Examples of Set-to-Set Image Classification

By J.M. Chang, M. Kirby, H. Kley, C. Peterson† AND R. Beveridge, B. Draper‡

Abstract

We present a framework for representing a set of images as a point on a Grassmann manifold. A collection of sets of images for a specific class is then associated with a collection of points on this manifold. Relationships between classes as defined by points associated with sets of images may be determined using the projection F-norm, geodesic (or other) distances on the Grassmann manifold. We present several applications of this approach for image classification.

1. Families of Patterns and the Grassmann Manifold

A collection of patterns with a common characterization may be viewed as a *family of patterns*. For example, snapshots of clouds and sky give rise to a family of images with a common characteristic while at the same time exhibiting significant variations across the images. Alternatively, a set of digital images of faces provides another example of an image family. Again, while members of this family clearly possess distinct features it is sensible to pose the question of whether an image, or set of unlabeled images, belongs to this family. So, we are interested in the question of whether an image or set of images belongs to a certain family of images. We may also consider hierarchical families. For example, the face of a single subject viewed under variations in illumination generates a family of images where the family is defined by the identity of the subject. Now the face recognition problem becomes the classification of image sets, each of which are associated with a specific family. The existence of similar structures across a set, or family, of patterns is often recognized as a footprint of low-dimensionality. The correlations between such images indicate that the full generality of an ambient coordinate system is not necessary. Such sets may be represented via best bases, i.e., bases that are tailored to the family of patterns. For example, for faces one may employ the singular value decomposition, see, e.g., Sirovich & Kirby (1987). An alternative to this absolute representation is the representation of different pattern families in a relative sense, i.e., where a pair of bases is associated with a pair of image sets and is constructed to extract similarities or differences in the image sets. The real Grassmann manifold $G(k,n)$ is the set of of k-dimensional vector subspaces of R^n, as induced by the bases described above, and consequently is a natural place to represent sets of images. For example, the projection F-norm distance between two points $A, B \in G(k,n)$ (i.e., two k-dimensional subspaces of R^n) on a Grassmann manifold is given by $d_k(A,B) = \|\sin\theta_\ell\|_2$ where $\theta_\ell = (\theta_1, \theta_2, \ldots, \theta_\ell)$ is the vector of principal angles between the subspaces A and B. The standard algorithm for computing the principal angles is described in Golub and Van Loan (1996). Additional details may be found in Chang *et al* (2006a) and Chang *et al* (2006b). As we shall see in

† Department of Mathematics, Colorado State University Fort Collins, Colorado, USA (`chang@math.colostate.edu`).
‡ Department of Computer Science, Colorado State University Fort Collins, Colorado, USA

TABLE 1. Example images from the male and female classes.

the applications, it is useful to define a further measure of nearness via the ℓ-truncated pseudo-distance, i.e., for any pair of vector subspaces A, B of R^n, $d_\ell(A, B) = \|\sin\theta_\ell\|_2$ so long as $\ell \leq \min\{\dim A, \dim B\}$. In summary, a set of images may be mapped to a point on a Grassmann manifold through the construction of a basis for the image set.

2. The Two Class Problem

Previously we have focused on the multi-class problem associated with face recognition under variations in illumination Chang *et al* (2006a), Chang *et al* (2006b). As we shall observe, two class problems studied in the manner proposed reveal important and potentially very useful properties of image sets. We begin by describing common features of the two class problems under consideration. In each case we have two sets of images associated with the classes that we denote C^+ and C^-. Further, each of these image sets is separated into a training and testing set. Consider the training set of N^+ images associated with pattern family C^+. Under mild generality assumptions, we may generate an array of points on the Grassmann manifold associated with class C^+ by generating subsets of k points; there are $N^+!/k!(N^+ - k)!$ such points. We may generate a similar set of points associated with class C^- on the Grassmann manifold. Following this strategy, we can also generate testing points for both and use the projection F-norm for example, or the l-truncated measure, to ascertain membership of the testing data.

3. Classification of Gender

It has been established that gender classification of digital images of faces is a tractable problem Cheng, O'Toole & Abdi (2001). Here we explore a slightly different question: Is a set of images drawn from the male or female population? To this end we use frontal images with neutral expressions and a single illumination setting taken from the CMU-PIE database Sim *et al* (2003). The complete image as supplied by CMU is used, and no cropping or resampling is performed. There are 50 men and 17 women in the CMU-PIE database. Therefore, 50 and 17 images are available for training and testing the male class C^+ and female class C^- respectively. The images of men are partitioned into 40 training and 10 test images. The images of women are partitioned into 10 training and 7 test images. For men, training points on the Grassmann manifold are generated by randomly sampling 8 out of 40 images. Thus, there are $76,904,685$ ways to construct labeled points associated with C^+. Test points are generated by randomly sampling 5 out

ℓ	1	2	3	4	5	6	7	8	9	10	mean±std
	Trial number										
1	0	0	0	0	0	0	22.5	0	10	0	3.25±7.4582
2	35	0	0	5	12.5	27.5	22.5	5	10	7.5	12.5±11.9606
3	30	0	0	0	0	27.5	27.5	5	22.5	0	11.25±13.6550

Table 2. Misclassification percentage out of 40 testing sets where 20 are of male class and 20 are of female class. Three male and three female image sets were used for the training set. Classification is based on the ℓ-truncated projection F-norm. The experiment is repeated ten times where the mean and standard deviation is reported in the last column of the table.

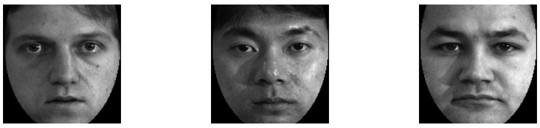
of 10 images: there are 252 possible points associated with C^-. For women, training points are generated by randomly sampling 8 out of 10 images and test points are generated by randomly sampling 5 out of 7 images. Clearly, the particulars of sampling could vary. Example images used to construct points on the Grassmann manifold are shown in Table 1.

As described above, many-to-many set comparisons were carried out for the male-female classification problem where male and female classes are represented by multiple sets of images. We present results for a training set consisting of 3 labeled points associated with each class using a nearest neighbor classifier. The resulting misclassification rates are presented in Table 2 where we vary the number of principal angles used to determine the ℓ-truncated distance. We found that the number of labeled points belonging to each class used on the Grassmann manifold impacts the classification outcome. Therefore ten trials were carried out to give a broader idea of how this set-to-set comparison behaves. A total of 40 testing points (sets of images) were used, where 20 are of male class and 20 are of female class. Admittedly the numbers of points in the testing and training sets are *ad hoc* and additional exploration is warranted.

4. Glasses versus No-glasses Classification

The data set used here is a subset of the lights portion of the CMU-PIE database, where images were captured in neutral expression under a single illumination condition with ambient lights on. Among the 67 subjects in this portion of the CMU-PIE database, 39 subjects are seen without glasses and 28 subjects were seen with glasses. Geometric normalization is performed with this data set. A total of 39 images are available for training and testing for the no-glasses class while 28 images were available for training and testing for the glasses class. For the glasses class, we divide the 28 images into mutually disjoint sets of 20 and 8 for training and testing, respectively. We further construct training sets by randomly selecting 10 images in the list of 20 so that there are $184,756$ ways to construct a set of images for the testing set. Similarly, we construct testing sets by randomly choosing 5 images in the list of 8 so that there are 56 ways to construct this testing set. For the no-glasses class, we divide the 39 images into mutually disjoint sets of 30 and 9 for training and testing, respectively. We then have 30,045,015 and 126 possibilities for the training set and testing sets, respectively. See Table 3 for the example images from the data set that are used to construct the training and testing sets for the glasses and no-glasses classes. We observed that the average classification errors are less than two percent for 3 training points (image sets) in each class and falls to zero with 10 training points per class.

(a)



(b)

TABLE 3. (a) Example images used to construct training and testing sets for the glasses class. (b) Example images used to construct training and testing sets for the no-glasses class.

5. Discussion

We have presented an approach for comparing sets of patterns for similarity. It is assumed that each set has a common attribute. For the illumination problem described in Chang *et al* (2006a), Chang *et al* (2006b) this attribute is the identity of the subject. We have presented additional illustrative examples here. We establish that we can determine whether a set of images of faces are associated with people wearing glasses or not, or whether the gender of the images is male or female. These examples are meant to illustrate how one can ascertain whether a set of data possesses a certain characteristic even though the images in the set have many different features as well. We envision a wide range of applications will fit into this framework.

Acknowledgements: This study was partially supported by the National Science Foundation MSPA-MCS Program (jointly supported by the DMS and DCCF) under award DMS-0434351 as well as the DOD-USAF-Office of Scientific Research under contract FA9550-04-1-0094. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the DOD-USAF-Office of Scientific Research.

REFERENCES

CHANG, J.-M., BEVERIDGE, J.R., DRAPER, B., KIRBY, M., KLEY, H. & PETERSON, C. 2006 Illumination Face Spaces are Idiosyncratic *IPCV 06* **2**, 390-396.

CHANG, J.-M., BEVERIDGE, J.R., DRAPER, B., KIRBY, M., KLEY, H. & PETERSON, C. 2006 A Low-Resolution Illumination Camera *submitted*.

CHENG, Y., O'TOOLE, A. J. & ABDI, H. 2001 Computational Approaches to Sex Classification of Adults' and Children's Faces *Cognitive Science* **25**, 819-838.

GOLUB, G. H. & VAN LOAN, C.F. 1996 Matrix Computations *Johns Hopkins University Press*

SIM, T., S. BAKER, & M. BSAT 2003 The CMU Pose, Illumination, and Expression Database *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25(12)**, 1615 - 1618.

SIROVICH, L. & M. KIRBY (1987), *A low dimensional procedure for the characterization of human faces*, *J. of the Optical Society of America* **A**, **4**, 519.