Class Dependent Kernel Discrete Cosine Transform Features for Enhanced Holistic Face Recognition in FRGC-II

Marios Savvides, Jingu Heo, Ramzi Abiantun, Chunyan Xie and B.V.K. Vijaya Kumar
ECE Department, Carnegie Mellon University, USA
{msavvid@ri, jheo@, raa@, chunyanx@, kumar@ece}.cmu.edu

ABSTRACT

Face recognition is one of the least intrusive biometric modalities that can be used to identify individuals from surveillance video. In such scenarios the users are under the least co-operative conditions and thus the ability to perform robust face recognition in such scenarios is very challenging. In this paper we focus on improving the face recognition performance on a large database with over 36,000 facial images from the Face Recognition Grand Challenge Phase-II data collected by University of Notre Dame. We particularly focus on Experiment 4 which is the most challenging and captured in uncontrolled conditions where the baseline PCA algorithm yields 12% verification rate at 0.1% FAR. We propose a novel approach using class-dependent kernel discrete cosine transform features which improves the performance significantly yielding a 91.33% verification rate at 0.1% FAR, and we also show that by working in the DCT transform domain for obtaining non-linear features is more optimal than working in the original spatial-pixel domain which only yields a verification rate of 85% at 0.1% FAR. Thus our proposed method outperforms the baseline by 79.33% in verification rate @ 0.1% False Acceptance Rate.

1. INTRODUCTION

Robust face recognition technology is in great demand for the ability to create automated facial recognition systems that can look for criminals in a watch-list using traditional camera surveillance infrastructure. However, current facial recognition algorithms have made significant progress but assume some level of user co-operation and the ability to capture a good set of enrollment images. However, it is intertesing and more challenging to work with few enrollment images and try to match faces which are captured in an uncontrolled scenariowhich can include variations in pose, expression, and lighting variations (such as overheard lighting). More importantly, most algorithms have been tested on databases where the number of images used is relatively small to be able to gather statistically significant recognition results for comparison in algorithms. To address this issue and push for development of next generation face recognition algorithms NIST has launched the Face Recognition Grand Challenge(FRGC)[1] which contains a database of images of over 36,000 images collected by the University of Notre Dame[1]. There are 6 experiments in FRGC, one of which is the most challenging where capture data is under uncontrolled conditions which included harsh lighting variations, expressions, pose variations. The challenge is to develop algorithms that can perform robust face recognition at very low false acceptance rate (FAR) conditions, specifically the measured results are the verification rates at 0.1% FAR. The PCA[2][4] baseline algorithm performed by NIST yields 12% at 0.1% FAR.

In our proposed methodology we combine ideas from Support Vector Machines[9-11], Synthetic Discriminant Functions[8], Correlation Filter designs[5][7], Image Transforms[2] and Feature extraction[6] techniques to improve this result to 91.33% at 0.1% FAR. More specifically we propose that extracting non-linear kernel features in the Discrete Transform Domain and using these features in a class-dependent feature analysis framework (CFA)[6] to reduce the dimensionality of the data coupled with on-line discriminant learning using Support Vector Machines leads to obtain the performance shown here. We show that these results greatly improve on the PCA based baseline which is a method to find a subspace for representing images in the least mean squared error sense.

2. FACE RECOGNITION GRAND CHALLENGE DATABASE

The Face Recognition Grand Challenge dataset has been collected at the University of Notre Dame[1] and is split into three datasets and we detail the specifications for Experiment 4.1

- The Generic Training Image set consisting of 222 people with a total of 12,776 images that can be used to build a global face representation (as done in global PCA).
- The Gallery set (also referred to as Target set) consists of ~16,000 facial images of 466 people.

- The Probe set (also referred to as Query set) consists of ~8,000 facial images.

The end goal is to compute a similarity matrix between the Gallery images and Probe images thus yielding a matrix of 16,000x8,000 elements. We then use this matrix to compute the verification rate at 0.1% FAR. Several images of the same person from Experiment 4 can be seen below in Figure 1.

Figure 1: Query images from FRGC Experiment 4 showing the uncontrolled conditions of a person captured under harsh over-head illumination. Variations such as expression, lighting, severe cast-shadows are present.

2. KERNEL DISCRETE COSINE TRANSFORM FEATURES AND HOW TO PERFORM DIMENSIONALITY REDUCTION

In this paper we show that nonlinear feature extracted in the Discrete Transform Domain provides optimal discrimination for face recognition performance on FRGC data. We also perturbed the DCT coefficients of images from the training set in order to generate more synthetic training images, this we show helps outperform performance compared to working in the original image domain. We perform 2D-DCT transformation[2] of a NxM image is defined as follows:

$$F(u,v)=\left(\frac{2}{N}\right)^{\frac{1}{2}}\left(\frac{2}{M}\right)^{\frac{1}{2}}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\Lambda(i)\Lambda(j)\cos\left[\frac{\pi u}{2N}(2i+1)\right]\cdots(1.1)$$
$$\cdots\cos\left[\frac{\pi u}{2N}(2j+1)\right]f(i,j)$$

where

$$\Lambda(\xi)=\begin{cases}\frac{1}{\sqrt{2}} & for\,\xi=0\\ 1 & otherwise\end{cases}$$

The DCT representation decomposes an image into a set of cosine basis functions of different spatial frequencies where to the top left corner represent the lowest spatial frequency content. As shown in Figure 1, we can see that most dominant signal energy is in the lowest frequency components, a well known fact used in JPEG compression methods. However we use all frequency components as for discrimination higher-frequency components provide more image detail that can discriminate between different people.

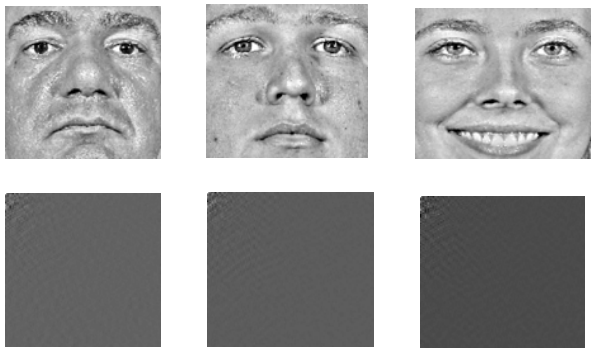

Figure 2: (Top Row) Images from 3 people in FRGC training set (bottom row) Their corresponding 2D-DCT representations.

The above representations do not perform any type of dimensionality reduction, i.e. the DCT components are still of size NxM as the same size the original images are. We can remove some high-frequency components with lowest signal energy, PCA can also be used to do dimensionality reduction, however we propose a novel approach that will perform dimensionality reduction yet preserve data that is most discriminative. We make use of the generic dataset of 222 people with 12,776 images. We use synthetic discriminant functions which are trained to produce specific projection outputs for each class:

- For each class of DCT transformed images we want their projections to yield +1) and all other remaining 221 classes to yield projection value 0.
- We show that we can apply kernel trick to these synthetic discriminant functions to provide higher discriminative power in higher-dimensional mapping space and yield a computationally simple closed form solution.
- We repeat this for all 222 classes to yield 222 Kernel Discrete Cosine Transform Synthetic Discriminants.

Let us define matrix **X** to contain the 2D-DCT features vectorized along the columns for all the generic training set, and define our synthetic discriminant projection as w, and our specified projection coefficients in row vector u. Thus we want to achieve the following:

$$\mathbf{X}^{+}\mathbf{w}=\mathbf{u} \qquad (1.2)$$

However, we would like to constrain **w** to lie in the span of the training images as follows:

$$\mathbf{w}=\mathbf{X}\boldsymbol{\alpha} \qquad (1.3)$$

Substituting Eq. (1.3) in Eq. (1.2) we get

$$\mathbf{X}^{+}\mathbf{X}\boldsymbol{\alpha} = \mathbf{u} \quad (1.4)$$

The Gram matrix $X^{+}X$ can be inverted assuming linear independent images to yield the linear combination coefficients $\boldsymbol{\alpha}$ which are then used to define the synthetic discriminant projection vector **w:**

$$\mathbf{w} = \mathbf{X}(\mathbf{X}^{+}\mathbf{X})^{-1}\mathbf{u} \quad (1.5)$$

To show how we can apply the kernel trick we define the projection coefficient c as follows given a test DCT feature vector **t,**

$$c = \mathbf{t}^{+}\mathbf{w} = \mathbf{t}^{+}\mathbf{X}(\mathbf{X}^{+}\mathbf{X})^{-1}\mathbf{u} \quad (1.6)$$

However these projections are linear and thus any decision boundaries found are only linear, to enhance the discrimination power we need to find non-linear mappings of these features that will discriminate between classes thus upon examining Eq.(1.6) we can apply the Kernel trick shown below

$$K(x,t) = <\Phi(x), \Phi(t)> \quad (1.7)$$

wherever we have inner-products between pairs of images to yield a Closed Form Kernel Discrete Cosine Transform Feature Discriminant, thus we can re-write Eq. (1.6) as follows to map the images to a higher dimensional mapping.

$$c = \Phi(\mathbf{t})^{+}\Phi(\mathbf{X})[\Phi(\mathbf{X})^{+}\Phi(\mathbf{X})]^{-1}\mathbf{u} \quad (1.8)$$

Applying the kernel trick in Eq. (1.7) we obtain the following kernel discriminant projection

$$c = K(\mathbf{t},\mathbf{X})[K(\mathbf{X},\mathbf{X})]^{-1}\mathbf{u} \quad (1.9)$$

Where K is the kernel matrix computed using a particular kernel mapping function. Any Kernel function Φ can be used as long as the produced kernel matrix satisfies Mercer's theorem (i.e. K must be symmetric, semi-positive definite) to ensure that the non-linear mapping of the feature in the higher dimensional feature space is a valid inner-product space. Examples of valid Kernel functions are:

Polynomial kernel:

$$K(\mathbf{t},\mathbf{x}) = (\mathbf{t}^{+}\mathbf{x} + 1)^{p} \quad (1.10)$$

Radial Basis Function kernel:

$$K(\mathbf{t},\mathbf{x}) = \exp\left(-\frac{\|\mathbf{t}-\mathbf{x}\|^2}{2\sigma^2}\right) \quad (1.11)$$

It also interesting to explore how new kernels can be formed by combining existing kernel functions. This has to be done however by ensuring that Mercer's Theorem is still valid. In the above examples, we need to find the optimal parameters such as the order of the polynomial and the sigma of the RBF kernel, these are all tunable parameters that can be refined to optimize overall performance of any kernel method approach.

4. DISCRIMINANT LEARNING IN REDUCED DIMENSIONAL FEATURE SPACE VIA SUPPORT VECTOR MACHINES

In the previous stages we showed how we used Kernel trick to extract non-linear kernel Discrete Cosine Transform Synthetic Discriminant Features.

- We build a Kernel DCT Discriminant Function for each 222 people in the generic dataset in a 1-against all setup. Thus we have 222 K-DCT-DFs.
- For each of the gallery and probe images we compute the 2D-DCT features and project into each of these 222 K-DCT-DFs to obtain a feature vector of length 222 for all DCT transforms of the gallery and probe.

In order to perform discriminant learning in this 222-dimensional feature space we train a support vector machine on the Gallery (Target) set.

- For each of the 466 people in the gallery set we train a single SVM which is trained to maximize the margin between its images and the images of the remaining 221 people.
- We then repeat this for all 222 people to yield 222 SVMs, and these are then
- For each gallery image we retrieve the SVM used to train it and evaluate all the K-DCT-DF projection features of the 8,000 faces from Probe set, the resulting projections are used to populate the similarity matrix.

We tried different kernel functions and kernel parameters and report the best results obtained in the next section.

5. RESULTS

We performed Experiment 4 as described in the previous sections by computing the K-DCT-DFs and producint the 222 dimensional feature vectors and then training a support vector machine for further discrimination in this reduced dimensional feature space. We also compared the results to not using the DCT and used the original spatial images in th same fashion, and we show that using DCT representation produces better non-linear features for discrimination compared to spatial images. Figure 3 in the previous stage shows how our proposed approach using the Kernel trick to extract non-linear kernel Discrete Cosine Transform features is superior to the PCA baseline algorithm yielding a verification rate of 91.33% at 0.1% FAR compared to PCA baseline of 12% @ 0.1 FAR. We also computed the ROC

curve if we did not use the Discrete Cosine Transform feature but rather presented the raw spatial image and we see that maximum performance obtained is 85% @ 0.1% FAR, which shows that our proposed approach produced a significant boost over baseline (79% boost) and possible similar spatial approach (6.33% boost).Note that we also did perform the experiment of combining PCA+SVM but the results were still very low (47% @ 0.1 % FAR nearly half the recognition rate that we obtained).

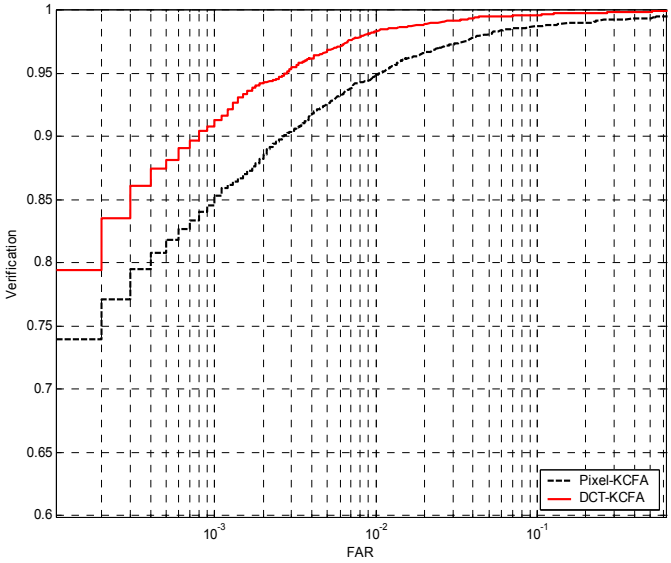

Figure 3: ROC curve showing the Verification rate vs FAR for FRGC experiment 4 for our proposed Kernel Discrete Cosine Transform Synthetic Discriminant Function compared to using spatial image domain image features.

5. CONCLUSIONS

In this paper we present a novel approach to perform extremely well on the hardest experiment in the face Recognition Grand Challenge Database collected by University of Notre Dame where the baseline benchmark algorithm given yields 12% verification @ 0.1 % FAR. We show that with our proposed Class Dependent Kernel Discrete Cosine Transform Discriminant Features we can obtain a significant boost in performance of about 79% over the baseline benchmark PCA algorithm. We also show that using the proposed Discrete Cosine Transform Representation is optimal and outperforms in trying to find non-linear features for discrimination in raw spatial images. Working in the DCT domain also allowed us to generate more generic training images by perturbing the DCT coefficients by adding a very small amount of noise in order to generate more training images (this was more effective than in the original image domain). In the near future work we will show how the performance changes by manually retaining only a very few DCT components and determining which parts of the DCT spectrum are most discriminative. We will also explore other image transformations to see their effect and compare the recognition performance and computational complexity.

ACKNOWLEDGEMENTS

We are thankful for the support of this research work sponsored by the United States Technical Support Working Group (TSWG) and in part by Carnegie Mellon CyLab.

REFERENCES

[1] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W.Worek, "Overview of the Face Recognition Grand Challenge," *IEEE Conf. Computer Vision and Pattern Recognition(CVPR)*, 2005

[2] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of *Cognitive Neuroscience*, Vol. 3, pp.72-86, 1991.

[3] "Fundamentals of Digital Image Processing" Prentice Hall, 1989

[4] P.Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. PAMI*, Vol.19. No.7, pp.711- 720, 1997.

[5] M. Savvides, B.V.K. Vijaya Kumar and P. Khosla, "Face verification using correlation filters," *Proc. Of Third IEEE Automatic Identification Advanced Technologie*s, Tarrytown, NY, pp.56-61, 2002.

[6] C. Xie, M. Savvides, and B.V.K. Vijaya Kumar, "Redundant Class-Dependence Feature Analysis Based on Correlation Filters Using FRGC2.0 Data," *IEEE Conf. Computer Vision and Pattern Recognition(CVPR)* , 2005

[7] A. Mahalanobis, B.V.K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt*. 26, pp. 3633-3630, 1987.

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (2nd Edition), *New York: Academic Pres*s, 1990

[9] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.

[10] B. Scholkopf, *Support Vector Learning*, Munich, Germany: Oldenbourg-Verlag, 1997.

[11] P. J. Phillips, "Support vector machines applied to face recognition," *Advances in Neural Information Processing Systems 11*, M. S. Kearns, S. A. Solla, and D. A. Cohn, eds., 1998