AN INTEGRATED APPROACH FOR GENERIC OBJECT DETECTION USING KERNEL PCA AND BOOSTING

Saad Ali   *Mubarak Shah*

Computer Vision Lab
School of Computer Science
University of Central Florida

ABSTRACT

In this paper we present a novel framework for generic object class detection by integrating Kernel PCA with AdaBoost. The classifier obtained in this way is invariant to changes in appearance, illumination conditions and surrounding clutter. A nonlinear shape subspace is learned for positive and negative object classes using kernel PCA. Features are derived by projecting example images onto the learned subspaces. Base learners are modeled using Bayes classifier. AdaBoost is then employed to discover the features that are most relevant for the object detection task at hand. Proposed method has been successfully tested on wide range of object classes (cars, airplanes, pedestrians, motorcycles etc) using standard data sets and has shown good performance. Using a small training set, the classifier learned in this way was able to generalize the intra-class variation while still maintaining high detection rate. In most object categories we achieved detection rates of above 95% with minimal false alarm rates. We demonstrate the comparative performance of our method against current state of the art approaches.

1. INTRODUCTION

Digital libraries have become an integral part of modern day applications in fields such as military, entertainment, academia, medical science, commerce etc. The world wide web is proving to be the driving force behind this explosion of multimedia content. Unfortunately utilization of this content is limited as current content-based image retrieval systems (CBIR) are not able to capture the semantics of scenes and objects present in images. This is, in part, because of the diverse visual appearances, poses, lighting conditions, and backgrounds in which an object can occur (Fig. 1). Therefore, industry mostly employs human indexers to assign keywords to their images so that a user can access them through simple text search. Hence, there is a pressing need for a methodology which can carry out automatic object detection and indexing across wide range of imagery.

Traditionally visual classification of objects is done in two steps. First, features are extracted from the image and object of interest is encoded using those features. Second, a classifier is learned using these features. Popular classifier employed for this task includes Support Vector Machines [17], Perceptron, Winnow etc. These are termed as linear classifiers or hyperplane classifiers, which work under the assumption that the data we are classifying is linearly separable. Unfortunately, in many cases, the representation of class and non-class images in feature space does not allow simple linear separation. For instance, when using image gray levels, color or texture as features, the separating surface between class and non-class example is highly non-linear and hence difficult to approximate[12, 8]. However still most of the current content based image retrieval systems use color, texture, orientation or blob features [3] and try to learn a linear classifier using them. Others try to compute similarity measure (L_1 or L_2 norm) between these high dimensional features to return the relevant images. But in high dimensions, data becomes very sparse and distance measures become increasingly meaningless. Therefore we see a degradation in the quality of results returned by CBIR.

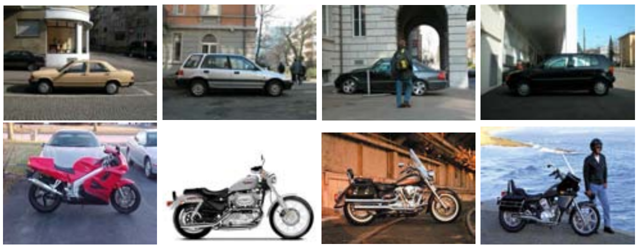

Fig. 1. Example of variation among object categories (Car and Motor Cycles) in terms of appearance, illumination condition, occlusion and background.

Principal Component Analysis (PCA) is an orthogonal basis transformation that can be effectively performed on a set of observations that vary linearly. However it fails to detect structure in a given data if the variations among the observations are non-linear which is the case when one is trying to extract features from object categories that vary in their appearance, pose and illumination conditions. Therefore any subsequent learning algorithm will have poor classification performance.

To overcome above mentioned shortcomings we propose an integrated framework of Kernel PCA [1] and AdaBoost. We demonstrate the feasibility of our approach on task of object detection on wide range of object categories. The essential idea is to employ Kernel PCA as a non-linear feature extractor by mapping input space to a higher dimensional feature space, through a non-linear

0-7803-9332-5/05/$20.00 ©2005 IEEE

map where the data is linearly separable. The justification of converting data to higher dimensional space is often alluded to Cover's theorem [7]. This theorem formalizes the intuition that number of separations increases with the dimensionality as we can have more views of the class and non-class data. Note that in practice we do not have to compute the expensive higher dimensional mapping as we can achieve the same effect by using the Kernel Trick [2]. This mapping will solve the problem of nonlinear distribution of low level image features. Once in the feature space, which is of high dimension, we uncover the patterns by selecting only the relevant (discriminative) dimensions using AdaBoost. This selectivity not only reduces the dimensionality but also speeds up online classification and retrieval. Classifier can be trained using small training set which is an added advantage. In short, our method overcomes many limitations of current CBIR in semantic content modelling.

2. RELATED WORK

Extensive research is being done in the area of video and image content modelling. Host of features such as color distributions, texture and shape have been explored. Few of the examples are IBM's QBIC system [4], BlobWorld [5] and VideoQ [6]. The retrieval on color usually returns images with similar colors but not necessarily similar semantic meaning. Retrieval on texture and shape features also return many irrelevant results.

PCA is a powerful technique for extracting global structure from high dimensional data set. It has been used to extract features for face recognition [10]. Kernel PCA [1] is proposed as a nonlinear extension of PCA, which computes the principal components in a hig dimensional feature space which is non-linearly related to the input space. Therefore it is able to extract non-linear principal components. Kernel PCA is used in Computer Vision community for modelling the variability in classes of 3D-shapes [13, 11]. In [14] they used Kernel PCA to learn the view subspaces for mutliview face detection. Recently [15] used it for recognition of facial expression using Gabor filters.

Our approach differs from above mentioned work as we propose an integrated framework of Kernel PCA with Boosting. This will enable us to exploit their strengths, first by modelling the nonlinear structure of object categories using Kernel PCA, and second by selecting highly discriminative features using Boosting. In addition, our method is able to handle multiple categories as oppose to above mentioned approaches that are restricted to just one category. We illustrate the robust performance of this approach on standard data sets.

3. KERNEL PCA FOR FEATURE EXTRACTION

3.1. Kernel PCA

Given a set of examples $x_i \in \Re^N$, i=1,...m, which are centered, $\sum_{i=1}^{m} x_i$ = 0, PCA finds the principal axis by diagonalizing the covariance matrix,

$$\mathbf{C} = \frac{1}{m}\sum_{j=1}^{m}\mathbf{x}_j\mathbf{x}_j^\top . \qquad (1)$$

To do this we solve the eigenvalue equation, $\lambda\nu = \mathbf{C}\nu$. We sort the eigenvalues in descending order, and use the first $M \leq N$ principal components ν_k as the basis vector of lower dimensional subspace, forming the transformation matrix **T**. The projection of example $\mathbf{x} \in \Re^N$ onto the M dimensional subspace can be calculated as $\beta = (\beta_1, ..., \beta_M)$ = $\mathbf{x}^\top\mathbf{T}$. The $\beta's$ represent the derived features for example **x**.

Now Kernel PCA is performed by first mapping the data from input space to a higher dimensional feature space i.e. $\phi : \Re^N \rightarrow$ **F**, and then performing a linear PCA in **F**. The covariance matrix in this new space **F** is,

$$\overline{\mathbf{C}} = \frac{1}{m}\sum_{j=1}^{m}\phi(\mathbf{x}_j)\phi(\mathbf{x}_j)^\top . \qquad (2)$$

The eigenvalue problem now becomes $\lambda\mathbf{V} = \overline{\mathbf{C}}\mathbf{V}$. As mentioned previously we do not have to explicitly compute the nonlinear map ϕ. We can achieve the same goal by using kernel function $k(x_i,x_j) = (\phi(x_i).\phi(x_j))$ which implicitly computes the dot product of vectors x_i and x_j in higher dimensional space.[2]. They can also be thought of as functions measuring similarity between instances. The kernel value will be greater if two samples are similar otherwise it falls off to zero if samples are distant. The most often used kernel types are polynomial and Gaussian kernels (Table 1). Now defining a Gram Matrix $K \in \Re^N$ where each entry $K_{i,j}$ is calculated using the kernel function $k(x_i, x_j)$, the eigenvalue equation can be written as (see [2]),

Gaussian Kernel	$k(x_i, x_j) = exp^{(\frac{-\|x_i - x_j\|)^2}{c}}$
Polynomial Kernel	$k(x_i, x_j) = (x_i.x_j + a)^d$, d=1,2..
Sigmoid Kernel	$tanh(k(x_i.x_j) + a)$

Table 1. Kernels

$$m\mathbf{A}\Lambda = \mathbf{K}\mathbf{A}, \qquad (3)$$

with **A**= $(\alpha_1, ..., \alpha_M)$ and Λ = $diag(\lambda_1, ..., \lambda_M)$. **A** is a *m* x *m* orthogonal eigenvector matrix and Λ is a diagonal eigenvalue matrix with diagonal elements in decreasing order. Since the eigenvalue equation is solved for α_j instead of eigenvector $\mathbf{V}_j$ of kernel PCA, we will have to normalize **A** to ensure that eigenvalues **V** have unit norm in the feature space, therefore $\alpha_j = \alpha_j/\sqrt{\lambda_j}$. The eigenvector matrix **V** of kernel PCA is computed as, **V** = **DA** where D =[$\phi(x_1)$ $\phi(x_2)$ ... $\phi(x_m)$] is the data matrix in feature space. Now let **X** be a test example whose map in the higher dimensional feature space is $\phi(X)$. The kernel PCA features of **X** are derived as follows:

$$\mathbf{F} = \mathbf{V}^\top\phi(\mathbf{X}) = \mathbf{A}^\top\mathbf{B}, \qquad (4)$$

where $\mathbf{B} = [\phi(x_1)\phi(\mathbf{X})\ \phi(x_2)\phi(\mathbf{X})\ ...\ \phi(x_m)\phi(\mathbf{X})]$.

3.2. Feature Extraction

Let **P** = $(p_1, p_2, ...p_m)$ and **N** = $(n_1, n_2, ...n_m)$ be the positive and negative images of the training set provided for learning. Gradient magnitudes are extracted from the images by convolving them with sobel gradient operator. Gradient provides better shape cues than gray level intensity or color texture patterns, which are more biased towards the visual appearance of the object, background and surrounding clutter. Gradient images are resized to 128 by 128 pixels, converted into column vector form and made zero mean and unit variance. We computed Gram matrix K_p and K_n using kernel function (polynomial or Gaussian). Eigenvector matrix A_p and A_n is calculated using eq. 3. Features for our base learners were obtained by projecting each positive and negative training example onto the positive and negative higher dimensional subspaces by plugging A_p and A_n in eq. 4 respectively. The feature vector for any particular example will be of the form, $\mathbf{f} = [d_1, d_2, ..., d_{w_1}, d_{d_1+1}, ..., d_{w_1+w_2}]$ where w_1 and w_2 are the number of principal components that we retained for each class. Therefore the total number of base learners are going to be $w_1 + w_2$.

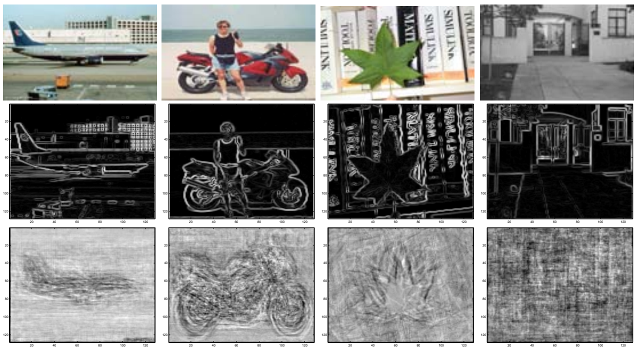

Fig. 2. In top to bottom form, each image is followed by its gradient and reconstructed image structure from top 150 eigenvectors. Categories are airplane, motorcycle, leave and background.

4. LEARNING CLASSIFIER WITH BOOSTING

The notion of focusing on the most relevant information in potentially high dimensional data is very important. Efficiency of the final system depends on whether we are able to discover the irrelevant features that hide the useful information in sea of noise or not. Specifically, in CBIR this step will determine the amount of time the system is going to spend in searching through overwhelming amount of image content for each user query. Features generated by kernel PCA lie in a high dimensional nonlinear subspace and we want to find out if all of those dimensions are useful for the classification task at hand or we can achieve the same goal by using subset of those dimensions. Therefore we integrate AdaBoost with Kernel PCA as a feature selection device. AdaBoost is an ensemble classifier learning algorithm that works by creating a sequence of base learners in each iteration, where each base learner is constructed based on the performance of the previous base learner on the training set. In each iteration the weight distribution over the training set is updated in a way that forces the base learners to focus on the example that are hard to classify. This results in a classifier with low training error and good generalization performance.

Note that one may be tempted to use nearest neighbor (NN) classifier or any similar classifier to categorize the features derived from kernel PCA without carrying out any feature selection. This will have adverse effect on the classification performance as NN uses all features for its distance computation which will include some features generated from noisy data. In addition the number of training examples required to reach given accuracy grows exponentially with number of irrelevant features in case of NN. On the other hand our framework guarantees to provide classifier based on subset of most discriminative features using small set of training examples.

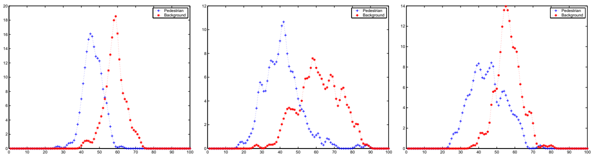

Fig. 3. Histograms of three different feature dimensions used in training of Pedestrian-Background Classification. Blue and Red represent pedestrian and background respectively.

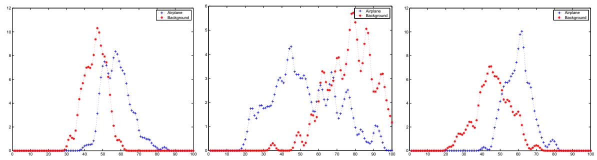

Fig. 4. Histograms of three different feature dimensions used in training of Airplane-Background Classification. Blue and Red represent airplane and background respectively.

We use the Bayes classifier as the base learner for AdaBoost. Let c_p and c_n be the positive and negative class respectively. The classification decision of *ith* classifier is taken as c_p if $P(c_p|d_i) > P(c_n|d_i)$. The posterior is given by Bayes rule, i.e., $P(c_p|d_i) = \frac{p(d_i|c_p)P(c_p)}{p(d_i)}$. The class conditional probability densities $p(d_i|c_p)$ and $p(d_i|c_n)$ are approximated through smoothed 1D histogram of the of the i^{th} dimension of the feature vector **f**. In order to have good discrimination, ranges and bin widths of these histograms need to be selected carefully. Examples of histogram of three different features for pedestrian/background and airplane/background feature vectors are given in Fig.3 and Fig. 4 respectively.

We used the boosting algorithm proposed by [16] to perform feature selection. Now, to test a new image, we preprocess it according to the specifications described in 3.2. The feature vector is obtained by projecting it onto the subspaces using A_p and A_n (eq. 4). Note that in eq. 4, $(x_1, x_1 ... x_m)$ is same training examples that were used to construct the nonlinear subspaces. We need to save them as they will be used for testing any new example.

5. RESULTS AND CONCLUSIONS

This section asses the performance of our object detection approach using standard data sets available in public domain. A description is given in Table 2. We performed the classification in the setting of one category versus the background in order to compare results with [9, 18].

5.1. Experiments

Experiments were carried out by splitting the data sets into two parts. One part is used for constructing the kernel PCA subspaces, base learners and strong classifier while the other part is used for testing. Each experiment was conducted using polynomial kernel (with d equal to 2 and 3) and Gaussian kernel (results in Fig. 6).

5.2. Conclusion

We have proposed a novel approach for object detection by integrating kernel PCA and boosting. Our approach has elegantly answered the the questions of which features to use for describing a semantic concept and how to combine those features. It has

Data Set	Training Images	Testing Images
UIUC Cars	150	200
Caltech Car Rear	170	480
Caltech Airplane	200	874
Caltech Motorcycles	200	626
Caltech Faces	100	350
Caltech Leaves	50	137
MIT CBCL Pedestrians	200	724
ETH Zurich Cars	50	50

Table 2. Data sets used in experiments.

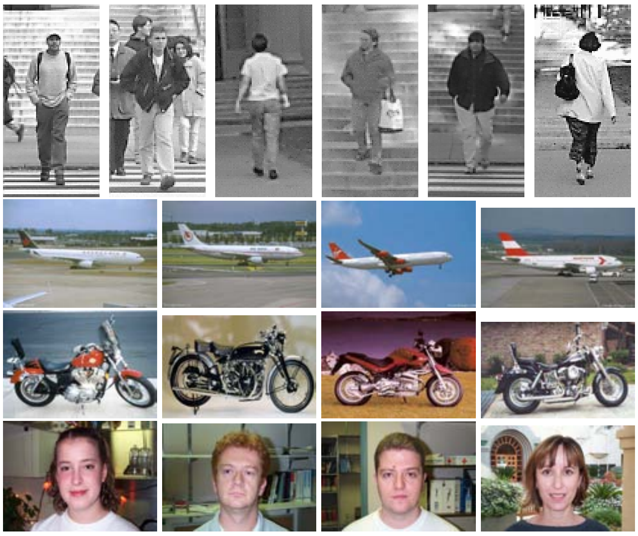

Fig. 5. Some example images from Pedestrian, Airplane, Motorcycle and Face data sets.

Data Set	Detection Rate	False Alarm Rate	Principal Components
UIUC Cars	99.5%	0.8%	140
Caltech Car Rear	99.4%	0.5%	150
Caltech Airplane	98.5%	0.1%	150
Caltech Motorcycles	99.84%	0.1%	150
Caltech Faces	100%	0.2%	90
Caltech Leaves	100%	0.1	45
MIT CBCL Pedestrians	100%	0%	150
ETH Zurich Cars	86%	9%	35

Table 3. Results in terms of detection rate and false alarm rate.

Data Set	**Our Method**	**Fergus**	**Boosting Conext**
Caltech Car Rear	99.4%	90.3%	97%
Caltech Airplane	98.5%	90.2%	92.7%
Caltech Motorcycles	99.84%	92.5%	73.9%
Caltech Faces	100%	96.4%	-
Caltech Leaves	100%		97.8%

Table 4. Comparison of our results with [9, 18].

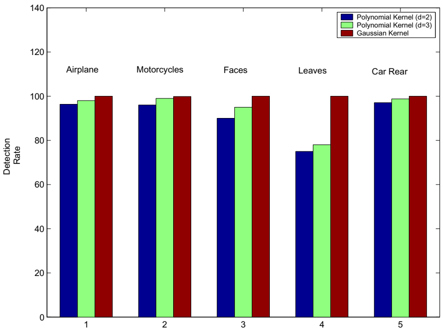

Fig. 6. Comparison of results obtained using polynomial kernel of degree 2 (Blue), degree 3 (Green), and Gaussian kernel (Red) on Airplane, Motorcycle, Face, Leaves and Car Rear data set.

several advantages. It is scalable and can be extended to any object category. It requires small set of examples for training. We showed its accuracy on challenging data sets and wide range of object classes. It performed better than many current state of the art methods.

6. REFERENCES

[1] B. Schölkopf et al., “ Nonlinear component analysis as a kernel eigenvalue problem, ” *Neural Computation*, 1998.

[2] B. Schölkopf and A. Smola, “ Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond, ”*MIT Press*, 2002

[3] R. C. Veltkamp et al., “ Image Retrieval Systems: A Survey,”*Technical Report Universiteit Utrecht*, 2002.

[4] W. Niblack et al., “ Querying Images by Content using Color, Texture and Shape,”*SPIE SRIVD*, 1993.

[5] C. Carson et al., “ Blobworld: A System for Region Based Image Indexing and Retrieval,”*VISUAL*, 1999.

[6] Shih-Fu Chang et al., “ VideoQ: An Automated Content Based Video Search System Using Visual Cues,”*ACM MULTIMEDIA* , 1997.

[7] T. M. Cover, “ Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition,”*IEEE Transaction on Electronic Computers*, 1965.

[8] M. Vidal-Naquet et al., “Object Recognition with Informative Features and Linear Classification, ”*ICCV*, 2003.

[9] R. Fergus et al., “ Object Class Recognition by Unsupervised Scale-Invariant Learning,”*CVPR*, 2003.

[10] M. Turk and A. Pentland, “ Eigenfaces for Recognition,”*Journal of Cognitive Neuroscience* , 3(1):71-86, 1991.

[11] S. Romdhani et al., “ A Multi-view Nonlinear Active Shape Model using Kernel PCA,”*BMVC* , 1999.

[12] D. Cremers et al., “ Nonlinear Shape Statistics via Kernel Spaces, ” *DAGM-Symposium*, 2001.

[13] C. J. Twining et al., “Kernel Principal Component Analysis and the Construction of Non-Linear Active Shape Models, ”*BMVC*, 2001.

[14] S. Z. Li et al., “Kernel Machine Based Learning for Multi-View Face Detection and Pose Estimation, ”*ICCV*, 2001.

[15] C. Liu, “Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition, ”*PAMI*, 26(5), 2004.

[16] SY. Freund and Robert E. Schapire, “ Experiments with a New Boosting Algorithm,”*ICML* , 1996.

[17] IBM Research TRECVID-2004 Video Retrieval System.

[18] J. Amores et al., “ Boosting contextual information in content-based image retrieval,”*ACM SIGMM*, 2004.