Sub-graph Selection and its Application to Improve 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. We develop a new training dataset partitioning methodology to improve face recognition for large datasets, and applying it to the Eigenface algorithm as an example of the face recognition techniques that suffer from this problem. Our algorithm represents all training face images as a fully connected graph, and our target is to divide the fully connected graph into simpler sub-graphs to enhance the overall recognition rate. The sub-graphs are generated dynamically, and we provide a comparison between different subgraph selection techniques including minimizing edge weight sums, random selection, and maximizing sum of edge weights inside the sub-graph. The optimized hierarchical dynamic technique developed in this paper with sub-graphs selection increases the recognition rate in large benchmark image dataset by more than 40% for rank 1 recognition rate compared to the original single large graph method. The approach developed in this paper is applicable not only to Eigenface technique, but also to other unsupervised face recognition techniques as well such as ICA, KPCA, Restricted Boltzmann Machine, and Orthogonal Laplacian etc.., and other datasets, spatially if the number of images per person in the training data are low (about one image per person).

Keywords: Face Recognition, Sub-Graph Selection, Hierarchical Techniques, Optimum Sub-Graphs

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 branches that handles this problem based on a given face image. In the last decade, many face recognition algorithms have been proposed and developed, for example, Direct Correlation, Principal Component Analysis (PCA),[1–3] Linear Discriminant Analysis (LDA),[4–9] Independent Component Analysis (ICA),[10–12] Kernel methods (KPCA, SVM, etc..),[13,14] other high dimension features methods as LBP, SURF, or 3D methods[15] etc.. Some of these methods are supervised (LDA, SVM, ... etc.), where the given data is divided into a training dataset and testing and validation datasets. The training dataset is then divided into labeled groups (classes), with each class containing images for one person. The other methods for face recognition are unsupervised (PCA, KPCA, ICA, etc.), where the given data is separated into training and testing data sets, but the training dataset is unlabeled and the algorithm takes care of the classes separation process. Many research and experiments have proven that simple recognition algorithms like Eigenface produces good recognition with small sized and accurately clipped face datasets (typically less than 100 images). If the number of images in the dataset increases to above a few hundred, then the recognition accuracy of these algorithms starts to decrease significantly. Some research approaches use indexing to handle this problem.[16] The main disadvantage of such method is that it can be used only with specific features and techniques, and is very sensitive to image normalization, orientation and features calculation. In the case of large datasets, a hierarchical partitioning technique can be used to improve recognition rate. This technique divides the given training dataset into small subgroups such that a comparison with input image is done with a small set of images at a time. The best matching results from each subgroup are fed to the next level of groups until a single new constructed group remains. Finally, the best result from this group determines a match or mismatch. Such hierarchical grouping principle on the training dataset has been used in,[17,18] but in these two papers a supervised grouping is used, where the training dataset is divided according to the image class (images of the same person are grouped together), and thus the algorithm achieves good separation between groups. However, main drawback of this approach is that if multiple images of

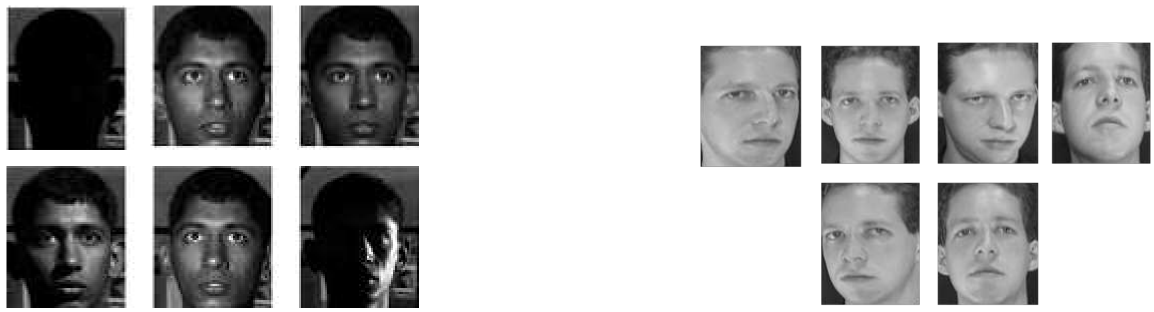

Figure 1. Sample of the datasets for used for the proposed technique

the same person are not available, then the group size is reduced to one. Further, the number of groups increases significantly as the number of images increase, especially if there are few images for each person. Unsupervised grouping technique can solve these problems because it will group images without considering person's identity. The main contribution of this work is:

1. Introducing a new unsupervised grouping technique for large training data.
2. Applying different grouping criteria on the proposed method.
3. Testing the proposed technique on two benchmark databases.

The remaining of the paper is divided into five sections. In the first section, the sub-graph selection process is explained . It is followed by a detailed explanation of the proposed hierarchical algorithm. The next section shows how we can further improve the recognition rate by optimizing the grouping process. This is followed by providing the results of the suggested techniques on the benchmark datasets. Finally, conclusions and future work are discussed.

2. SUB-GRAPH SELECTION PROCESS

The sup-graph selection process is the process of selecting a sub-graph $k_o$from a graph G that has a spsific critiria. And to apply it to images it should work as following, assuming that all training face images are a full connected graph (G) with number of nodes (N) and the edge between every two nodes w_{ij}equal the Euclidean distance between the features of these two nodes i and j. Our goal is to find the best sub-graphs set $S = \{k_1, k_2, k_3, ..., k_m\}$each of them has a number of nodes (l) , where $k_o$is the sub-graph number o, and m is the total number of reconstructed sub-graphs that will be used in the hierarchical technique. We investigate different strategies for this sub-graph selection process: 1)minimizing sum of edges weight within the entire sub-graph; 2) random chosen nodes for the sub-graph or 3) maximizing the sum of edges weight within the entire sub-graph. After sub-graphs groups are created, regular face recognition technique (Eigenface in this case) is applied to each full connected sub-graph to select the top best matches from each group. These set of matches from the first sub-graphs level forms the next level of subgroups. This process is repeated until a single small full connected graph of (l) nodes remains. We present this hierarchical grouping algorithm and explore different variations of it by testing it on benchmark datasets to prove that it can improve recognition rate on the all full connected images graph. Figure 1 shows a sample for the used datasets for the proposed sub-graph selection algorithm

3. HIERARCHICAL RECOGNITION TECHNIQUE

From testing many face recognition algorithms, we have determined that the recognition rate in unsupervised algorithms like standard Eigenface technique drops down if the number of images in the dataset is roughly above 100. This leads to the question, what will be the case if we divide the entir training dataset to small sub-sets? assuming that the face images are full connected graph as explained in the previews section(G) with number of nodes (N) and our goal is to find the best sub-graphs set $S = \{k_1, k_2, k_3, ..., k_m\}$each of them has a number of nodes ($l \leq 100$) , where m is the number of reconstructed sub-graphs, will this improve the recognition rate over

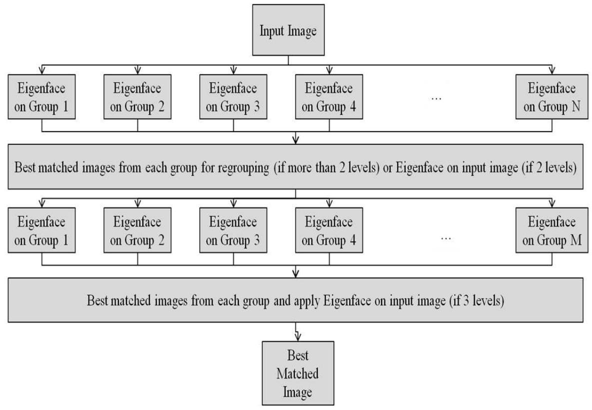

Figure 2. The Proposed Hierarchical System

the hierarchical technique?. So we will do our test as following: First, applying recognition algorithm over each of these sub-graphs (groups), we pick a few top matched nodes from each sub-graph (group) (2 to 5), then create new groups from these top matches. Depending on the number of images in the dataset, a number of hierarchical levels will be created. Second, recognition algorithm (e.g. Eigenface) is applied on each level group. Then in 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 2 shows a block diagram of the proposed hierarchical technique when the Exigence is the recognition method. The main challenge in this technique is to determine the best sub-graphs selection strategy to improve the overall face recognition rate. There are three possible grouping strategies: Similarity Grouping (minimizing 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). Random Grouping: Assigning the images to the groups (sub-graphs) randomly. Dissimilar Grouping (maximizing sum of weights in the entire sub-graph, or 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). The other challenge is to find the suitable number of levels in the hierarchical system, and the number of matched images to be selected from each level to feed into the next level in the hierarchy. To solve the above challenges, the three possibilities have been tested on a large dataset (Extended Yale B+)[19] having different positioning and illumination levels to check which approach will lead to best results with the hierarchical face recognition technique.

4. OPTIMIZED DISSIMILARITY SUB-GRAPH SELECTION TECHNIQUE

As will be shown in the results section, the simple dissimilarity measurement (maximizing the distance between group images by taking the mean image as a reference) worked better than the other two grouping techniques (similarity and random selection). But this method has some drawbacks. 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 them 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, which will lead to a poor grouping. One can see 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. The above conclusion mathematically stated is that the variance between all the sub-graph (group) nodes over all basis should be maximized. So applying this method to a face image dataset sub-graphs selection will lead to the following equation:

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

where m, n are the number of rows and columns of the face image respectively (assuming that the pixels gray level are the image features) and N is the number of extracted sub-graphs. And σ_{lk} is the standard deviation of image dimension k in the sub-graph l. This equation (1) will work perfectly if the number of hierarchical grouping levels is 2. But if dataset is large, then it is required to regroup again to the third or more levels. In this case another term should be added to equation (1) to guarantee that the variance of the next grouping stage will also be maximized. The term will deal with extra sub-graphs' mean (the difference between means of different groups), which will lead to making groups far from each other to have the maximum variance between its group members.

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

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) will be our 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) measurement works better than L2 (Euclidean distance) to measure the distance between two projected images in the Eigen space, so we test the effect of both of these metrics in the grouping process to get the best recognition rate. The optimized group generation from the minimization of equation 4 can be done in a separate process than the face recognition process through the use of meta heuristics such as Simulated Annealing, Tabu search, or Ant Colony Optimization (ACO). ACO has very good parallel execution properties as all ants in the colony can be launched in parallel, thus causing less overhead on the overall face recognition execution time

5. RESULTS

The results section is divided to two parts. Part one: for ORL AT&T, and part two: for Extended B+ Yale. Every part provides comparison between the different grouping techniques as well as the difference when using optimum dissimilarity metric. All proposed techniques have been implemented in MATLAB on Ubuntu 12.04 OS.

ORL AT&T dataset

The total number of images in this 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 for this dataset. For the first grouping level, the group size is 50 images. We got the following results for rank 1 best match:

- The recognition rate of the original Eigenface algorithm without any grouping is 92.5%.
- Similarity Grouping: Images are grouped using K-means clustering algorithm with L2 as a metric. In this case the recognition rate improved to 94.5%.
- Random Grouping: Images are randomly grouped. In this case the recognition rate has been improved to 94%.
- Dissimilarity Grouping: Images are grouped based on dissimilar (based on the L2 distance from the mean image of the entire dataset). In this case the recognition rate improved to 93.5%.
- Optimum Dissimilarity (based on L2 metric): Images are grouped based on stated objective function, with L2 metric. This method improved the recognition rate to 95%.
- Optimum Dissimilarity (based on L1 metric): Images are grouped based on stated objective function, with L1 metric. This recognition rate obtained is 94%. Figure 3 shows a comparison of the results on AT&T dataset.

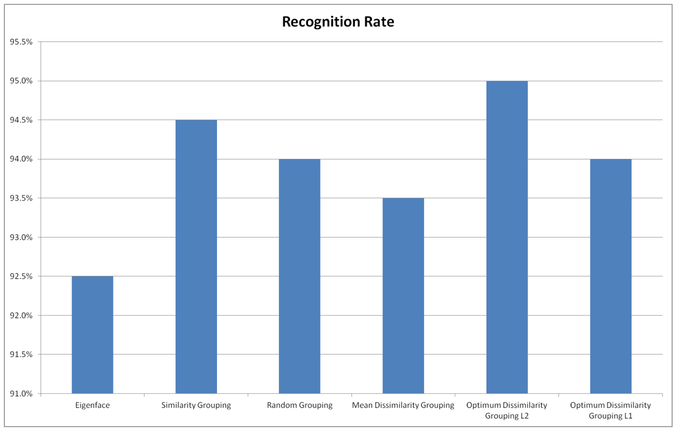

Figure 3. Rank 1 recognition rate of different techniques forORL AT&T dataset.

Extended B+ Yale dataset

The number of images in this dataset is 14,800. Again images have been divided into two sets. One set of 7400 images for training. And another set of 7400 images is for testing. Because of the large number of images, a three level hierarchy has been used (the two levels recognition did not give significant improvement in recognition rate). The training images have been divided into 140 groups, each of them with about 50 images. The three different grouping strategies have been tested. We obtained the following results for rank 1 best match:

- The original Eigenface algorithm recognition rate is 55%.

- Similarity Grouping: The recognition rate improves to about 77% - 83.5% depending on 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 we reselect best 5 matches from each subgroup, and apply a final Eigenface step on these to find the best match. Results are shown in Fig. 6. 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 between 88% - 88.65% depending on the number of best matched images selected from the first level groups(from 2 to 5), and the number of groups in the regrouping step in the second level. From the results, we can see that the recognition rates are less dependent on the number of best images selected from each group. Further, the execution time is almost similar for any number of best images selected from a group, and number of regroups.

- Dissimilar Grouping (based on the L2 distance from the mean image on the training set): The recognition rate improves to order 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. Again from the results, we can see that 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 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 again to 93.6%. Also, for this dataset, the probability that the correct person appears in the best top 10 images is tested (rank 10), and is shown in Figure 5. The results in Figure 4 show that the recognition rate increased rapidly when the hierarchical technique was used, especially for large databases (Extended B+ Yale). The recognition rate has been improved further by using our proposed optimum dissimilarity grouping criteria. From this result, and the results in[20–23]

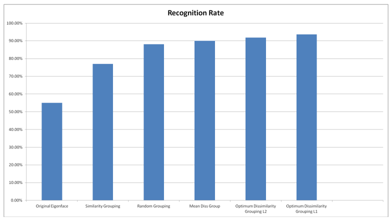

Figure 4. Rank 1 recognition rate of different techniques for Extended B+ Yale dataset

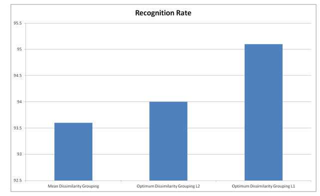

Figure 5. Comparison between rank 10 recognition rate of dissimilarity grouping techniques for Extended B+ Yale dataset

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 our algorithm is much better. Further, our optimized dataset grouping technique can be used with other powerful recognition methods such as ICA or LDA, and not just with the PCA based Eigenface technique. Another advantage of our 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.

6. CONCLUSIONS

We have presented a hierarchical sub-graph selection algorithm that overcomes the large dataset limitation of the standard face recognition algorithms such as Eigenface. Our algorithm is based on creating small sub-graphes, selecting best matches from each sub-graph, and then dynamically creating next-level sub-graphes until a single group is remaining. The best match from this last group is the rank 1 final result of face recognition. We have also investigated the best approach to creating sub-graphes by developing an objective function that can be used for best dissimilarity between groups in all levels. Detailed testing on large benchmark datasets indicates that our proposed method produces best results with a sub-graph size of around 50 nodes (images) for the Eigenface technique. As compared to the standard Eigenface algorithm, our 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. Our future work involves applying the hierarchical technique to other unsupervised face recognition algorithms such as Independent Component Analysis, KPCA, and other algorithms.

REFERENCES

1. M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991. [Online]. Available: http://dx.doi.org/10.1162/jocn.1991.3.1.71
2. ——, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, Jun 1991, pp. 586–591.
3. A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, Jun 1994, pp. 84–91.
4. C. Liu and H. Wechsler, "Comparative assessment of independent component analysis (ica) for face recognition," in *International Conference on Audio and Video Based Biometric Person Authentication*, 1999, pp. 22–24.
5. K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of Optical Society of America A*, vol. 14, pp. 1724–1733, 1997.
6. P. N. Belhumeur, J. a. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997. [Online]. Available: http://dx.doi.org/10.1109/34.598228
7. W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, Apr 1998, pp. 336–341.

8. A. M. Martínez and A. C. Kak, "Pca versus lda," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001. [Online]. Available: http://dx.doi.org/10.1109/34.908974
9. J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using lda-based algorithms," *Trans. Neur. Netw.*, vol. 14, no. 1, pp. 195–200, Jan. 2003. [Online]. Available: http://dx.doi.org/10.1109/TNN.2002.806647
10. H. Moon and P. J. Phillips, "Computational and performance aspects of pca-based face-recognition algorithms," *Perception*, vol. 30, no. 3, pp. 303–321, 2001. [Online]. Available: http://www.perceptionweb.com/abstract.cgi?id=p2896
11. M. Bartlett, J. R. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 13, no. 6, pp. 1450–1464, Nov 2002.
12. F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, Mar. 2003. [Online]. Available: http://dx.doi.org/10.1162/153244303768966085
13. M.-H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, ser. FGR '02. Washington, DC, USA: IEEE Computer Society, 2002, pp. 215–. [Online]. Available: http://dl.acm.org/citation.cfm?id=874061.875432
14. B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998. [Online]. Available: http://dx.doi.org/10.1162/089976698300017467
15. J. Lu, K. Plataniotis, and A. Venetsanopoulos, "Boosting linear discriminant analysis for face recognition," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 1, Sept 2003, pp. I–657–60 vol.1.
16. S.-H. Tse and K.-M. Lam, "Efficient face recognition with a large database," in *Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on*, Dec 2008, pp. 944–949.
17. J. Lu and K. Plataniotis, "Boosting face recognition on a large-scale database," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 2, 2002, pp. II–109–II–112 vol.2.
18. M. Kyperountas, A. Tefas, and I. Pitas, "Face recognition via adaptive discriminant clustering," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, Oct 2008, pp. 2744–2747.
19. [Online]. Available: http://vision.ucsd.edu/~leekc/ExtYaleDataset/ExtYaleB.html.
20. Y. Tang, R. Salakhutdinov, and G. Hinton, "Robust boltzmann machines for recognition and denoising," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2264–2271.
21. B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, "Recognizing faces with PCA and ICA," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 115–137, 2003. [Online]. Available: http://dx.doi.org/10.1016/S1077-3142(03)00077-8
22. I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 11, pp. 2106–2112, Nov 2010.
23. A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 643–660, Jun 2001.