ILLUMINATION ROBUST DICTIONARY-BASED FACE RECOGNITION

Vishal M. Patel[1], *Tao Wu*[1], *Soma Biswas*[2], *P. Jonathon Phillips*[3] *and Rama Chellappa*[1]

[1]UMIACS, University of Maryland, College Park, MD
[2] University of Notre Dame, Notre Dame, IN
[3] National Institute of Standards and Technology, Gaithersburg, MD
{pvishalm,taowu,rama}@umiacs.umd.edu sbiswas@nd.edu jonathon@nist.gov

ABSTRACT

In this paper, we present a face recognition method based on simultaneous sparse approximations under varying illumination. Our method consists of two main stages. In the first stage, a dictionary is learned for each face class based on given training examples which minimizes the representation error with a sparseness constraint. In the second stage, a novel test image is projected onto the span of the atoms in each learned dictionary. The resulting residual vectors are then used for classification. Furthermore, to handle changes in lighting conditions, we use a relighting approach based on a non-stationary stochastic filtering framework to generate multiple images of the same person with different lighting. As a result, our algorithm has the ability to recognize human faces with good accuracy even when only a single or a very few images are provided for training. The effectiveness of the proposed method is verified on publicly available databases and it is shown that this method is efficient and can perform significantly better than many competitive face recognition algorithms.

***Index Terms*—** Face recognition, illumination variation, albedo, relighting, simultaneous sparse signal representation.

1. INTRODUCTION

Face recognition is a challenging problem that has been actively researched for many years [1]. Current systems work very well when the test image is captured under controlled conditions. However, their performance degrades significantly when the test image contains variations that are not present in the training images. Some of these variations include illumination, pose, expression, cosmetics, and aging.

In recent years, the theories of Sparse Representation (SR) and Compressed Sensing (CS) have emerged as powerful tools for efficiently processing data in non-traditional ways. This has lead to a resurgence in interest in the principles of SR and CS for face recognition [2, 3, 4, 5, 6]. Wright *et al.* [3] introduced an algorithm, called Sparse Repersentation-based Classification (SRC), where the training face images are the dictionary and a novel test image is classified by finding its sparse representation with respect to this dictionary. This work was later extended to handle pose and illumination variations [4], [5]. Also, an expression-invariant face recognition method based on ideas from the distributed compressed sensing and joint sparsity models was proposed in [6]. Phillips [2] proposed matching pursuit filters for face feature detection and identification. The filters were designed through a simultaneous decomposition of a training set into a 2D wavelet expansion designed to discriminate among faces. It was shown that the resulting algorithm was robust to facial expression and the surrounding environment.

There are a number of hurdles that face recognition systems based on sparse representation must overcome. One is designing algorithms that are robust to changes in illumination; a second is that algorithms need to efficiently scale as the number of people enrolled in the system increases. In some of the above approaches, the former mentioned challenge is met by collecting a set of images of each person that spans the space of expected variations in illumination. The SRC approach recognizes faces by solving an optimization problem over the set of images enrolled into the database. This solution trades robustness and size of the database against computational efficiency.

In this paper, we present an algorithm to perform face recognition across varying illumination based on learning class specific dictionaries. Using a relighting method, we add many elements to the dictionary so that robustness to illumination changes can be realized. Our method consists of two stages. In the first stage, given training samples from each class, class specific dictionaries are trained with some fixed number of atoms. Then, a novel test image is projected onto the span of the atoms in each learned dictionaries. The residual vectors are then used for classification. The effectiveness of this method is demonstrated on experiments showing robustness to changes in illumination.

The organization of the paper is as follows. Our dictionary-based face recognition algorithm is detailed in section 2. Section 3 presents the experimental results on various datasets and Section 4 concludes the paper with a brief summary and discussion.

2. DICTIONARY-BASED RECOGNITION

Let $\tilde{\mathbf{D}} = [\mathbf{d}_1, \cdots, \mathbf{d}_K] \in \mathbb{R}^{N\times K}$ be a redundant dictionary with K atoms represented as columns $\mathbf{d}_j \in \mathbb{R}^N$ with $K \gg N$. The atoms have unit Euclidean norm. The choice of dictionary usually depends on the specific application. For instance, a dictionary may be chosen such that it favors sparse approximations or it can be chosen to resemble the structure that may appear in the input samples. For face recognition, in [3] the dictionary consisted of the gallery images while in [2] the dictionary contained steerable wavelet bases elements.

Given a data matrix $\mathbf{B} = [\mathbf{x}_1, \cdots, \mathbf{x}_m] \in \mathbb{R}^{N\times m}$ and a fixed dictionary $\tilde{\mathbf{D}} \in \mathbb{R}^{N\times K}$, simultaneous sparse approximation attempts to find a matrix Γ such that $\mathbf{B} \simeq \mathbf{D}\Gamma$. Where $\mathbf{D} \in \mathbb{R}^{N\times P}$, $P < N$, is a dictionary matrix whose atoms are selected from $\tilde{\mathbf{D}}$ and $\Gamma = [\gamma_1, \cdots, \gamma_m]$ is the matrix whose columns γ_i are the coefficients corresponding to each data vector $\mathbf{x}_i$. In other words, simultaneous sparse approximation attempts to approximate all the samples in $\mathbf{B}$ at once as a linear combination of a common subset of atoms with cardinality much smaller than N. In fact, by keeping the sparsity low enough, one can eliminate the internal variation of the samples in $\mathbf{B}$ which may lead to more robust representation. One algorithm for simultaneous sparse approximation based on greedy pursuit, called simultaneous orthogonal matching pursuit (SOMP), was proposed in [7]. SOMP uses a predefined dictionary. However, it has been shown that instead of using a predetermined dictionary, learning dictionaries from the training data provides much better representation and hence can improve the performance of reconstructive approach to discrimination.

Learning Class Specific Reconstructive Dictionaries: Designing dictionaries based on training is a much recent approach to dictionary design which is strongly motivated by the advances in the sparse representation theory [8]. We now briefly describe the K-SVD [9] algorithm for learning dictionaries for face images.

Given a set of examples $\mathbf{B} = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$, the goal of the K-SVD algorithm is to find a dictionary $\mathbf{D}$ and a sparse matrix Γ that minimize the following representation error

$$(\hat{\mathbf{D}}, \hat{\Gamma}) = \arg\min_{\mathbf{D},\Gamma} \|\mathbf{B} - \mathbf{D}\Gamma\|_F^2 \text{ subject to } \forall i \ \|\gamma_i\|_0 \leq T_0 \quad (1)$$

where γ_i represent the columns of Γ and the ℓ_0 sparsity measure $\|.\|_0$ counts the number of nonzero elements in the representation. Here, $\|\mathbf{A}\|_F$ denotes the Frobenius norm defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} \mathbf{A}_{ij}}$. The K-SVD algorithm alternates between sparse-coding and dictionary update steps. In the sparse-coding step, $\mathbf{D}$ is fixed and the representation vectors γ_is are found for each example $\mathbf{x}_i$. Then, the dictionary is updated atom-by-atom in an efficient way.

Classification based on Learned Dictionaries: Suppose that we are given C distinct face classes and a set of m_i training images per class, $i \in \{1, \cdots, C\}$. We identify an $l \times q$ grayscale image as an N-dimensional vector, $\mathbf{x}$, which can be obtained by stacking its columns, where $N = l \times q$. Let $\mathbf{B}_i = [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{im_i}] \in \mathbb{R}^{N \times m_i}$ be an $N \times m_i$ matrix of training images corresponding to the i^{th} class.

For training, we first learn C class specific dictionaries, $\mathbf{D}_i$, to represent the training samples in each $\mathbf{B}_i$, with some sparsity level T_0, using the K-SVD algorithm. Once the dictionaries have been learned for each class, given a test sample $\mathbf{y}$, we project it onto the span of the atoms in each $\mathbf{D}_i$ using the orthogonal projector $\mathbf{P}_i = \mathbf{D}_i(\mathbf{D}_i^T\mathbf{D}_i)^{-1}\mathbf{D}_i^T$. The approximation and residual vectors can then be calculated as

$$\hat{\mathbf{y}}^i = \mathbf{P}_i\mathbf{y} = \mathbf{D}_i\alpha^i \quad (2)$$

and

$$\mathbf{r}^i(\mathbf{y}) = \mathbf{y} - \hat{\mathbf{y}}^i = (\mathbf{I} - \mathbf{P}_i)\mathbf{y}, \quad (3)$$

respectively, where $\mathbf{I}$ is the identity matrix and $\alpha^i = (\mathbf{D}_i^T\mathbf{D}_i)^{-1}\mathbf{D}_i^T\mathbf{y}$ are the coefficients. Since the K-SVD algorithm finds the dictionary, $\mathbf{D}_i$, that leads to the best representation for each examples in $\mathbf{B}_i$, we suspect $\|\mathbf{r}^i(\mathbf{y})\|_2$ to be small if $\mathbf{y}$ were to belong to the i^{th} class and large for the other classes. Based on this, we can classify $\mathbf{y}$ by assigning it to the class, $d \in \{1, \cdots, C\}$, that gives the lowest reconstruction error, $\|\mathbf{r}^i(\mathbf{y})\|_2$:

$$d = \text{identity}(\mathbf{y}) = \arg\min_i \|\mathbf{r}^i(\mathbf{y})\|_2. \quad (4)$$

2.1. Rejection rule for non-face images

For classification, it is important to be able to detect and then reject invalid test samples. To decide whether a given test sample is valid or not, we define the following rejection rule.

Given a test image $\mathbf{y}$, for all classes in the training set, the score s_{yi} of the test image $\mathbf{y}$ to the i^{th} class is computed as $s_{yi} = \frac{1}{\|\mathbf{r}^i(\mathbf{y})\|_2^2}$, where $\mathbf{r}^i(\mathbf{y})$ is the residual vector as defined in (3). Then, for each test image $\mathbf{y}$, the score values are sorted in the decreasing order such that $s'_{y1} \geq s'_{y2} \geq \cdots \geq s'_{yC}$. The corresponding sorted classes are the candidate classes for each test image. The first candidate class is the most likely class that the test image belongs to. We define the ratio between the score of the first candidate class to the score of the second candidate class: $\lambda_y = \frac{s'_{y1}}{s'_{y2}}$ as a measure of the reliability of the recognition rate. Based on this, a threshold τ can be chosen such that, $\mathbf{y}$ is accepted as a good image if $\lambda_y \geq \tau$, otherwise rejected as an invalid image. Since the score values to all the candidate classes are sorted, the score values of the third and the higher order candidates are less than or equal to the score of the second candidate class. Hence, a high ratio λ_y for the test image $\mathbf{y}$ would show that the score of the first candidate class is significantly greater than all the other scores. Therefore, the identification result can be claimed to be reliable.

2.2. Image Relighting

Recognizing faces under varying illumination given a single training image is a difficult problem. In this section, we propose a method to deal with this illumination problem. The idea is to capture the illumination conditions that might occur in the test sample in the training samples.

We assume the Lambertian reflectance model for the facial surface. The surface normals, albedo and the intensity image are related by an image formation model. For Lambertian objects, the diffused component of the surface reflection is modeled using the Lambert's Cosine Law given by

$$I = \rho \max(\mathbf{n}^T\mathbf{s}, 0), \quad (5)$$

where I is the pixel intensity, $\mathbf{s}$ is the light source direction, ρ is the surface albedo and $\mathbf{n}$ is the surface normal of the corresponding surface point. Using this model, a nonstationary stochastic filtering framework was recently proposed in [10] to estimate the albedo from a single image. We adapt this method to first estimate the albedo map from a given face image. Then, using the estimated albedo map, we generate new images under any illumination condition using the image formation model (5). This can be done by combining the estimated albedo map with the average facial information [11].

It was shown in [12] that an image of an arbitrarily illuminated object can be approximated by a linear combination of the image of the same object in the same pose, illuminated by nine different light sources placed at preselected positions. The nine pre-specified light source directions are given by [12]

$$\phi = \{0, 49, -68, 73, 77, -84, -84, 82, -50\}^\circ$$
$$\theta = \{0, 17, 0, -18, 37, 47, -47, -56, -84\}^\circ.$$

Hence, the image formation equation can be rewritten as $I = \sum_{i=1}^9 a_i I_i$ where $I_i = \rho \max(\mathbf{n}^T\mathbf{s}_i, 0)$, and $\{\mathbf{s}_1, \cdots, \mathbf{s}_9\}$ are the pre-specified illumination directions. Since, the objective is to generate gallery images which will be sufficient to account for any illumination in the probe image, we generate images under the nine pre-specified illumination conditions and use them in the gallery. As a result, our algorithm has the ability to recognize human faces with good accuracy even when only a single or a very few images are provided for training. Fig. 1 shows some relighted images and the corresponding input images.

We summarize our dictionary-based face recognition (DFR) algorithm is in Fig. 2. Note that a K-SVD based face recognition algorithm was recently proposed in [13], but we differ in a few key areas.

Fig. 1. Examples of the original images (first column) and the corresponding relighted images with different light source directions from the PIE data set.

Unlike [13], we do not take discriminative approach to face recognition. Our method is a reconstructive approach to discrimination and does not require multiple images to be available. Another difference is that our algorithm can identify and reject non-face images.

Given a test sample $\mathbf{y}$ and C training matrices $\mathbf{B}_1, \cdots, \mathbf{B}_C$ where each $\mathbf{B}_i \in \mathbb{R}^{N \times m_i}$ contains m_i training samples.
Procedure:
1. For each training image, use the relighting approach described in section 2.2 to generate multiple images with different illumination conditions and use them in the gallery.
2. Learn the best dictionaries $\mathbf{D}_i$, to represent the face images in $\mathbf{B}_i$, using the K-SVD algorithm.
3. Compute the approximation vectors, $\hat{\mathbf{y}}^i$, and the residual vectors, $\mathbf{r}^i(\mathbf{y})$, using (2) and (3), respectively for $i = 1, \cdots, C$.
4. Identify $\mathbf{y}$ using (4).

Fig. 2. DFR algorithm.

3. RECOGNITION EXPERIMENTS

In this section, we present experimental results on some of the publicly available databases for face recognition such as Extended Yale B dataset [14] and PIE dataset [15]. The comparison with other existing face recognition methods in [3] suggests that the SRC algorithm is among the best. Hence, we treat it as state-of-the-art and use it as a bench mark for comparisons in this paper. See [16] for more details on our method and additional experimental results.

In all of our experiments, the K-SVD [9] algorithm is used to train the dictionaries with 15 atoms. The performance of our algorithm is compared with that of five different methods: SRC, nearest neighbor (NN), nearest subspace (NS), support vector machines (SVM) and class dependent principal component analysis (CDPCA) [17]. Our algorithm is also tested using several features, namely, Eigenfaces, Fisherfaces, Randomfaces, and downsampled images[1].

Results on Extended Yale B Database: There are a total of $2,414$ frontal face images of 38 individuals in the Extended Yale B database. These images were captured under various controlled indoor lighting conditions. They were manually cropped and normalized to the size of 192×168 [12]. Our first set of experiments on the Extended Yale B data set consist of testing the performance of our algorithm with different features and dimensions. The objective is to verify the ability of our algorithm in recognizing faces with different illumination conditions. We follow the experimental setup as considered in [3]. The feature space dimensions of 30, 56, 120, and 504 corresponding to the downsampling ratios of, $1/32$, $1/24$, $1/16$, and $1/8$, respectively are computed. We randomly select 32 images per subject (i.e. half of the images) for training and the other half for testing. The best recognition rates of different methods with different dimensions and features are compared in Table1.

[1] This means that, we first transform the given images into a feature space. We then train dictionaries on the feature space.

Table 1. Recognition Rates (RR) (in %) of different methods on the Extended Yale B database [16].

Method	DFR	SRC	NN	NS	SVM	CDPCA
RR	**99.17**	98.1	90.7	94.1	97.7	98.83

The maximum recognition rates achieved by DFR are 95.99%, 97.16%, 98.58% and 99.17% for all $30, 56, 120$ and 504 dimensional feature spaces, respectively. The maximum recognition rate achieved by SRC is 98.1% with 504D randomfaces [3]. Also, NN, NS, SVM and CDPCA achieve the maximum recognition rates of 90.7%, 94.1%, 97.7%, and 98.83%, respectively [3],[16]. As can be seen from this experiment that DFR performs favorably over some of the competitive methods for face recognition on the Extended Yale B database.

Results on PIE Database: The PIE database contains face images of 68 subjects. The images were captured under 13 different poses and 21 flashes under pose, illumination and expression variations. The face images were cropped with the size 48×40. In this experiment, our objective is to perform recognition across illumination with images from one illumination condition forming the gallery while images from another illumination condition forming the test set. In this setting, there is just one image per subject in each gallery and probe set. See [18] for more details on how the training and test data sets are created for this experiment. The average rank-1 results obtained using our method are reported in Table 2. The average rank-1 recognition rate achieved by our method is 99% and it outperforms the other competitive methods that follow similar experimental setting.

Table 2. Average rank-1 recognition rates (RR) (in %) of different methods on the PIE database [16].

Method	DFR	MA[18]	MB[18]	[10]
RR	**99**	93	96	94

Recognition with partial face features: In this section, we report the ability of our algorithm in recognizing faces from the partial face features. Partial face features have been used in recovering the identity of human faces before [3], [19]. We use the images in the Extended Yale B database for this experiment. We replicate the experimental setup of [3]. For each subject, 32 images are randomly selected for training, and the remaining images are used for testing. The region of eye, nose and mouth are selected as partial face features. For this experiment, we omitted the relighting step of our algorithm. Examples of these features are shown in Fig. 3. Table 3 compares the results obtained by using our method with the other methods presented in [3]. As can be seen from the table, our method achieves recognition rates of 99.3%, 98.8% and 99.8% on eye, nose and mouth region, respectively and it outperforms the other methods such as SRC, NN, NS and SVM [3].

Eye Nose Mouth

Fig. 3. Examples of partial facial features.

Rejecting non-face images: In this section, we demonstrate the effectiveness of our method in dealing with invalid test images with and without block occlusion. We test our rejection rule, described in Section 2.1, on the Extended Yale B data set. We use Subsets 1 and 2 for training and Subset 3 for testing. We simulate varying

Table 3. Recognition results with partial facial features.

	Right Eye	Nose	Mouth
Dimension	5,040	4,270	12,936
DFR	**99.3%**	**98.8%**	**99.8%**
SRC	93.7%	87.3%	98.3%
NN	68.8%	49.2%	72.7%
NS	78.6%	83.7%	94.4%
SVM	85.8%	70.8%	95.3%

levels of occlusion by replacing a randomly chosen block of each test image with random noise. We include only half of the subjects in the training set. This way, half of the subjects in the test set are new to the algorithm. We plot the Receiver Operating Characteristic (ROC) curve according to different τ values in Figure 4. As can be seen from this figure, that simple rejection rule performs quite well. It performs nearly perfectly at 10% occlusion and without any occlusion. Even at 50% occlusion, it performs better than making a random decision. This performance, can be further improved by applying our DFR method on different features such as PCA and Fisher.

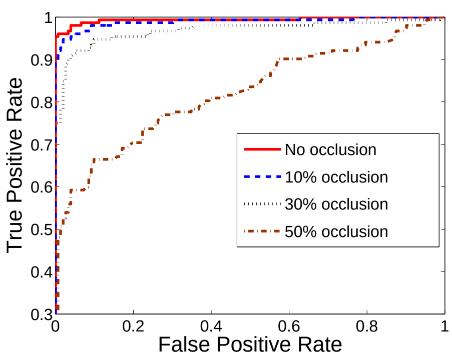

Fig. 4. ROC curves corresponding to rejecting invalid test samples experiment. The solid curve is generated by the DFR method based on our rejection rule. The dotted curves correspond to the cases when different levels of occlusion has been added to the test images.

Efficiency: Using a unix system with Intel Xeon E5506/2.13 GHz processor, on average our algorithm takes about 0.3 seconds to train a dictionary of 15 atoms for a gallery matrix containing 32 images of size 24×21. Recognizing a test sample from 38 classes takes on average about 0.01 seconds.

4. DISCUSSION AND CONCLUSION

We have proposed a new face recognition algorithm based on dictionary learning methods that is robust to changes in lighting. This entails using a relighting approach based on a robust albedo estimation. Various experiments on popular face recognition data sets have shown that our method is efficient and can perform significantly better than many competitive face recognition algorithms.

Even though, in this paper, we took a reconstructive approach to dictionary learning, it is possible to learn discriminative dictionaries for the task of face recognition. It remains an interesting topic for future work to develop a discriminative dictionary learning algorithm that is robust to pose, expression and illumination variations.

5. REFERENCES

[1] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Computing Surveys*, pp. 399–458, Dec. 2003.

[2] P. J. Phillips, “Matching pursuit filters applied to face identification,” *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 150–164, 1998.

[3] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[4] J. Huang, X. Huang, and D. Metaxas, “Simultaneous image transformation and sparse representation recovery,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, Ancorage, Alaska, June, 2008.

[5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, “Towards a practical face recognition system: Robust registration and illumination by sparse representation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 597–604, Miami, FL, 2009.

[6] P. Nagesh and B. Li, “A compressive sensing approach for expression-invariant face recognition,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1518–1525, Miami, FL, June 2009.

[7] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. part i: Greedy pursuit,” *Signal Processing*, vol. 86, no. 3, pp. 572–588, March 2006.

[8] R. Rubinstein, A. M. Bruckstein, and M. Elad, “Dictionaries for sparse representation modeling,” *Proceedings of the IEEE*, submitted 2009.

[9] M. Aharon, M. Elad, and A. M. Bruckstein, “The k-svd: an algorithm for designing of overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[10] S. Biswas, G. Aggarwal, and R. Chellappa, “Robust estimation of albedo for illumination-invariant matching and shape recovery,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 884–899, Mar. 2009.

[11] V. Blanz and T. Vetter, “Face recognition based on fitting a 3d morphable model,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, Sept. 2003.

[12] K. Lee, J. Ho, and D. J. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, May 2005.

[13] Q. Zhang and B. Li, “Discriminative k-svd for dictionary learning in face recognition,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2691–2698, 2010.

[14] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Ilumination cone models for face recognition under variable lighting and pose,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, June 2001.

[15] T. Sim, S. Baker, and M. Bsat, “The cmu pose, illumination, and expression database,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[16] V. M. Patel, T. Wu, S. Biswas, R. Chellappa, and P. J. Phillips, “Dictionary-based face recognition,” *UMIACS Tech. Report, UMIACS-TR-2010-07/CAR-TR-1030, University of Maryland, College Park*, July 2010.

[17] M. Deriche, “A simple face recognition algorithm using eigeneyes and a class-dependent pca implementation,” *International Journal of Soft Computing*, vol. 3, pp. 438–442, 2008.

[18] S. K. Zhou, G. Aggarwal, and D. W. Jacobs, “Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 230–245, Feb. 2007.

[19] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, “Face recognition by humans: nineteen results all computer vision researchers should know about,” *Proc. IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.