EFFICIENT OBJECT DETECTION USING PCA MODELING AND REDUCED SET SVDD

Rudra N. Hota and Venkataramana Kini B.
Research and Technology Unit, Honeywell Technology Solutions Lab, Bangalore, India

Keywords: Object Detection, PCA modeling and reduced set SVDD.

Abstract: Object detection problem is traditionally tackled as two class problem. Wherein the non object classes are not precisely defined. In this paper we propose cascade of principal component modeling with associated test statistics and reduced set support vector data description for efficient object detection, both of which hinge mainly on modeling of object class training data. The PCA modeling enables quick rejection of comparatively obvious non object in initial stage of the cascade to gain computation advantage. The reduced set SVDD is applied in latter stages of cascade to classify relatively difficult images. This combination of PCA modeling and reduced set support vector data description leads to a good object detection with simple pixel features.

1 INTRODUCTION

The object detection is a process of isolating the object of interest from it's surroundings, e.g. detection of people in an image snapshot. We human beings can do this effortlessly but when it comes to doing the same automatically by using machines we encounter many problems, in terms of scales at which they are present, orientation, illumination etc.

Traditionally in machine learning paradigm of object detection problem is tackled using two class problem, i.e. by having positive and negative class (object and non-object classes respectively) training data sets. The training procedure strives to find a optimal boundary between these two classes in the space of features with the expectation of good generalization. This kind of approach is followed successfully ((Viola and Jones, 2001), (Romdhani et al., 2001), (Rowley et al., 1997)) by the researchers by having variety of object class data in terms of several thousands and hundred thousands from non-object class. It is natural to include variety of data (by applying different affine transformations such as orientation, scale and with different illumination) for positive class, but for negative class we can not define precisely what is variety of data to collect. Hence, the negative class remains ill defined. We could collect few hundreds of thousands (or more) of negative class data, but it is insufficient as negative class is of virtually infinite in size.

Another framework for object detection problem often used by the machine learning community is of developing a model /data description (which describes structure in the data) only for the positive class. The negative class data is used mostly for fine tuning of the boundary around positive class. The classification (detection in true sense) involves looking for object features in given image. If such features are found, then object is detected otherwise the image is labeled as non-object.

The simple model for this kind of approach is Principal Component Analysis(PCA). This appears to be natural way of describing the objects. The PCA features in the form of eigenfaces (Moghaddam and Pentland, 1997) has been applied successfully for face recognition and related tasks. In this paper we utilize PCA features for describing the structure in target object class data. The test data is projected onto this model represented by major principal directions. These principal directions being few in number, leading to quick rejection of obvious negative samples. We utilize the test statistics which are traditionally used in fault detection (Venkatasubramanian et al., 2003) (Yue and Qin, 2001) community. These statistics provide thresholds which are effectively used to discard non-object images. Here, it should be noted that the PCA model is used as a coarse model for object class data, the different directions might have different discriminatory power (Sun et al., 2002), but overall model for object class based on dominant prin-

N. Hota R. and Kini B. V. (2008).
EFFICIENT OBJECT DETECTION USING PCA MODELING AND REDUCED SET SVDD.
In *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pages 222-227
DOI: 10.5220/0001079402220227

cipal directions gives an envelop around object class data and provides strategy for quick rejection of non-object class data.

The kernel methods have been proposed for the task developing data based models as well. The traditional data based modeling technique PCA is extended to handle higher order correlations in the data by mapping into higher dimensional feature space - Kernel Principal Component Analysis(KPCA) (Scholkopf et al., 1998). It is simple enough compared to PCA in terms of just finding eigen value decomposition. It finds the uncorrelated features in higher dimensional space, explaining structure of positive class data. But in object detection problems as we use many thousands of object class data for training of KPCA, the run-time computational complexity blows up.

In this work, we applied cascaded structure for object detection, in which the removal negative samples are taken care in different stages according to their degree of closeness to positive class distribution. Therefore, it it required to have strong classifiers in the later stages for separating difficult negative samples from positive class. Hence, for discriminating non-objects which are like objects, we need to resort to strong classifiers (Heisele et al., 2003). Traditionally artificial neural networks are developed as strong classifier. However, neural networks demand lengthy training, convergence of the training process is sometimes uncertain and choice of network architectures remains somewhat of an art. In early 1990s, the kernel methods (Vapnik, 1999) such as Support Vector Classifiers(SVC), regressors are developed for classification and function approximation tasks. The advantage of these methods over neural network methods is that they implicitly solve the nonlinear problem. Also they exhibit good generalization capability because of their regularization properties.

Another data based modeling technique in kernel feature space is Support Vector Data Description (SVDD)(Tax and R.P.W, 2004). It tries to find a enclosing sphere of minimal volume for positive class data in high dimensional feature space unlike SVC, which tries to find the hyperplane between positive and negative class training data. This kind of model is particularly suited for object detection problem (Seo and Ko, 2004) (Tax and R.P.W, 2004). But, the daunting disadvantage with SVDD when applied to object detection is the number of kernel computations involved. The number of kernel computations involved is order of the number of support vectors generated during training procedure. In typical problems of object detection (face detection and people detection) that we are targeting, these support vectors are as high as few thousands. Because of these high number of Support Vectors (SVs) the computational cost may be sometimes between half a minute to few minutes.

In this paper to tackle this computational cost, we propose to "leverage the technique of reducing the number of support vectors (Romdhani et al., 2001) into SVDD". The number of support vectors can be reduced to few hundreds from thousands without compromising much on accuracy.

The quick rejection of non-object data using linear PCA and associated test statistics, followed by Reduced Set SVDD leads to a good balance between speed and accuracy. Hence, we propose a efficient (both in terms of speed and accuracy) method for the problem of object detection by novel cascade of linear PCA modeling and series of Reduced Set SVDD (RSSVDD) with increasing number of reduced set SVs.

The outline of the paper is as follows. The method for quick rejection based on PCA modeling is explained in next section. The RSSVDD is explained in section 3. The overall approach of cascaded PCA and RSVDDs is explained in section 4. Section 5 gives experiments and results of object detection (specifically on face data). In section 6 we draw some conclusions of this work and plans for future work.

2 PCA MODELING OF OBJECT CLASS AND THRESHOLDING STATISTICS

PCA is a versatile data analysis tool. It can be considered as data modeling tool, the major principal components capturing most of the variance in the covariance matrix of data. The rest of the components are assumed to represent noise in the data. The steps involved in PCA modeling is summarized in below algorithm. PCA based feature extraction has received considerable attention in computer vision area. In previous works (Moghaddam and Pentland, 1997) the image is represented by features in a low dimensional space spanned by the principal components. These features are further utilized in classifier. PCA is predominantly used for extracting features and dimensionality reduction, the modeling perspective is missing.

PCA is traditionally applied by Chemometrics with modeling perspective for the purpose of fault detection. Fault detection using PCA models is normally accomplished by applying two statistics. The squared prediction error SPE , which indicates the amount by which a sample deviates from the model, is defined

Table 1: Algorithm for PCA model training on Object class data.

1. Given the data set $X = \mathbf{x}_1 \dots \mathbf{x}_n$ from object class find the mean vector $\mu = \frac{1}{n}\sum_1^n \mathbf{x}_i$,
2.Find the covariance matrix $C = \frac{1}{n-1}\sum_1^n (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)'$
3.Eigen value decompose the covariance matrix $C = P\Lambda P$
4. Choose the principal components $\hat{P}$ corresponding dominant eigen values and remaining eigen vectors $\tilde{P}$ represent minor components.

by

$$SPE = \mathbf{x}'\tilde{P}\tilde{P}'\mathbf{x} \tag{1}$$

Hotellings T^2 statistic, which measures deviation of a sample inside the model, takes the form

$$T^2 = \mathbf{x}'\hat{P}\Lambda^{-1}\hat{P}'\mathbf{x} \tag{2}$$

Both of these indices follow chi square distribution under the normality assumption of object class data. The δ^2 and τ^2 limits (for SPE and T^2 respectively)are found for given confidence level. It is to be noted that SPE departs from limits whenever the covariance structure breaks down, where as T^2 violation happens whenever the different contributors (PCA scores) go out of range. For object detection problem, whenever any one of these indices crosses limits, we can infer that image is from non-object class. Therefore it makes sense to combine them. (Yue and Qin, 2001) has proposed combination index defined as:

$$\rho = c\frac{SPE}{\delta^2} + (1-c)\frac{T^2}{\tau^2}, \quad c \in (0,1) \tag{3}$$

Since ρ is linear combination of T^2 and SPE the limits can be found using chi square distribution.

3 REDUCED SET SVDD

First we introduce the notation by explaining SVDD (Tax and R.P.W, 2004) in input space. Given the positive class data-set, SVDD tries to find enclosing sphere around data which has minimum volume. By minimizing the sphere volume, SVDD minimizes the detection error (i.e. the chance of accepting outlier objects).

Given positive class data $X = \{\mathbf{x}_1 \dots \mathbf{x}_n\}$ the SVDD attempts to find a hyper-sphere with center c and radius R which encloses most of the data, i.e. the volume of the hyper-sphere is minimized. That is to minimizing the objective function:

$$F(R,c) = R^2 \quad s.t. \|\mathbf{x}_i - c\|^2 \leq R^2 \quad \forall i \tag{4}$$

The positive class data set might contain some outliers. To allow such possible outliers, introducing slack variables ξ_i into 4 leads to primal solution:

$$F(R,c) = R^2 + C\sum_i \xi_i \;\; s.t. \|\mathbf{x}_i - c\|^2 \leq R^2 + \xi_i \;\; \xi_i \geq 0 \quad \forall i \tag{5}$$

The parameter C is a regularization parameter. Introducing Lagrange multipliers and setting the partial derivatives w.r.t. R, c and ξ_i leads to dual solution:

$$L = \sum_i \alpha_i(\mathbf{x}_i \cdot \mathbf{x}_i) - \sum_{i,j} \alpha_i\alpha_j(\mathbf{x}_i \cdot \mathbf{x}_j) \tag{6}$$

When $0 < \alpha_i < C$ implies $\|\mathbf{x}_i - c\|^2 = R^2$, i.e. these data points lie on boundary of SVDD solution. For any test data point $\mathbf{x}$, label is decided to be belonging to object class based on distance to center smaller or equal than the radius:

$$\|\mathbf{x} - c\|^2 = (\mathbf{x}\cdot\mathbf{x}) - 2\sum_i \alpha_i(\mathbf{x}\cdot\mathbf{x}_i) - \sum_{i,j}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j) \leq R^2 \tag{7}$$

R^2 is found by substituting for $\mathbf{x}$ with any of the SVs.

For flexible data descriptions the dot products are replaced by corresponding kernel dot products, i.e. $x \mapsto \Phi(x)$. Also the negative data objects can be utilized in fine tuning the data description boundary of SVDD (Tax and R.P.W, 2004).

As discussed in introduction, (section 1) the solution generated by SVDD involves large percentage of data points as SVs, resulting in high run-time complexity in object detection problems.

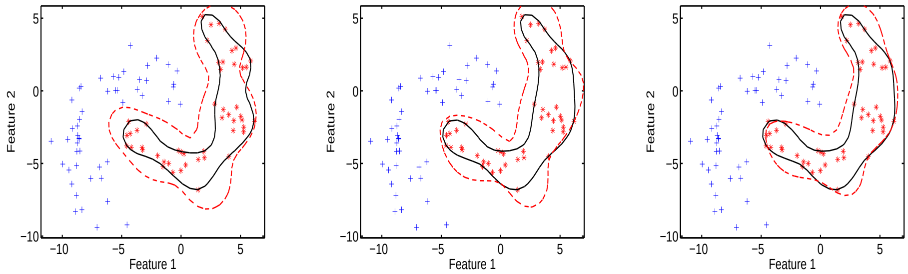
In this work, the methodology adopted to reduce the number of SVs is derived from (Romdhani et al., 2001). Let N_s number of SVs obtained for SVDD training. The sum of SVs weighted by non-zero α is given by

$$\Psi = \sum_{i=1}^{N_s} \alpha_i \Phi(\mathbf{x_i}) \tag{8}$$

By approximating the N_s number of SVs by new N_{rs} number of SVs we can reduce the computational complexity. The approximation of Ψ is:

$$\hat{\Psi} = \sum_{i=1}^{N_{rs}} \beta_i \Phi(\hat{\mathbf{x}}_\mathbf{i}) \tag{9}$$

Where, $\beta \in \Re$ and $\hat{\mathbf{x}}_\mathbf{i}$ are approximate SVs. We can usually achieve $N_{rs} \ll N_s$ with out loss of much accuracy. To achieve this, (Romdhani et al., 2001) has suggested to minimize the norm $\|\Psi - \hat{\Psi}\|^2$, which can



(a) (b) (c)

Figure 1: Dotted boundary represents (a) RSVDD with 30% SVs (b) RSVDD with 40% SVs (c) RSVDD with 60% SVs.

Table 2: Algorithm for RSSVDD training on Object class.

1. Given the data set $X = \mathbf{x}_1 \ldots \mathbf{x}_n$ find find SVs using SVDD algorithm
2.Find the reduced set SVs by minimizing $\|\Psi - \hat{\Psi}\|^2$
3. Choose the cascade of RSSVDDs starting from a minimum number of reduced set SVs decided based on classification performance on training data set

be rewritten in terms of only kernel dot products. They also provide an iterative algorithm for finding reduced SVs, starting from one SV. The Figure 1 shows a simple banana data example, wherein star shaped positive class data is modeled using different percentages of original SVs. The solid line envelop around positive class represents original SVDD boundary. The algorithm for RSSVDD is summarized in Table 2.

4 OUR APPROACH - EFFICIENT CASCADE OF PCA AND SVDD ALGORITHM FOR OBJECT DETECTION

In object detection problems, such as face detection, the object can occur at any position of the image and in several scales. Under such cases the automated object detection algorithm has to search the image either with a fixed size of sub-window in a pyramidal structure or by varying the scale of the search window starting from a desired scale. For a standard size image the percentage of sub-windows containing target object is usually less than one percent of windows to be scanned for the whole image. This problem demands a coarse-to-fine cascade classifier strategy to search the target object, i.e. in the initial stage of the detector it should be able to remove as many negative patterns as possible, retaining almost all positive patterns. As the level increases the task of separating the object like negative class objects from the target object becomes tougher. Hence cascaded classifier with better accuracy at later stage (where we can afford higher computational complexity) of the detector is suitable. Our approach uses cascade of linear PCA model followed by series of one-class RSSVDD classifiers with increasing number of support vectors. There are many possible good features for face detection, presented in literature. For simplicity we consider only pixel values of the image sub-window as feature vector. Our method can also be applied with other sophisticated features

Table 3: Algorithm for Object detection using PCA and RSVDD cascaded classifiers.

1. **Training**: Given target object class image data $X = \{\mathbf{x}_1 \ldots \mathbf{x}_n\}$ perform
 (1.a) PCA model training (Table 1) and find the threshold by using equation (3)
 (1.b) SVDD training and find Reduced set SVs (Table 2), decide on minimum number of RSSV
2. **Testing**: Given the test image, form the sub-windows and for each vector x representing sub-window
 (2.a) Project x onto principal directions found in (1.a) and if it crosses the threshold,
  discard the sub-window as non-object and continue with new sub-window,
 else
  Start with RSSVDD with least number of SVs in cascade, if accepted by current RSVDD continue with next RSVDD in cascade until final RSVDD is reached, label accordingly.

(Gabor, Haar wavelet, etc.) to improve the performance. The overall approach (training and testing) is explained in below algorithm.

5 EXPERIMENTS AND DISCUSSION

Usually, in object detection problem the number negative sample is much more as compared to that of positive samples. This is because the negative class is always ill defined as we have discussed in earlier section. To discard comparatively easier negative samples we applied PCA modeling and more difficult false alarms are handled by reduced set SVDD in cas-

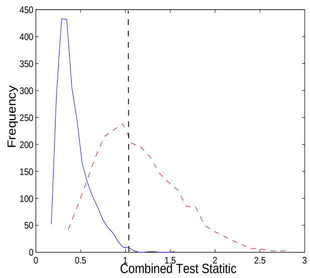


Figure 2: Quick rejection of non faces using PCA with test statistics.

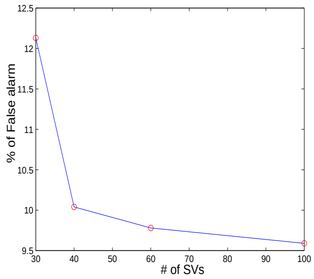


Figure 3: Percentage of reduced SVs vs false alarm rate.

caded structure.

Data set: In our experiments of face detection we used CBCL face data set [http://www.ai.mit.edu/projects/cbcl/ software-dataset/index.html], which contains 2429 faces and 4585 non face samples for training and 472 faces and 23573 non face sample for testing. For our experiment we have selected randomly 2000 each from face and non-face samples from training set for training purpose and rest of the samples are used for testing.

After applying initial filtering by PCA model with associated combined test statistics the negative patterns which are difficult to classify are fed to SVDD. The PCA model is able to discard more than 50% of non-face class data, retaining almost all face class data. The number of principal components retained is equal to the components which captures 90% of the variance (which is equal to 8 principal directions). The threshold based on chi square distribution is found to be equal to 1.08 (with $c = 0.5$ as shown in Figure 2 by dashed vertical line. The two frequency curves (solid and dashed respectively) represents frequency with which face class data and non-face class data appears corresponding to combined statistic. Here, randomly 2000 number face and non face class data is utilized to show the quick rejection capability of non-faces at the same retaining almost all face class data. Out of these 2000 face class data

Table 4: Performance of RSSVDD on Face detection.

% of SV retained	Detection rate (%)	False Alarm (%)
100	94	9.5
60	94	10
40	94	9.8
30	94	12

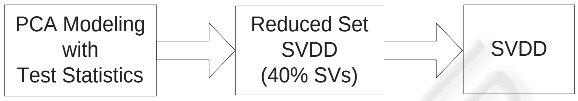


Figure 4: Cascade structure for face detection.

700 data points which have higher value of ρ (eqn 3)is retained for SVDD training and non-face data which have ρ less than 1.5 retained for fine tuning of SVDD boundary in next stage of cascade. The detection rate on test data the by applying PCA model is found to be 99.3% with false alarm rate of 58%.

With the data retained by PCA we analyzed the performance of SVDD by increasing the number of reduced support vectors (as explained in section 3). By using 100% of the support vectors for the trained SVDD we achieved 94% detection accuracy during testing with 9.5% of false alarm. By reducing the support vector to 30% (originally SVDD training produced 139 SVs from 700 face class training data) the false alarm rate increases to near about 12% with retaining same accuracy of detection rate. With 40% and 60% of SVs, the false alarm rate further reduces to around 10% and 9.8% respectively. The plot of percentage of support vectors vs false alarm rate is depicted in Figure 3. Further increase in reduced set of support vectors lead to better performance in terms of lesser false alarm at the same level of detection rate. This can be because of some generalization capability provided by approximate reduced SVs. When we reduce the SVs by certain percentage, the runtime computational complexity also reduced by approximately same percentage. Hence, we could reduce the runtime computational complexity of SVDD with out much compromising on the accuracy (false alarm increased by just 0.3% when SVs reduced to 40% from 100% with same detection rate). Therefore, we propose a cascade structure shown in Figure 4 for face detection problem. The combination of PCA modeling and RSSVDD drastically reduced overall computational complexity by about 80% as compared to a single monolithic SVDD classifier.

It is to be noted that in (Seo and Ko, 2004) by using the color information in SVDD false alarm rate was as low as 1% at same detection rate as that of our experiments. Hence, by making use of color features or other sophisticated features and preprocessing techniques the false alarm rate can be reduced substantially. However, in this work the goal was to show the efficacy of the PCA and RSSVVD cascade approach for the problem of face detection.

6 CONCLUSIONS AND FUTURE WORK

In this paper we proposed cascade of PCA modeling with associated test statistic and reduced set support vector data description for efficient object detection. The PCA modeling enabled quick rejection up to $40-50\%$ of comparatively obvious non-faces to gain computation advantage. The reduced set SVDD applied in later stage of cascade to classify relatively difficult images. This novel combination of PCA modeling and RSSVDD lead to good face detection at reduced computational cost by using only simple pixel features.

Motivated by these results, in future we plan to apply this approach to other object detection tasks such as vehicle, people detection. Further in future we plan to improve the performance of object detection using more sophisticated feature set.

REFERENCES

Heisele, B., Serre, T., Prentice, S., and Poggio, T. (2003). Hierarchial classification and feature reduction for fast face detection with support vector machines. In *Pattern Recognition Society*.

Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Romdhani, S., Torr, P., Scholkopf, B., and Blake, A. (2001). Computationally efficient face detection. In *8th Int. Conf on Computer vision*.

Rowley, H. A., Baluja, S., and Kanade, T. (1997). Rotation invariant neural network based face detection. In *Computer Science technical report CMU-CS-97-201*.

Scholkopf, B., Smola, A., and Muller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. In *Neural Computation, 10(5):1299-1319*.

Seo, J. and Ko, H. (2004). Face detection using support vector domain description in color images. In *ICASSP, 54, 4566,*.

Sun, Z., Bebis, G., and Miller, R. (2002). Object detection using feature subset selection. In *Pattern Recognition, 37, 21652176*.

Tax, D. and R.P.W, D. (2004). Support vector data description. In *Machine Learning, 54, 4566,*.

Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 2nd edition.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S. N. (2003). A review of process fault detection and diagnosis. In *Part I:Quantitative model-based methods. Comput. Chem*.

Viola, P. and Jones, M. (2001). Rapid object detection using boosted cascaded of simple features. In *Conf. on Computer Vision and Pattern Recognition*.

Yue, H. and Qin, S. J. (2001). Reconstruction based fault identification using combined index. In *I & EC Reasearch, 40, 4403-4414*.