is tested (rank 10), as shown in Figure 5.

The results in Figure 4 indicate that the recognition rate increased significantly when the hierarchical technique was used, especially for large databases (Extended B+ Yale). The recognition rate has improved further by the proposed optimum dissimilarity grouping criteria. In summary, compared to the results in[12,4,10,5] where the ICA and Boltzmann machines are used on the same datasets (around 82% for ICA and 83% for Boltzmann approach), the recognition rate of the proposed algorithm is superior. Further, the proposed optimized dataset grouping technique is compatible with other powerful recognition methods such as ICA or LBP, and not just with the PCA based Eigenface technique. Another advantage of this algorithm is that it uses all the training datasets as one bulk with the unsupervised grouping technique, which is completely independent of the face background and illumination levels.

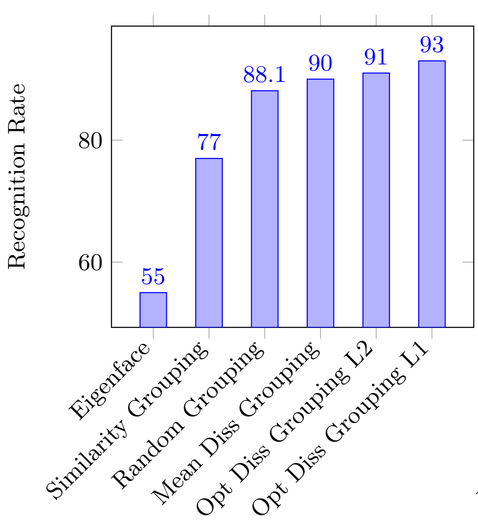

Fig. 4. Rank 1 recognition rate of different techniques for Extended B+ Yale dataset

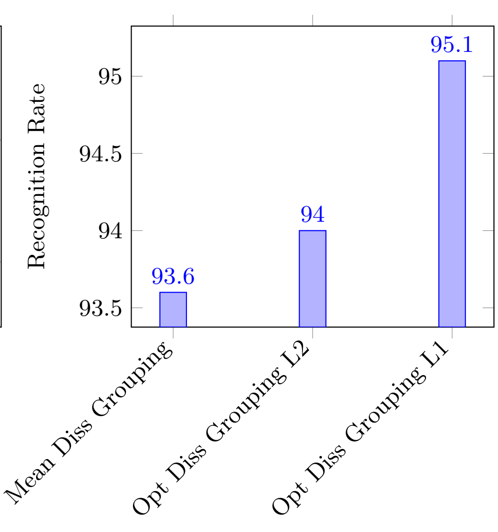

Fig. 5. Comparison between rank 10 recognition rate of dissimilarity grouping techniques for Extended B+ Yale dataset

6 Conclusions

The paper presented a hierarchical sub-graph selection algorithm that aims at overcoming the large dataset limitation of the standard face recognition algorithms. The algorithm is based on creating small sub-graphs, selecting best matches from each sub-graph, and then dynamically creating next-level sub-graphs until a single group remains. The best match from this last group is accepted as the rank 1 final result of face recognition. The study also investigated the best approach for creating sub-graphs by developing an objective function that can be used for best dissimilarity between groups at all levels. Detailed testing on large benchmark datasets indicates that the proposed method produces best results with a sub-graph size of approximately 50 nodes (images) for the Eigenface technique. Compared to the standard Eigenface algorithm, the new hierarchical sub-graph selection algorithm improves the recognition rate by more than 40% on the original Eigenface algorithm, and by more than 2% on the mean based dissimilarity method. The future work involves applying the hierarchical technique to additional unsupervised face recognition algorithms such as Independent Component Analysis (ICA), KPCA, LBP, SIFT and other computer vision algorithms that suffer from degradation in recognition rate due to large dataset size.

References

1. http://vision.ucsd.edu/~leekc/ExtYaleDataset/ExtYaleB.html.
2. Bach, F.R., Jordan, M.I.: Kernel independent component analysis. J. Mach. Learn. Res. 3, 1–48 (Mar 2003), http://dx.doi.org/10.1162/153244303768966085
3. Bartlett, M., Movellan, J.R., Sejnowski, T.: Face recognition by independent component analysis. Neural Networks, IEEE Transactions on 13(6), 1450–1464 (Nov 2002)
4. Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.R.: Recognizing faces with PCA and ICA. Computer Vision and Image Understanding 91(1-2), 115–137 (2003), http://dx.doi.org/10.1016/S1077-3142(03)00077-8
5. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: illumination cone models for face recognition under variable lighting and pose. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23(6), 643–660 (Jun 2001)
6. Kyperountas, M., Tefas, A., Pitas, I.: Face recognition via adaptive discriminant clustering. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. pp. 2744–2747 (Oct 2008)
7. Lu, J., Plataniotis, K.: Boosting face recognition on a large-scale database. In: Image Processing. 2002. Proceedings. 2002 International Conference on. vol. 2, pp. II–109–II–112 vol.2 (2002)
8. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using lda-based algorithms. Trans. Neur. Netw. 14(1), 195–200 (Jan 2003), http://dx.doi.org/10.1109/TNN.2002.806647
9. Moon, H., Phillips, P.J.: Computational and performance aspects of pca-based face-recognition algorithms. Perception 30(3), 303–321 (2001), http://www.perceptionweb.com/abstract.cgi?id=p2896

10. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(11), 2106–2112 (Nov 2010)
11. Schölkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. 10(5), 1299–1319 (Jul 1998), http://dx.doi.org/10.1162/089976698300017467
12. Tang, Y., Salakhutdinov, R., Hinton, G.: Robust boltzmann machines for recognition and denoising. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 2264–2271 (June 2012)
13. Tse, S.H., Lam, K.M.: Efficient face recognition with a large database. In: Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on. pp. 944–949 (Dec 2008)
14. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on. pp. 586–591 (Jun 1991)
15. Yang, M.H.: Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition. pp. 215–. FGR '02, IEEE Computer Society, Washington, DC, USA (2002), http://dl.acm.org/citation.cfm?id=874061.875432
16. Zhao, W., Chellappa, R., Krishnaswamy, A.: Discriminant analysis of principal components for face recognition. In: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. pp. 336–341 (Apr 1998)