ROBUST MULTI-DIMENSIONAL NULL SPACE REPRESENTATION FOR IMAGE RETRIEVAL AND CLASSIFICATION

Xu Chen, Dan Schonfeld and Ashfaq Khokhar

Department of Electrical and Computer Engineering, University of Illinois at Chicago
xchen27, dans, ashfaq@uic.edu

ABSTRACT

This paper presents a novel system for image retrieval and classification based on a robust multi-dimensional view invariant representation and a linear classifier algorithm. Specifically, multi-dimensional Null Space Invariant (NSI) matrix representation has been derived. Moreover, the robustness of null space representation has been investigated with perturbation analysis in terms of the error ratio and SNR. The proposed representation is invariant to affine transformations and preserves the null space matrix. We use principal component null space analysis (PCNSA) of the NSI operator for recognition, indexing, and retrieval of 2D and 3D images. We rely on PCNSA to determine the distance of a query image to the centroid of the class, which is a statistical information vector in the PCNSA algorithm representing the corresponding class of the object. Our results shows that NSI provides a robust and powerful approach to image recognition and classification even when the query image or the stored image has been subjected to unknown affine transformations due to camera motions.

Index Terms— Image Retrieval, Affine View Invariant, Perturbation Analysis, Principal Component Analysis, Image Recognition and Classification

1. INTRODUCTION

Developing an efficient indexing and retrieval system for a large collections of images is a challenging task in the presence of affine transformations [1]. In this paper, we present a novel system for image retrieval and classification based on null space invariants (NSI). In [2], the mathematical form of null space invariants has been derived. In this paper, we rely on the theoretical formulation of null space invariants to demonstrate its enormous potential in image retrieval and classification. Specifically, we demonstrate the invariance of the null space representation of 2D images produced from objects with different unknown camera orientation. We present the invariance of the null space representation of 3D images with unknown rotation and translation. We subsequently investigate the robustness of the proposed approach to noisy data by using tools from perturbation theory. Finally, we demonstrate the performance of the proposed approach in the application of image indexing, retrieval and classification. As far as we know, this is the first use of null space invariants in multi-dimensional image-based classification and retrieval applications.

2. MULTI-DIMENSIONAL NULL SPACE REPRESENTATION

Let $P_i = (x_i, y_i, z_i)$ for $i = 0, ..., n-1$ be an ordered set of n non-coplanar points in R^3 for 3D objects, and consider the $4 \times n$ matrix M given by

$$M = \begin{pmatrix} x_0 & x_1 & ... & x_{n-1} \\ y_0 & y_1 & ... & y_{n-1} \\ z_0 & z_1 & ... & z_{n-1} \\ 1 & 1 & ... & 1 \end{pmatrix} . \tag{1}$$

For a 3-D object whose feature set consists of the points $P_0, ..., P_{n-1}$, an (n-4)-dimensional linear subspace K^{n-4} of R^n can be represented as:

$$\begin{aligned} &K^{n-4} = \{p = (p_0, ..., p_{n-1})^T \in R^n \\ &Mp = (0,0,0,0)^T\} , \end{aligned} \tag{2}$$

where K^{n-4} is regarded as affine invariant. Moreover, since $K^{n-4} \subset H^{n-1}$, the points are determined by K^{n-4} in the Grassmannian, $Gr_R(n-4, n-1)$ of (n-4)-planes in (n-1)-space, which is a well understood manifold of dimension 3n-12. One can show that K^{n-4} is spanned by the vectors $v_i = (p_0^{(i)}, ..., p_{n-1}^{(i)})^T$, $i = 4, ...n-1$, where

$$p_0^{(i)} = -det \begin{pmatrix} x_1 & x_2 & x_3 & x_i \\ y_1 & y_2 & y_3 & y_i \\ z_1 & z_2 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} , \tag{3}$$

$$p_1^{(i)} = det \begin{pmatrix} x_0 & x_2 & x_3 & x_i \\ y_0 & y_2 & y_3 & y_i \\ z_0 & z_2 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} , \tag{4}$$

$$p_2^{(i)} = -det \begin{pmatrix} x_0 & x_1 & x_3 & x_i \\ y_0 & y_1 & y_3 & y_i \\ z_0 & z_1 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} , \quad (5)$$

$$p_3^{(i)} = det \begin{pmatrix} x_0 & x_1 & x_2 & x_i \\ y_0 & y_1 & y_2 & y_i \\ z_0 & z_1 & z_2 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} , \quad (6)$$

$$p_i^{(i)} = -det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix} , \quad (7)$$

and $p_i^{(j)} = 0$, j=4,...,i-1,i+1,...,n-1. Let us assume the noisy data as $\tilde{P}_i = (x_i+\epsilon_{x,i}, y_i+\epsilon_{y,i}, z_i+\epsilon_{z,i})$, where $\epsilon_{x,i}, \epsilon_{y,i}, \epsilon_{z,i}$ have Gaussian distribution with zero mean and the variance δ^2. Relying on the first order perturbation, we obtain the difference of the perturbed null operator $\tilde{p}_0^{(i)}$ and the true null operator:

$$\tilde{p}_0^{(i)} - p_0^{(i)} = -[det \begin{pmatrix} \epsilon_{x,1} & \epsilon_{x,2} & \epsilon_{x,3} & \epsilon_{x,i} \\ y_1 & y_2 & y_3 & y_i \\ z_1 & z_2 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} + det \begin{pmatrix} x_1 & x_2 & x_3 & x_i \\ \epsilon_{y,1} & \epsilon_{y,2} & \epsilon_{y,3} & \epsilon_{y,i} \\ z_1 & z_2 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} + det \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ \epsilon_{z,1} & \epsilon_{z,2} & \epsilon_{z,3} & \epsilon_{z,i} \\ 1 & 1 & 1 & 1 \end{pmatrix}] . \quad (8)$$

$$E\|\tilde{p}_0^{(i)} - p_0^{(i)}\|_F^2 = \sum_{j,k,m} [(det \begin{pmatrix} y_j & y_k & z_m \\ z_j & z_k & z_m \\ 1 & 1 & 1 \end{pmatrix})^2 + (det \begin{pmatrix} x_j & x_k & x_m \\ y_j & y_k & y_m \\ 1 & 1 & 1 \end{pmatrix})^2 + (det \begin{pmatrix} x_j & x_k & x_m \\ z_j & z_k & z_j \\ 1 & 1 & 1 \end{pmatrix})^2]\delta^2 , \quad (9)$$

where $j = 1,2; k = 2,3; m = 3, i$ and $j \neq k \neq m$. We evaluate the sensitivity of the system by the output noise produced from the unit input noise, namely, the error ratio $\tau = \frac{E\|\tilde{Q}-Q\|_F^2}{E\|Z\|_F^2}$, where Z is the $4 \times N$ input noise matrix and the perturbed matrix $\tilde{Q} = [\tilde{v}_4, ..., \tilde{v}_n]$. Therefore, we have:

Lemma 1:

$$E\|\tilde{Q} - Q\|_F^2 = \{(n-4) \sum_{j,k,m} [(det \begin{pmatrix} x_j & x_k & x_m \\ y_j & y_k & y_m \\ 1 & 1 & 1 \end{pmatrix})^2 + (det \begin{pmatrix} x_j & x_k & x_m \\ 1 & 1 & 1 \\ z_j & z_k & z_m \end{pmatrix})^2 + (det \begin{pmatrix} 1 & 1 & 1 \\ y_j & y_k & y_m \\ z_j & z_k & z_m \end{pmatrix})^2] + 2 \sum_{i=4}^{n-1} \sum_{p,q} [(det \begin{pmatrix} x_p & x_q & x_i \\ y_p & y_q & y_i \\ 1 & 1 & 1 \end{pmatrix})^2 + (det \begin{pmatrix} x_p & x_q & x_i \\ 1 & 1 & 1 \\ z_p & z_q & z_i \end{pmatrix})^2 + (det \begin{pmatrix} 1 & 1 & 1 \\ y_p & y_q & y_i \\ z_p & z_q & z_i \end{pmatrix})^2]\}\delta^2 , \quad (10)$$

where $j = 0,1; k = 1,2; m = 2,3$ and $j \neq k \neq m$, $p = 0,1,2; q = 1,2,3$ and $p \neq q$. Thus, the error ratio can be computed as

$$\tau = \frac{E\|\tilde{Q} - Q\|_F^2}{E\|Z\|_F^2} = \frac{E\|\tilde{Q} - Q\|_F^2}{3N\delta^2} . \quad (11)$$

It can be seen from the equation (11) that the error ratio only depends on the system itself and independent of noise. Besides the error ratio, SNR is also widely used to compare the level of desired signal to the level of background noise. We can define the SNR of the system with different matrix norms. Here we choose F norm: $SNR = \frac{E\|Q\|_F^2}{E\|\tilde{Q}-Q\|_F^2}$. Based on (3)-(7) and (10), the exact expression of SNR can be obtained. With a similar approach as 3D images, the error ratio and SNR for 2D images can be computed. Moreover, it is easy to extend the result of the error ratio and SNR into multi-dimensions. Relying on basic perturbation theorem [3], we obtain:

Lemma 2:

$$\frac{\|Q\|}{\|\tilde{Q} - Q\|} \geq \frac{1}{\|M^{-1}Z\|} , \quad (12)$$

$$\frac{\|\tilde{Q} - Q\|}{\|Z\|} \leq \|M^{-1}\|\|Q\| , \quad (13)$$

where $\|.\|$ denotes a matrix norm and a consistent vector norm. In general, the inequalities (12) and (13) can be considered as the bound for the relative error of the multi-dimensional null space invariants.

3. VIEW INVARIANT FEATURE REPRESENTATION FOR RETRIEVAL AND CLASSIFICATION

The choice of a low-dimensional but highly discriminatory feature space is crucial to developing efficient and scalable indexing and classification system. We use scale invariant feature transform (SIFT) to obtain feature points based on SIFT algorithm in [4]. SIFT extracts and connects feature points in images which are invariant to image scale, rotation and changes in illumination. Since the feature points of images in a class may have different numbers, we normalize the length of the vector formed by feature points by selecting the minimum number of feature points from all the images.

Once we generate the feature points in each image, we compute the null space representation $NSI_{n\times(n-3)}$ for 2D images or $NSI_{n\times(n-4)}$ for 3D images based on the null operator. We can rely on numerous methods for indexing and classification. We choose a method for dimensionality-reduction and classification based on PCNSA [1]. Note that the term null space used in PCNSA refers to Approximate Null Space (ANS) for representing each class which is computed using minimal eigenvectors within the class and thus minimizes the intra-class variance. However, this process is not intended to capture the null operator and is unrelated to the null space invariant proposed in Section 2. We first convert NSI into $P = n(n-3)$ or $P = n(n-4)$ column vector Y_p which is assumed in the class C_i. We further denote μ_i as the class mean and $\Sigma_{full,i}$ as the class covariance matrix and assume there are C classes and each class has K images. Then we apply **PCNSA Algorithm** as follows:

1. **Obtain PCA space:** Evaluate the total covariance matrix Σ_{full} which is given by $\Sigma_{full} = \Sigma_{full,w} + \Sigma_{full,b}$, where the within class covariance matrix $\Sigma_{full,w} = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{K}\sum_{i=1}^{K}[Y(i,k) - \mu(i)][Y(i,k) - \mu(i)]^T$ and the between class covariance matrix $\Sigma_{full,b} = \frac{1}{C}\sum_{i=1}^{C}(\mu_{full,i} - \mu_{full})(\mu_{full,i} - \mu_{full})^T$, then apply PCA to the Σ_{full} to find $(W_{PCA})_{P\times L}$, whose columns are the L leading eigenvectors.

2. Project the data vectors, class means, and class covariance matrices into the corresponding data vectors, class means, and class covariance matrices in the PCA space.

$$(X)_{L\times 1} = W_{PCA}^T(Y - \mu_{full}) \quad (14)$$
$$(\mu_i)_{L\times 1} = W_{PCA}^T(\mu_{full,i} - \mu_{full}) \quad (15)$$
$$(\Sigma_i)_{L\times L} = W_{PCA}^T \Sigma_{full,i} W_{PCA} \quad (16)$$

3. **Obtain ANS:** Find the Approximate Null Space, $(N_i)_{L\times M_i}$ for each class i by choosing M_i smallest eigenvalues of corresponding eigenvectors.

4. **Obtain Valid Classification Directions in ANS:** Say $N_i = (e_{i,1}|e_{i,2}|...|e_{i,M_i})_{L\times M_i}$. Any direction e_i which satisfies $|(\mu_i - \mu_j)e_i| > \delta_2\|\mu_i - \mu_j\|$, is assumed a valid direction and used to build valid ANS, $W_{NSA,i}$.

5. **Classification:** PCNSA finds distances from a query image to all the classes $d_i(X) = \|W_{NSA,i}(X - \mu_i)\|$. We choose the smallest distance for classification of X.

Fig. 1. 24 different face poses used for each face from the UMIST face database.

4. EXPERIMENTAL RESULTS

Traditionally, the 2D feature points in images and 3D coordinates in objects are associated with 3×4 projection matrix and many research work focus on designing projection matrix [5], which incorporates three coordinate systems: camera, image and world. While our retrieval and classification system relying on null space invariants does not require any information of the projection matrix. The applications of our system are twofold: 1) retrieval and classification for 2D images with unknown views 2) retrieval and classification for 3D images with unknown translation and rotation. For simplicity, we demonstrate the performance of our system with 2D images. The proposed classification and retrieval system is implemented and evaluated using the UMIST database [6] as in Figure.1, and available at $http://images.ee.umist.ac.uk/danny/database.html$ and consists of facial images of 20 people, namely, 20 different classes. In each class, we choose 24 images, captured from different poses. We extract 36 feature points from each image with SIFT. Subsequently, the NSI for these feature points are computed, which is invariant to affine transformation since it preserves the values against rotation and transformation. For all the simulations, we have chosen random noise with zero mean and variance $\delta_1^2 = 0.1$ for noisy image data, $\delta_2 = 5 \times 10^{-5}$ and 35 largest eigenvectors in obtaining the PCA space. Figure.2 depicts accuracy of the proposed classification system versus number of classes for perfect images and noisy images. Simulation shows that our systems preserves its efficiency and robustness even with large number of classes and noisy data. Figure.3 depicts accuracy values versus increase in the number of images in a class for perfect images and noisy images. Simulation results show that the performance of our system deteriorates slightly with the increase in the number of images within a class. This problem can be resolved by using a recursive partition, where we repeatedly divide all the images into smaller groups until each image in the group contains enough and consistent

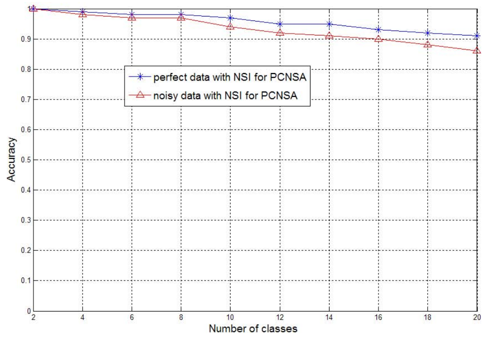

Fig. 2. Accuracy of classification with increasing number of classes.

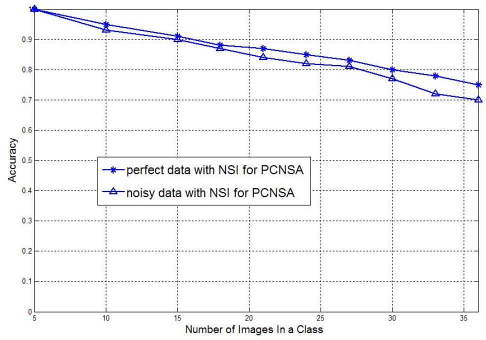

Fig. 3. Accuracy for classification with increasing number of images within the class.

feature points. Recursive partitions stop until the subtractions of the angles of the faces are smaller than 30 degrees. We can utilize null space representation for each class in the nested hierarchical tree structure to obtain a scalable representation with robust performance when the number of images in the system increases. Figure.4 shows PR curves for indexing and retrieval problem for 18 different classes of images, each class having 20 images. We compute the distance of the query image to any other image using PCNSA on NSI as $D(X_i, W) = \|W_{NSA,i}(X_i - W)\|$, where W is the query image. This distance is then used to find γ nearest images, where γ is a user specified parameter. There are 3 curves in figures. One corresponds to directly using PCA on NSI for retrieval, which has the worst performance. The other two curves correspond to using PCNSA on NSI for perfect and noisy data, both of which are much superior compared to the curve without using PCNSA on NSI.

5. CONCLUSION

We have demonstrated the efficiency and robustness of the NSI operator as a powerful tool for image classification

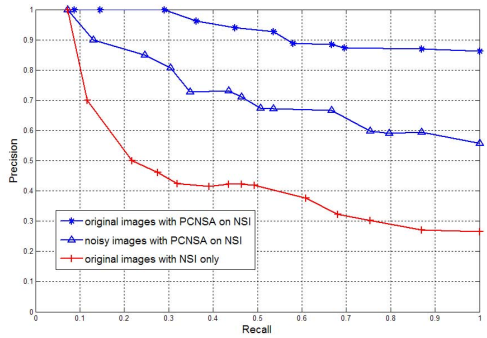

Fig. 4. Precision-Recall curve for image retrieval using PCNSA on NSI for original and noisy images.

and retrieval. We have derived the error ratio and SNR for 3D NSI operators with perturbation analysis, which can easily be extended into n-dimensions. Moreover, we have proposed the general bound for the relative error based on multi-dimensional NSI operators. Note that any classification algorithms can be used relying on the PCA presentation of the multi-dimensional NSI operators. We have further demonstrated the performance of retrieval and classification of perfect and noisy images based on our algorithm.

6. REFERENCES

[1] N.Vaswani and R.Chellappa, “Principal components null space analysis for image and video classification,” *IEEE Trans. Image Processing*, 2006.

[2] P.F.Stiller, C.A.Asmuth, and C.S.Wan, “A general theroy of single view recognition- the affine case- with applications to indexing image databases for content based retrieval,” 1994, International Journal of Computer Vision.

[3] G.W.Stewart, “Matrix algorithms volume i: Basic decompositions,” 1998.

[4] D.Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, 2004, vol. 60, pp. 91–110.

[5] R.Mohamed, A.Ahmed, A.Eid, and A.Farag, “Support vector machines for camera calibration problem,” International Conference on Image Processing, 2006, pp. 1029–1032.

[6] D.Graham and N.Allinson, “Characterizing virtual eigensignatures for general purpose face recognition,” in Face Recognition: From Theory to Application, 1998, pp. 446–456, New York: Springer.