Computing Eigen Space from Limited Number of Views for Recognition

Paresh K. Jain, Kartik Rao P.*, and C.V. Jawahar

Centre for Visual Information Technology,
International Institute of Information Technology,
Hyderabad 500032, India
paresh@research.iiit.ac.in, kartikr@kritikalsolutions.com,
jawahar@iiit.ac.in

Abstract. This paper presents a novel approach to construct an eigen space representation from limited number of views, which is equivalent to the one obtained from large number of images captured from multiple view points. This procedure implicitly incorporates a novel view synthesis algorithm in the eigen space construction process. Inherent information in an appearance representation is enhanced using geometric computations. We experimentally verify the performance for orthographic, affine and projective camera models. Recognition results on the COIL and SOIL image database are promising.

1 Introduction

Recognition is an active area of research in computer vision. The problem of view-independent object recognition has received considerable amount of attention in recent years [1,2,3,4,5]. Recognition techniques can be broadly classified into (a) Shape representation based and (b) Appearance based techniques. Shape representation based techniques are popular for specific categories of applications, where the object's structure is more important than the intensity information within its boundaries. Plenoptic function captures object appearance across views and allows to completely model an object. However, estimation of this appearance representation is not a viable intermediate step in recognition. A popular alternative is to model the subspace that will contain all views of the object. Appearance based matching techniques attempt to model this space. Since images are bulky in nature, dimensionality reduction is usually sought to reduce the complexity of the appearance based representation. One of the very popular approaches for this purpose is the eigen image representation [5,6].

Following the successful application of Eigen spaces for face recognition [6], a real-time system [5] was built to recognize hundred objects imaged from multiple view points. This system employed parametric hypersurfaces constructed in the eigenspace, to model the appearance of the objects in different views. View based models are built from a large number of training images in [7]. Correlation between views of an object is exploited to construct the appearance models.

* Currently with Kritikal Solutions, New Delhi, India.

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 662–673, 2006.

Performance and applicability of these systems is often limited by the number of views. With reduction in number of views, the parametric representations also become poorer approximations of the real appearance models. A direct method to address this problem is to capture or synthesize additional views of the object and construct the appearance model from these images. When an object is viewed with multiple cameras, there holds some constraints in the geometry of the views [8] which allows the synthesis of novel views.

This paper presents a new approach to solve the view-independent recognition problem from limited number of views. We construct an eigen space, from a limited number views, equivalent to the one obtained from a large number of views. We show that these two are mathematically equivalent, except for the occluded pixels. The following sections have been written with the emphasis on linear Principal Component Analysis (PCA) but it may be extended to nonlinear techniques like Kernel PCA. Note that the output of the proposed technique is a data matrix of new views interpolated in the pose space. A linear PCA, Kernel PCA and many other nonlinear Component Analyses can equally well be applied over this data matrix.

The rest of the paper is organized as follows. Section 2 revisits Principal Component Analysis and introduces the notation used. Section 3 describes the details of the eigen space construction process in the proposed scheme for simple case of orthographic camera projection model. Section 4 discusses its extension to other camera projection models and application to object recognition. In Section 5, we demonstrate that the eigen space constructed from limited views practically approximates the ideal one for orthographic, affine and projective camera models. Performance of recognition is verified on a set of synthetic images and images from the COIL [9] and SOIL [10] databases.

2 Eigen Images for Representation

Eigen space representation is very popular for compression [11] and recognition [5,6]. Eigen space compactly represents the appearance in the presence of variations in instantiation of the object, say human face, due to pose or illumination [6]. A new set of bases vectors (eigenvectors of the covariance matrix) along the direction of maximal variance is employed to build the representation. The decorrelation of features achieved by PCA, allows discarding of features that contribute less to the content of an image, without significant loss of information.

Given a set of images $\{I_1, I_2, \ldots, I_P\}$, each of size $N = h \times w$, the eigen space is obtained as follows: Each image is arranged as a vector by concatenating the pixels in row order. The image vectors are normalized by subtracting the mean vector from each of the images, i.e., $\widetilde{I_i} = I_i - \mu$ where $\mu = \frac{1}{P}\sum_{i=1}^{P} I_i$. These normalized images are arranged to form a data matrix A of dimensions $N \times P$.

$$A = \left[\widetilde{I_1}\ \widetilde{I_2} \ldots \widetilde{I_P} \right] \tag{1}$$

The matrix A is multiplied with its transpose, A^T, to yield the scaled version of the covariance matrix Σ. Eigenvectors corresponding to the larger eigenvalues

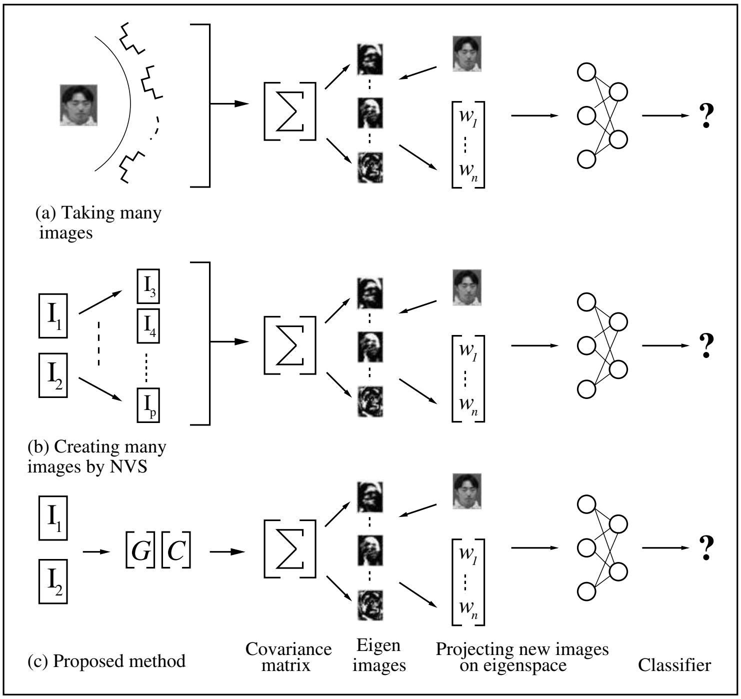


Fig. 1. Three different methods to construct Eigenspace representation. Popular method (a) capture multiple images of the object under different pose. We propose two alternate methods for eigen space construction. Method (b) constructs eigen space by synthesis of many additional images. Method (c) does the same without explicit synthesis of novel views from two views of an object.

act as the basis vectors for the new representation. If the images are not mean centered, we will get a correlation matrix whose eigenvectors are directly related to those of Σ.

To build a view-independent recognition system, a large number of images of an object is needed and the eigenspace is typically constructed [5] from these images. We propose two alternate schemes for constructing high resolution appearance space from limited number of input images. Recent advances in multiview geometry [8] permit us to interpolate or extrapolate from two input views I_1 and I_2 to obtain new views $I_3, \ldots, I_P$. Incorporation of the geometric information into the appearance representation adds the otherwise missing information, which could not be obtained from limited number of images. We propose to construct the eigen space from a set including the synthesized ones. On a closer look, synthesis of novel views seems to be a redundant step in the construction of the eigen space. In the next section, we demonstrate that without the intermediate views, one could directly construct the eigen space and hence carry out recognition from limited views. A conceptual explanation of these alternate procedures is presented in Fig. 1. Although appearance of objects from multiple views can be learnt with method (a) (Fig. 1); it generally requires sophisticated camera(s) and/or light(s) setup. There are a good number of scenarios (for example surveillance) where such freedom is unavailable. Hence there is a need for techniques such as proposed in this paper; which can build a denser eigenspace with limited available views in a computationally efficient manner.

3 Eigenspace Construction for Orthographic Cameras

We start with a simple case, where two orthographic cameras related by a Y-axis rotation provides the input. Let I_1 and I_2 be the input images of size $h \times w$. If x_1 and x_2 are x-coordinates of corresponding points in the two images, then the corresponding point x_3 in a third (novel) image I_3 is given by (see [12])

$$x_3 = ax_1 + bx_2, \tag{2}$$

where a and b depend on the translation and rotation that the camera undergoes to image I_1, I_2 and I_3. The y-coordinate remains the same. Note that it is the coordinates of the corresponding points, which are linearly related, and this does not directly imply anything on the eigenspace constructed from the intensity values. More over, since a and b can be fractional values, x_3 need not be an integer. The intensity at a point in the novel view can be expressed as a linear combination of intensities in its neighborhood, which can be computed from the original image.

Let $G_j^T = \left[I_1(j,1),\, I_1(j,2),\, \ldots\, I_1(j,w) \right]$ denote the intensity values in row j of image I_1. If $C_j^l = \left[y_1\; y_2\; \ldots\; y_w \right]$ represents a vector of interpolation coefficients for the lth pixel in the jth row, then the intensity at any point in row j can be expressed as

$$I_3(j,l) = \sum_{k=1}^{w} I_1(j,k) \cdot y_k = G_j^T \cdot C_j^l \tag{3}$$

Let (x_{k_1}, j), (x_{k_2}, j) and (x_{k_3}, j) be corresponding points in I_1, I_2 and the novel view I_3. These are related by Equation (2). To synthesize the digital image, we need to interpolate from the synthesized real coordinates. The point (x_{k_3}, j) contributes to the intensity of $I_3(l,j)$ only if the distance between the two points is less than one unit on the integer grid. The contribution varies inversely with distance. Thus the element y_k, $1 \leq k \leq w$ is given by

$$\begin{matrix} 1 - dist & \text{if} & dist < 1 \\ 0 & \text{if} & dist \geq 1 \end{matrix}, \tag{4}$$

where $dist = |a_i x_{1k} + b_i x_{2k} - l|$. The interpolation vector C_j^l is determined for each pixel in the novel view. Then, the product $G_j^T C_j^l$ is computed for each point (j,l) in the novel view. These products are arranged to obtain the vector form of the novel view I_3. Hence the novel view I_3 is given as

$$\begin{aligned} &\left[G_1^T C_1^1 \;\ldots\; G_1^T C_1^w \;\ldots G_h^T C_h^1 \;\ldots\; G_h^T C_h^w \right]^T \\ &= \mathcal{G} \left[C_1^1 \;\ldots\; C_1^w \;\ldots\; C_h^1 \;\ldots\; C_h^w \right]^T = \mathcal{G}\mathcal{C}_3, \end{aligned}$$

where $\mathcal{G}$ is a $hw \times hw^2$ matrix obtained by appropriately arranging the G_i's and $\mathcal{C}_3$ is a $hw^2 \times 1$ vector obtained by column order concatenation of C_j^l's. Each row of $\mathcal{G}$ can be considered as the concatenation of hw vectors of size $w \times 1$. In each row, $hw - 1$ of these hw vectors are zero. The remaining vector is assigned the value of G_j. It may be noted that $\mathcal{G}$ is highly sparse and the non-zero elements can be efficiently computed.

Novel views can also be generated in a similar manner for different placements of the camera and the corresponding values of a and b. A new variable i is introduced into the notation of the interpolation vector C_j^l to index the novel views. Thus, the view I_i is synthesized using the interpolation vector C_{ij}^l as

$$\mathcal{G}\left[\,C_{i1}^1 \ldots C_{i1}^w \ldots C_{ih}^1 \ldots C_{ih}^w\,\right] = \mathcal{G}\mathcal{C}_i \tag{5}$$

Eigen space representation of the set of images involves computation of eigenvectors of AA^T as explained in Section 2. The data matrix A is rearranged as product of two matrices,

$$A = [I_1, I_2, \ldots I_P] = \mathcal{G}[\mathcal{C}_1\mathcal{C}_2 \ldots \mathcal{C}_P] = \mathcal{G}\mathcal{C} \tag{6}$$

where $\mathcal{C}$ is obtained by arranging $\mathcal{C}_1, \ldots, \mathcal{C}_P$ as columns of a matrix. Even though the dimension of these two matrices are huge, the number of operations required to compute the product is small. This is so because both $\mathcal{G}$ and $\mathcal{C}$ are highly sparse.

Since we are considering camera rotation around one axis, let α and θ be the angles between the camera plane for generating I_2 and I_3 from that of I_1. Given these angles, Ullman and Basri [12] show that

$$a = \frac{\sin(\alpha - \theta)}{\sin(\alpha)} \quad \text{and} \quad b = \frac{\sin(\theta)}{\sin(\alpha)}. \tag{7}$$

Given the two input views I_1 and I_2, and the angle α, a and b are computed for different values of θ using Equation (7). The choice of resolution of θ depends on the tradeoff between density of eigenspace needed and the computational effort needed. Let the value of a and b corresponding to $\theta = \theta_i$ be denoted by a_i and b_i respectively. For each pair (a_i, b_i), C_{ij}^l is computed for each pixel in the novel image using Equation (4). From the intensity values of the images and C_{ij}^l, the matrices $\mathcal{G}$ and $\mathcal{C}$ are generated and then the eigenvectors corresponding to A are computed as explained above.

4 Extension to Other Camera Models

Affine Camera The process of construction of eigen space for an object imaged with an affine camera is similar to an object imaged with an orthographic camera. The difference is in the determination of the constants a and b, which is done as follows. We assume that the world is imaged with an affine camera P_1 to generate image I_1. The world is imaged again after transforming P_1 with a transformation

T_1, which is rotation by an angle α, to yield I_2. For a given θ, we compute T_2, the rotation matrix about Y-axis by θ. Now we assume the world is imaged after P_1 is transformed by T_2 to yield I_3. Since Equation (2) holds for corresponding points in the three views, we obtain a system of equations in terms of a, b, P_1, T_1 and T_2. These system of equations can be solved to obtain the values of a and b. Using this procedure, a_i's and b_i's are generated and the remaining steps are followed as for an orthographic camera.

Perspective Camera. The novel view synthesis with general perspective cameras has been explored in [13]. In this technique, trilinear relationships between two views are created and then a tensorial operator is used to create the tensor relating the desired and two initial views. This tensor (say α_k^{ij}) is later used to get the coordinates of a point in new views using the coordinates in the initial view. The relation used is as follows

$$x'' = \frac{x'\alpha_i^{31}p^i - \alpha_i^{11}p^i}{\alpha^{13}p^i - x'\alpha_i^{33}p^i}, y'' = \frac{y'\alpha_i^{32}p^i - \alpha_i^{12}p^i}{\alpha^{13}p^i - y'\alpha_i^{33}p^i}.$$

This relation can be used to modify *dist* as $dist = \sqrt{(x'' - x)^2 + (y'' - y)^2}$, where x'' and y'' are given by above equations. Note that in this case, since the y-coordinate changes as well, the process for creating one pixel would be modified to include the whole image (as raster scanned row vector), instead of just one row as earlier. This would also result in an increase in the coefficient vector of one pixel. This coefficient vector would correspondingly be of $hw \times 1$ size as well. The rest of the process of stacking up pixel to build $\mathcal{G}$ and $\mathcal{C}$ matrices remains the same except for the above change. In this case, the dimensionality of $\mathcal{G}$ and $\mathcal{C}$ would also increase to $hw \times h^2w^2$ and $h^2w^2 \times n$ respectively. However the matrices are still highly sparse and their product can be computed efficiently. It should be noted that perspective novel view synthesis (NVS) relationship is general and is applicable for general motion and not just y-axis rotation as in the previous cases.

4.1 Eigenspace Construction from More Than Two Views

If there are more than two views, even then the eigenspace can be constructed with minor modifications to the algorithm described above. Let there be m views of an object, $I_1 \dots I_m$, imaged with an orthographic or affine camera. We will get $(m-1)$ pairs of consecutive views. Also, let α_n be the angle between the camera plane for I_n and I_{n+1} , $1 \leq n \leq m-1$. For every pair of consecutive views, the matrices $\mathcal{G}$ and $\mathcal{C}$ are computed. For the nth pair of consecutive views, let these be denoted by $\mathcal{G}_n$ and $\mathcal{C}_n$. The data matrix A will be defined as

$$A = [\mathcal{G}_1\mathcal{C}_1 \ \dots \mathcal{G}_{m-1}\mathcal{C}_{m-1}] = \mathcal{G}_1[\mathcal{C}_1, \frac{\mathcal{G}_2}{\mathcal{G}_1}\mathcal{C}_2 \dots \frac{\mathcal{G}_{m-1}}{\mathcal{G}_1}\mathcal{C}_{m-1}]$$

$$= \mathcal{G}_1[\mathcal{C}_1, \widetilde{\mathcal{C}}_2 \dots \widetilde{\mathcal{C}}_{m-1}] = \mathcal{G}\mathcal{C}.$$

Since we want to express the data matrix as the product of two matrices, $\mathcal{G}$ and $\mathcal{C}$, $\mathcal{G}_1$ is brought out as a common factor. After bringing out $\mathcal{G}_1$ as the common factor, we get $\mathcal{G}_n\mathcal{C}_n = \mathcal{G}_1\widetilde{\mathcal{C}}_n$. To compute $\widetilde{\mathcal{C}}_n$, we multiply with the pseudo-transpose of $\mathcal{G}_1$ on both sides of the above equation and compute the matrix $\widetilde{\mathcal{C}}_n$. Having arranged A as a product of the matrices $\mathcal{G}$ and $\mathcal{C}$, the eigenvectors are computed.

4.2 Application to Object Recognition

Suppose there are m objects, that need to be recognized. Also, assume there are at least two views for each object and the angle by which the camera has been rotated for each pair of consecutive views is known. For the nth object, $\mathcal{G}_n$ and $\mathcal{C}_n$, are computed, $1 \leq n \leq m$. The data matrix A is arranged as a product of $\mathcal{C}$ and $\mathcal{G}$, by arranging all the $\mathcal{G}_n$ and $\mathcal{C}_n$, $1 \leq n \leq m$, and then performing the necessary transformations. The eigen subspace for A is constructed by computing its eigenvectors and discarding eigenvectors corresponding to lower eigenvalues. Recognition is performed by projecting a test sample into this space and then classifying.

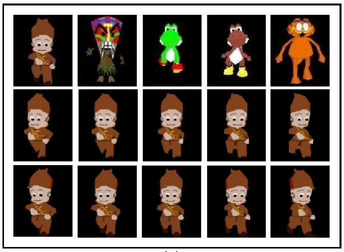
5 Results and Discussions

To analyze the performance of the proposed formulation, we have considered synthetic and real-images. Synthetic models allow us to conduct the experiments in a controlled manner to systematically study the performance. Real-images are taken from Columbia Object Image Library(COIL) [9] and Surrey Object Image Library (SOIL) [10].

In all the experiments, eigen space is created using the method described in the previous section. Though the eigen space can be constructed from any arbitrary views, for better analysis and understanding, we use input images from cameras rotated around Y axis and separated by α (explained later). We then construct eigen space corresponding to images at θ from the first input image. For example, θ may take values ranging from -20° to $+20^\circ$ at increments of 1°. This means that the eigen space is created for a 40° view cone with a resolution restricted by the 1° difference between consecutive views.

5.1 Synthetic Models

The proposed scheme is useful when a reasonably accurate eigenspace is required when there are few input images available for each object. We conducted experiments with synthetic models to test the accuracy of the eigenspace built with the proposed method. We used the eigenspace for recognition problem. Five synthetic objects shown in Fig. 2 (i):(a-e) are considered. These models(O1-O5), were imaged by rotating the camera around the object about the Y-axis. Out of these, two images were taken for construction of the eigen space. From these two images an eigen space corresponding to the images in the range -20 to $+20$ from the first image (out of these two) is constructed. We found that even in the presence of similar objects (Fig. 2(i):(c) and (d)), the eigenspace constructed as per the proposed scheme gives 100% accuracy for a variety of cases (illustrated



(i)

α	N	O1	O2	O3	O4	O5	Total(%)
5^o	2	100	100	100	100	100	100
10^o	2	100	100	100	100	100	100
15^o	2	100	100	100	100	96.66	99.33
20^o	2	100	100	100	82.22	100	96.44
30^o	2	100	100	91.11	77.77	100	93.77
20^o	3	100	100	100	100	100	100
30^o	3	100	100	100	100	100	100

(ii)

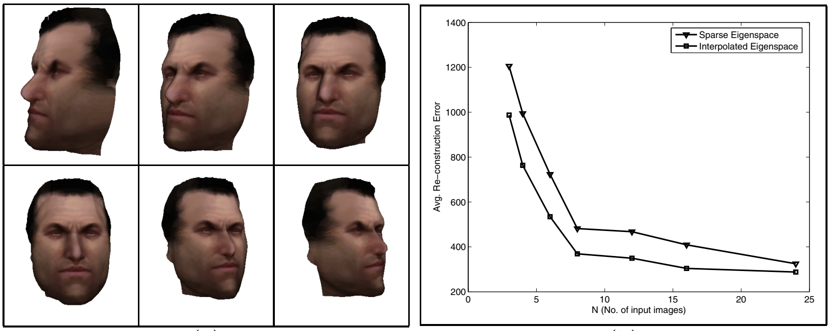
Fig. 2. (i): The first row (a-e) shows the five synthetic models considered for the experiment. Second row (f-j) shows the novel views of the Policeman(object 1) corresponding to the eigen space representation constructed. Final row (k-o) shows the images which a real-camera could have seen. (ii) Recognition results on objects in (a). As the angle between two views(α) is increased, there is a reduction in performance. But with minimal additional views, this can be compensated. N is the number of views.

in Fig. 2 (ii) and explained in detail below). This experiment underlines the accuracy and utility of the eigenspace created by the proposed scheme.

The test set contained 450 images, 90 images per object. The views ranged from -45^0 to 44^o at steps of 1^o. These images were obtained by explicitly rotating the camera and capturing the views thus obtained. These test images (Fig. 2(k - o)), in fact, deviate from the synthesized images for the same angle(Fig. 2(f - j)). The test set was projected into the eigenspace and classified using a Nearest Neighbor algorithm. Eigen space is spanned by 15 eigen vectors corresponding to the largest eigen values. The recognition accuracy was equal to or near 100% for the various values of α. The results of some experiments on this set are provided in Table 2. When α is increased to $30°$, the accuracy is found to be above 90% with only two views. When three views were used, 100% accuracy was achieved.

This experiment indicates that for large α the accuracy can be further improved by using additional views. The eigenspace created by using more number of seed views provides a better estimation of the true eigenspace. Although the accuracy with two seed views is high, additional views can be used to improve the performance to suit practical applications.

Interpolated eigenspace Vs Sparse eigenspace. The motivation behind the current work is to capture the information contained in views from different orientation by doing Novel View Synthesis (NVS). In many real life applications very few images per object are available. NVS allows creation of a *denser* set of images from a relatively sparse set. Intuitively such an interpolated eigenspace (created by NVS) would be *closer* to the actual eigenspace of an image (which could be created by exhaustively taking images at various orientation). We validated this intuition by considering a 3D Face model (see Fig. 3 (a)). We considered 48 equally spaced images of this 3D Face model by moving the camera around the model from $-60°$ to $+60°$. It can be assumed that these 48 images make up the *true* eigenspace of this object. N ($\in \{4, 6, 8, 12, 16, 24\}$) equally spaced images



(a) (b)

Fig. 3. (a) Few images of the 3D Face model used to compare the sparse eigenspace with interpolated eigenspace. (b) A comparison between the average re-construction error per image for sparse eigenspace and interpolated eigenspace.

were selected from this set to create a sparse set of images. Each image was segmented to a uniform size of 70×60 pixels. We built an eigenspace limited to this sparse set and calculated the average re-construction error for each of the 48 original images. We performed NVS on this sparse set and synthesized $25 \times \frac{48}{N}$ new images between every pair of consecutive images in this sparse set. We created an eigenspace with these images and calculated the re-construction error for this *interpolated* eigenspace. In both cases top eigenvectors corresponding to energy factor $k = 0.99$ ($k = \sum_i^k \sigma_i / \sum_i^T \sigma_i$, where σ_is are the eigenvalues and T is the total number of eigenvalues) were picked. For the intermediate set of $N = 6$ initial images the error with interpolated images was 534.4 while with sparse eigenspace, it was 723.3. This error is over a total of $60 \times 70 = 4200$ pixels. In Fig. 3 (b) more detailed comparison of the re-construction errors between sparse and interpolated eigenspaces is shown. Clearly, the interpolated eigenspace represents the appearance of objects better than the sparse eigenspace.

5.2 Application to Recognition

We conducted recognition experiments on two datasets (e.g. COIL-20 [9], SOIL-47 [10]) to show the applicability of the proposed eigenspace in the context of object recognition. Good performance with recognition accuracy in the range of $> 90\%$ was achieved.

Experiments on COIL images. We verified the performance of the proposed scheme on COIL [9] images (Fig. 4 (a)). The Columbia Object Image Library was developed for conducting experiments on object recognition [5]. This database has been widely used by researchers for verification of object recognition algorithms. This library provides images of objects rotated about the Y-axis at intervals of $5°$.

The training set consisted of 180 images, with views ranging from -20^o to 20^o (inclusive) at steps of 5^o. In fact, only two images per object are used for

(a)

(b)

α	5	10	15	α	5	10	15	α	5	10	15
O1	100	100	100	O8	100	100	100	O15	100	84.21	73.68
O2	100	100	100	O9	100	89.47	84.21	O16	100	100	100
O3	100	100	100	O10	100	100	100	O17	89.47	100	100
O4	100	100	100	O11	100	100	100	O18	100	73.68	84.21
O5	73.68	100	100	O12	100	73.68	85.73	O19	100	100	100
O6	100	68.42	84.73	O13	94.74	100	89.47	O20	100	94.74	73.68
O7	100	84.21	52.63	O14	100	100	100	Total(%)	97.89	93.42	91.42

Fig. 4. (a)All COIL-20 Objects used for the recognition experiment. (b) Recognition results on COIL images. With only two images, most of the images in 90° view cone can be recognized with good accuracies. With minimal increase in the number of images, 100% accuracy is achieved. See text for details.

constructing the eigen space. i.e., in reality only 40 images were used to get an effective representation, which we could have computed from all the 180. The test set consisted of 380 images, 19 images for each object at increments of 5^o from -45^o to 45^o (inclusive). The results of the experiment on the real world images are provided in Fig. 4 (b).

The recognition accuracy ranging from 88% to 95%, is achieved by using two images per object. This outperforms the direct methods for recognition by explicitly constructing eigen spaces. Crowley [14] achieved around 90% recognition from four views. Their results were verified only for a smaller view cone, compared to that of ours. Additionally, we have exhaustively tested images beyond the scope of the training images. i.e., even if we construct the eigen space for $-20°$ to $+20°$, applicability was verified for $-45°$ to $+45°$. In general, the proposed method is found to give good results even for such test images. Additionally the recognition accuracy is found to improve if the number of initial seed images is increased.

Experiments on SOIL images. The Surrey Object Image Library (SOIL [10]) consists of 25 planar and 22 complex shaped objects (some of which are shown in Fig. 5 (a)). This dataset is also widely used for testing recognition performance in literature along with COIL-20. There are 2 sets, SOIL-47A and SOIL-47B which differ in overall illumination. We conducted our experiments on SOIL-47A. There are 20 images per object taken at approximately 9° intervals spanning 180°.

We considered 2 images of all 47 objects in the 90° middle sector. Here we varied initial angle α as 9° and 18° and generated 18 $(2\alpha, \alpha = 9)$ and $36(2\alpha, \alpha = 18)$ images respectively per object, separated at 1° intervals spanning the central

(a)

(b)

α	9	18	α	9	18	α	9	18	α	9	18
O1	83.66	83.66	O13	100	83.33	O25	100	83.33	O37	100	100
O2	66.67	66.67	O14	83.33	100	O26	100	100	O38	100	100
O3	100	83.33	O15	100	83.33	O27	100	100	O39	100	100
O4	100	100	O16	66.67	83.33	O28	66.67	83.33	O40	100	100
O5	100	100	O17	100	83.33	O29	100	100	O41	100	100
O6	66.67	83.33	O18	83.33	83.33	O30	100	100	O42	100	100
O7	83.33	66.67	O19	100	100	O31	100	100	O43	66.67	83.33
O8	100	83.33	O20	83.33	83.33	O32	100	100	O44	100	100
O9	100	100	O21	83.33	66.67	O33	100	100	O45	100	100
O10	83.33	66.67	O22	83.33	83.33	O34	66.67	83.33	O46	100	100
O11	66.67	100	O23	83.33	83.33	O35	66.67	83.33	O47	100	100
O12	83.33	100	O24	100	83.33	O36	100	100	Total(%)	90.43	90.07

Fig. 5. (a) Few images taken from the SOIL-47 database. (b) Recognition results on SOIL-47 images. With only two images, most of the images in the 90° cone are recognized. See text for details.

2α section. These ($47 \times 18 = 846$, $47 \times 36 = 1692$) images were taken as the training set. The testing set consisted of images in the range -2α to 2α (5 and 7 images respectively per object for 47 objects). Recognition accuracies above 90% validate the use of the proposed technique. More detailed results are in Fig. 5 (b). The recognition experiments show that the eigenspace estimated from view in a limited range can be used to recognize the views outside the range. It is evident that the multi-view relationships can be used to enhance appearance models to enable view independent recognition of objects.

6 Conclusion

The major contribution of this work is in construction of eigenspace from limited number of views. The algorithms proposed for the construction of eigenspace involve matrices that are very sparse. Efficient algorithms for performing operations on sparse matrices are used for implementation. A detailed analysis of the computational complexity of these algorithms is beyond the scope of the current work. The reconstruction error per pixel for various camera models are found to be less than 1%, validating the correctness of eigenspace construction process. Further, the recognition experiments conducted on both synthetic and real world data ascertain that the approach presented can be used to build view independent recognition systems. Future work would focus on the applicability of this for accurate pose estimation from limited views for deformable objects.

Acknowledgments

One of the authors, Paresh K. Jain, would like to acknowledge the financial support provided by GE (through the GE Foundation Scholar-Leaders Program 2004-06) while carrying out this work.

References

1. Wang, Y., Chua, C.S.: Face Recognition Across Views From 2D and 3D Images. Asian Conference on Computer Vision **2** (2004) 730–735
2. Mori, G., Malik, J.: Recognizing Objects in Adversial Clutter: Breaking a visual CAPTCHA. IEEE Conference on Computer Vision and Pattern Recognition **1** (2003) 134–144
3. Liu, X., Srivastava, A., Gallivan, K.: Optimal Linear Representation of Images for Object Recognition. IEEE Conference on Computer Vision and Pattern Recognition **1** (2003) 229–234
4. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. IEEE Conference on Computer Vision and Pattern Recognition (1994) 84–91
5. Nayar, S.K., Nene, S.A., Murase, H.: Real Time 100 Object Recognition System. IEEE Int'l Conference on Robotics and Automation **3** (1996) 2321–2325
6. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. IEEE Conference on Computer Vision and Pattern Recognition (1991) 586 – 591
7. Cootes, T., Wheeler, G., Walker, K., Taylor, C.: Coupled View Active Appearance Models. British Machine Vision Conference **1** (2000) 52–61
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)
9. Nene, S.A., Nayar, S.K., Murase, H.: Columbia Object Image Library(COIL-20). Technical Report CUCS-005-96 (1996)
10. J. Burianek, A. Ahmadyfard, J. Kittler: SOIL-47, The Surrey Object Image Library, Centre for Vision, Speach and Signal processing, Univerisity of Surrey. (at `http://www.ee.surrey.ac.uk/Research/VSSP/demos/colour/soil47/`)
11. Chatterjee, S., Banerjee, S., Biswas, K.K.: Reconstruction of Local Features for Facial Video Compression. Int'l Conf. on Image Processing **2** (2000) 211–214
12. S.Ulmann, R.Basri: Recognition by Linear Combination of Models. IEEE Trans. Pattern Anal. Machine Intell. **13** (1991) 992–1006
13. Avidan, S., Shashua, A.: Novel view synthesis by cascading trilinear tensors. IEEE Transactions on Visualization and Computer Graphics **4** (1998) 293–306
14. de Verdiere, V.C., Crowley, J.L.: Visual Recognition Using Local Appearance. European Conference on Computer Vision **1** (1998) 640–654