

Cost Parameter Analysis and Comparison of Linear Kernel and Hellinger Kernel Mapping of SVM on Image Retrieval and Effects of Addition of Positive Images

Swathi Rao G

Anuj Sharma

ABSTRACT

In this paper we have brought out the analysis and comparison of cost parameter validation in Support vector machine using two different kernel mappings i.e. the linear and the Hellinger kernel. This paper also shows and discusses the results of the addition of positive images to the respective class of images with different cost parameters. The analysis is carried out using Matlab R2009a and C environment. The results obtained show that the increase in cost parameter for linear kernel gives much better results whereas for Hellinger kernel the performance decreases as cost parameter is increased. In the other hand, two classes of images are taken and they are tested by increasing the number of positive images gradually and the results show that the addition of positive class of images to a database can increase the performance of the system employed.

KEYWORDS

Cost parameter, Hellinger Kernel, Image Classification, Image Retrieval, Kernel Functions, Linear kernel, SIFTS, Support Vector Machine.

1 INTRODUCTION

The growth of the World Wide Web have led to the huge online digital images and videos, so there is a strong demand for developing an efficient technique for image retrieval to exploit maximum benefit from this huge amount of digital information. With the advancement in technology, a large amount of information in the form of images is being generated daily in various fields like architecture engineering designs biometrics satellite imagery. Also with the increasing capacity of the storage devices, the database of the image information is expanding allowing a huge amount of database images to be stored quite easily. Even though storing these images is now easier, accessing or retrieving these images as per the requirement is a tedious job. Image retrieval has been an active topic for research for the past three decades. The goal of an image retrieval system is to retrieve a set of images from a collection of images such that this set meets the user's requirements.

The user's requirements can be specified in terms of similarity to some other image or a sketch, or in terms of keywords. An image retrieval system provides the user with a way to access, browse and retrieve efficiently and possibly in real time, from these databases. Well-developed and popular international standards, on image coding have also long been available and widely used in many applications. The challenge to image indexing is studied in the context of image database, which has also been researched by researchers from a wide range of disciplines including those from computer vision, image processing, and traditional database areas for over a decade. This paper analyses the most important parameter i.e. the cost parameter of the support vector machine for various classes of images. The cost parameter is significant in a SVM classifier and has to be studied very carefully to train and test the dataset and obtain satisfactory results.

The other sections of the paper can be put up as follows: SECTION 2 describes the Support Vector Machine and the Significance of the cost parameter for various kernels, SECTION 3 Provides information about the Database, SECTION 4 explains Feature Extraction, SECTION 5 gives the results and discussion and finally SECTION 6 gives the conclusions.

2 SUPPORT VECTOR MACHINE

A Support Vector Machine (SVM) performs classification by constructing an N-dimensional hyper plane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network.

Support Vector Machine (SVM) models are a close cousin to classical multilayer perceptron neural networks. Using a kernel function, SVM's are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training.

In the parlance of SVM literature, a predictor variable is called an attribute, and a transformed attribute that is used to define the hyper plane is called a feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case (i.e., a row of predictor values) is called a vector. So the goal of SVM modeling is to find the optimal hyper plane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other side of the plane (Fig 1). The vectors near the hyper plane are the support vectors.

The way of constructing a hyper plane to get binary classifiers done that can separate members of one class from others, but most real data hardly separate because the hyper plane that can successfully separate the members of the two classes in most case does not exist. One measure to solve this problem is to map the data into a higher dimensional space (Fig 2), where the members of the two classes can separate by a hyper plane. However, the traditional classifier is not good at in high dimensional vector. It is extremely expensive in terms of memory and time. Support Vector Machines can solve this problem. SVM avoid over fitting the data by choosing a hyper plane from the many that can separate the data. That maximizes the minimum distance from the hyper plane to the closest training point. Such a hyper plane calls the maximum margin hyper plane. Another advantage of the SVM is the compact representation of the decision boundary, so the number of support vectors is small as compared to the number of points in the training set.

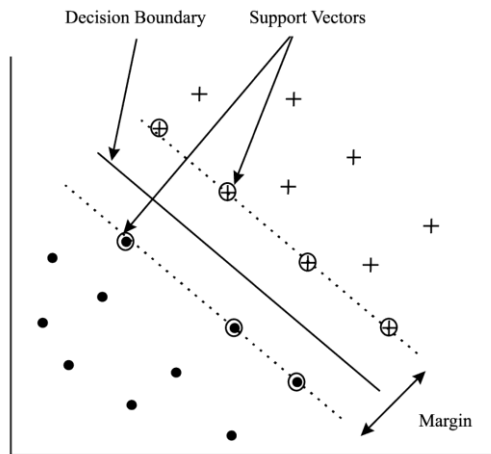


Fig 1 Support vector machine classifier with separating hyper planes and their equations

SVM generally are capable of delivering higher performance in terms of classification accuracy than the other data classification algorithms. However, for some datasets, the performance of SVM is very sensitive to how the cost parameter and kernel parameters are set. As a result, the user normally needs to conduct extensive cross validation in order to figure out the optimal parameter setting. This process is commonly referred to as model selection. One practical issue with model selection is that this process is very time consuming. We have experimented with a number of parameters associated with the use of the SVM algorithm that can impact the results. These parameters include choice cost parameter for different image categories and also its influence on the number of positive images added to the database. In the support-vector networks algorithm one can control the trade-off between complexity of decision rule and frequency of error by changing the parameter C .

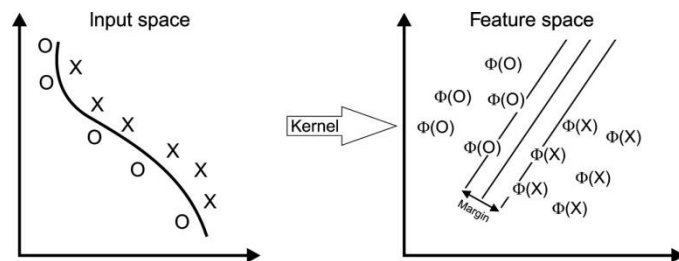


Fig 2 Support vector machine kernel mapping

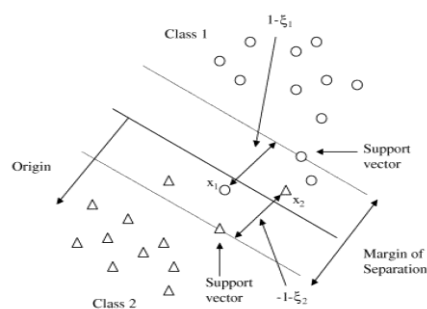


Fig 3 SVM- Linearly Non Separable Case

Ideally an SVM analysis should produce a hyper plane that completely separates the feature vectors into two non-overlapping groups. However, perfect separation may not be possible, or it may result in a model with so many feature vector dimensions that the model does not generalize well to other data; this is known as over fitting. To allow some flexibility in separating the categories, SVM models have a cost parameter, C , that controls the tradeoff between allowing training errors and forcing rigid margins. It creates a soft margin that permits some misclassifications. Increasing the value of C increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well.

3 DATABASE PREPARATIONS

The given classes of images are tested one by one with each linear kernel and Hellinger kernel with different cost parameters. The database consists of five image classes' airplanes, motorbikes, and people, horses and cars. It also contains background image which doesn't contain the above mentioned classes. The data is divided as per the below mentioned table which consists of the training and testing images for each class. The sample of Image database can be seen in Fig 4.

Table 1 Distribution of Image Database

	Bikes	Airplanes	People	Horses	cars
Training	120	112	1025	180	150
Testing	125	126	983	148	148
Total	245	238	2008	328	298

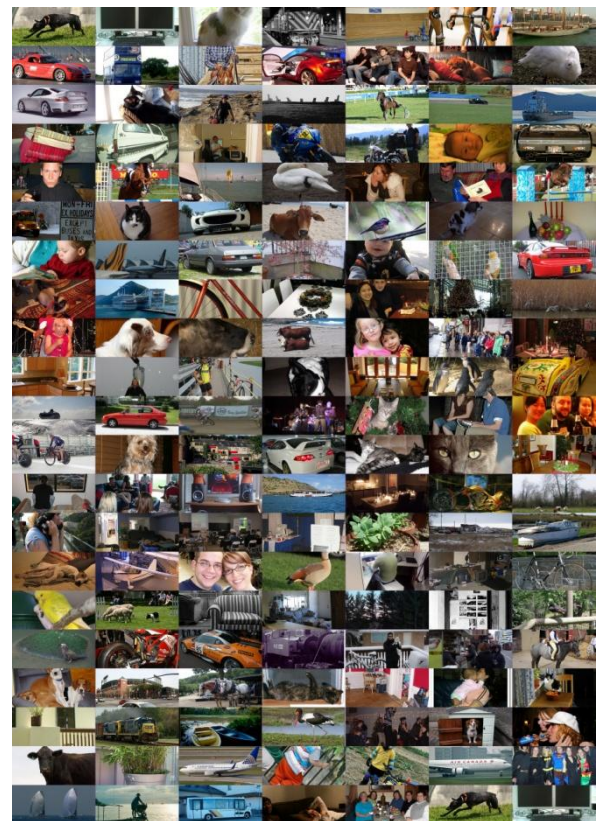


Fig 4 Sample of Image database

4 FEATURE EXTRACTIONS

It computes a dense set of multi-scale SIFT descriptors from a given input image. Vocabulary learning is then used to cluster a few hundred thousand visual descriptors into a vocabulary of 10^3 visual words. A spatial histogram calculates the joint distribution of appearance and location of the visual words in an image.

The feature vector consists of SIFT features computed on a regular grid across the image ('dense SIFT') and vector quantized into visual words. The frequency of each visual word is then recorded in a histogram for each tile of a spatial tiling. The final feature vector for the image is a concatenation of these histograms.

The Scale Invariant Feature Transform (SIFT) is probably the most popular feature used in computer vision. Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. The algorithm was published by David Lowe in 1999. It detects salient image regions (key points) and extracts discriminative yet compact descriptions of their appearance (descriptors). For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

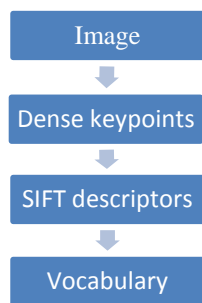


Fig 5 Feature extraction



Fig 6 Obtaining Spatial Histogram

The figs 5 & 6 illustrate the process of feature extraction in the corresponding steps. This may be used as a basis for a three-dimensional reconstruction of the scene. Alternatively, key points with discredited descriptors can be used as visual words as an intermediate image characterization. Histogram of visual words can then be used by a classifier to map images to abstract visual classes (e.g. car, cow, horse). Despite its popularity, the original SIFT implementation is available only in binary format. Dense SIFT is a fast algorithm for the computation of a dense set of SIFT descriptors.

The results of all the above cases are plotted using the precision and recall curve which gives statistical comparison of the methods. To assess the retrieval effectiveness, we have used the precision and recall as statistical comparison parameters.

It is given by

$$\text{Precision} = \frac{\text{No. of Relevant Images Obtained}}{\text{Total No. of Images Retrieved}}$$

$$\text{Recall} = \frac{\text{No. of Relevant Images Retrieved}}{\text{Total No. of Relevant Images Retrieved}}$$

5 RESULTS AND DISCUSSION

The below tables compare various classes of image database and their retrieval performance with linear kernel and Hellinger kernel with different cost parameters and effects of addition of positive images to the class of horses and cars.

The sample of the image database for both the Horse and Cars is shown separately and their best precision recall plot is also displayed.

The best values are also plotted considering the best cost parameter and by cross validating them.

Table 2
Results with Linear kernel

Classes	C	Test A.P	No. of images retrieved in Top 36	PR on Train	PR on Test
Airplanes	100	0.55	30	61.31	54.76
	500	0.58	31	64.33	57.80
	1000	0.60	32	65.66	59.77
Bikes	100	0.29	15	36.74	28.65
	500	0.30	16	38.45	29.96
	1000	0.31	15	43.68	31.41
People	100	0.61	26	68.63	61.35
	500	0.63	26	71.09	63.21
	1000	0.63	26	71.36	63.37
Horses	100	0.19	9	74.71	19.20
	500	0.20	12	76.05	20.29
	1000	0.20	10	76.22	19.76
Cars	100	0.42	21	73.66	41.62
	500	0.45	24	77.01	44.84
	1000	0.44	24	77.20	44.28

C-> Cost parameter

A.P -> Average Precision

PR-> Precision- Recall

Table 3
Results with Hellinger Kernel

Classes	C	Test A.P	No. of images retrieved in Top 36	PR on Train	PR on Test
Airplanes	100	0.66	30	100	65.62
	500	0.66	29	100	65.55
	1000	0.66	30	100	66.33
Bikes	100	0.69	34	100	69.42
	500	0.69	33	100	69.43
	1000	0.68	34	100	68.44
People	100	0.77	34	100	76.77
	500	0.76	35	100	76.39
	1000	0.76	34	100	76.45
Horses	100	0.44	28	100	43.28
	500	0.42	27	100	42.25
	1000	0.41	28	100	41.45
Cars	100	0.34	16	100	34.09
	500	0.37	21	100	36.93
	1000	0.36	19	100	36.12

It can be clearly seen from the above Table 2 & 3 that in linear kernel mapping of SVM, as we increase the cost parameter 'C', the average precision values both in testing and training increase significantly. But in the case of Hellinger kernel mapping, the performance seen is different. The precision recall values decrease slightly and then increase again back to the original. This proves that while using Hellinger kernel the cost parameter doesn't greatly influence the results but whereas in linear kernel care has to be taken in selecting the appropriate cost parameter and it has to be cross validated for obtaining satisfactory results. The plots proving the above tables are shown in the upcoming figures.

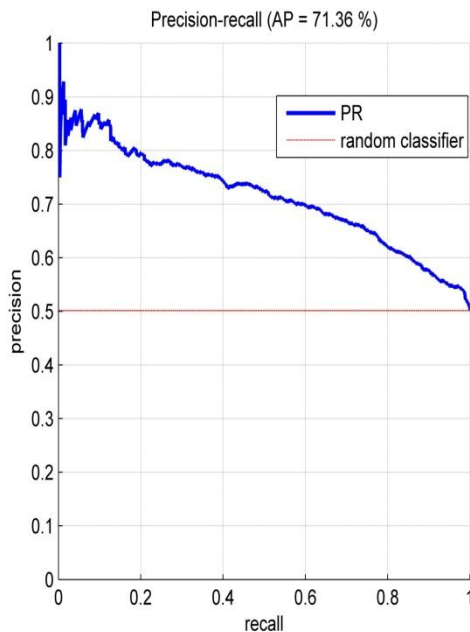


Fig 7 Best Precision Recall for people class with linear kernel at training

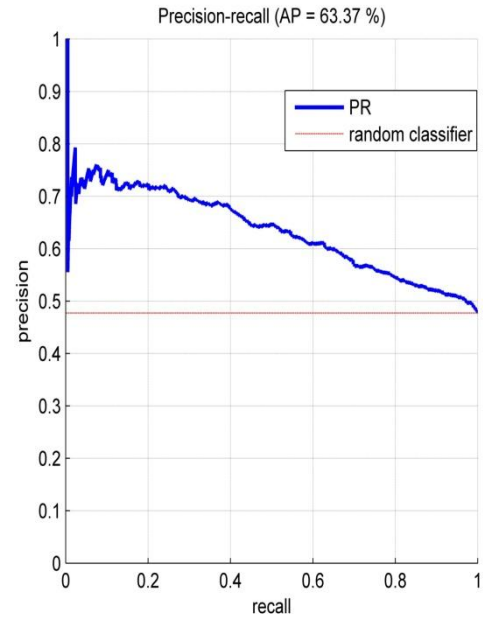


Fig 8 Best Precision Recall for people class with linear kernel at testing

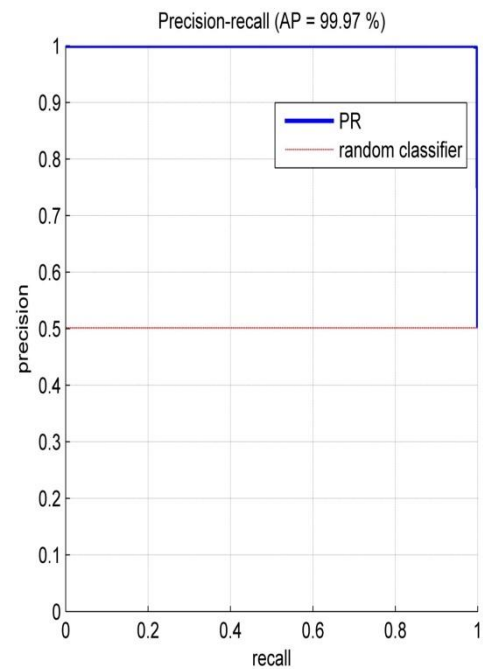


Fig 9 Best Precision Recall for people class with Hellinger kernel at training

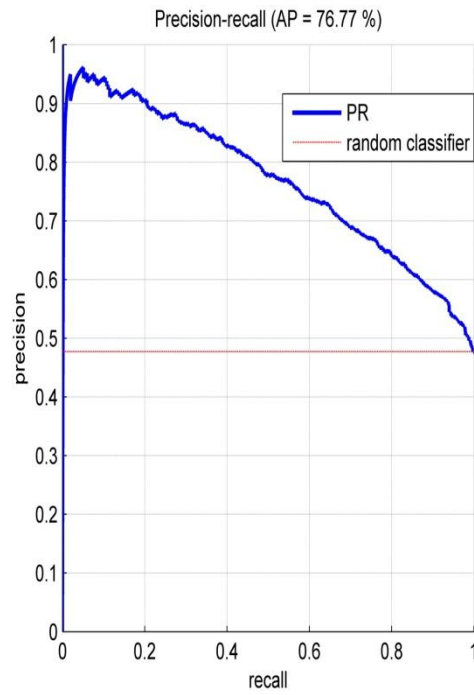


Fig 10 Best Precision Recall for people class with Hellinger kernel at Testing



Fig 12 Sample of car database



Fig 11 Sample of horse database

Table 4

Results with increase in positive images for linear kernel

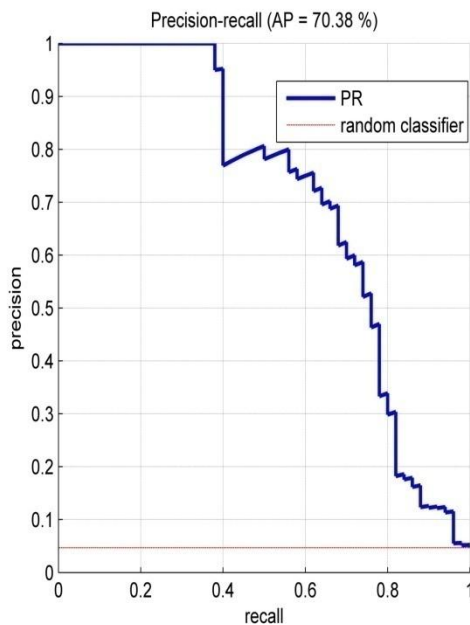
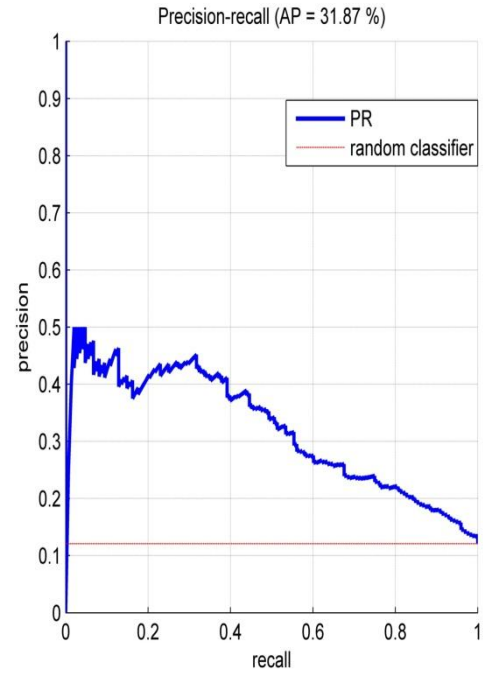
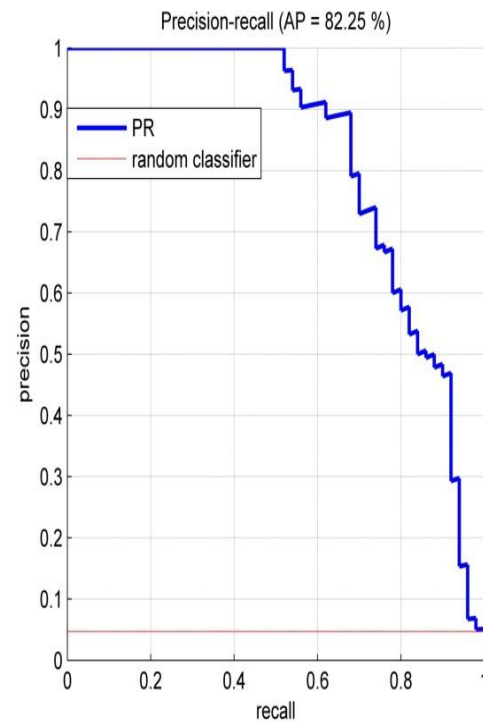
Classes	No. of positive images	Test A.P	No. of images Retrieved in the top 36	PR on train	PR on Test
Horses	5	0.20	11	10.31	20.19
	10	0.21	12	18.71	21.81
	20	0.23	13	23.51	23.31
	50	0.25	15	69.95	24.97
Cars	5	0.24	9	42.10	24.12
	10	0.25	11	25.36	25.59
	20	0.30	13	57.08	30.36
	50	0.39	16	80.11	38.69

Table 5**Results with increase in positive images for Hellinger kernel**

Classes	No. of positive images	Test A.P	No. of images Retrieved in the top 36	PR on train	PR on Test
Horses	5	0.32	18	100	31.92
	10	0.40	30	100	40.62
	20	0.45	32	100	45.92
	50	0.51	35	100	51.40
Cars	5	0.33	16	100	32.50
	10	0.34	18	100	34.45
	20	0.39	24	100	38.93
	50	0.46	26	100	45.67

From table 4 & 5 it infers that the increase in the number of positive images of a specific class increases the overall performance of the retrieval system and it can be seen in both the Linear and Hellinger kernel case.

The positive images as added increases the Test A.P and number of images retrieved in the top 36. The images added should not be the same as already added but should be different from the formerly added. The success of the system lies in the number of images it retrieves in the top 36.

**Fig 13 Precision Recall for horse class with linear kernel at training****Fig 14 Precision Recall for horse class with linear kernel at testing****Fig 15 Precision Recall for car class with linear kernel at training**

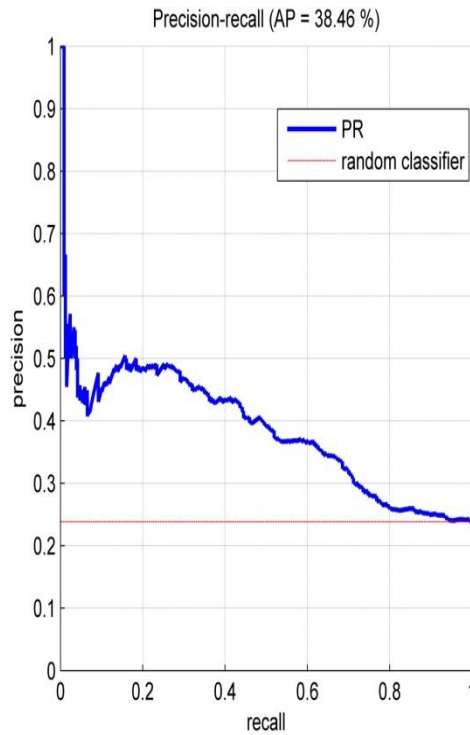


Fig 16 Precision Recall for car class with linear kernel at testing

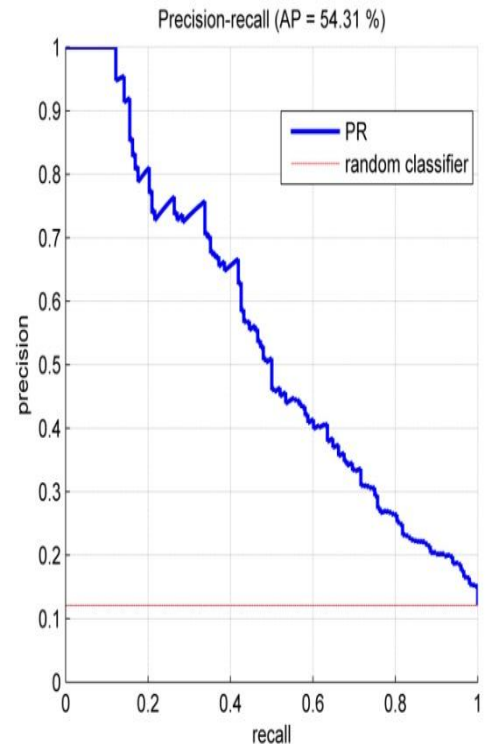


Fig 18 Precision Recall for horse class with Hellinger kernel at testing

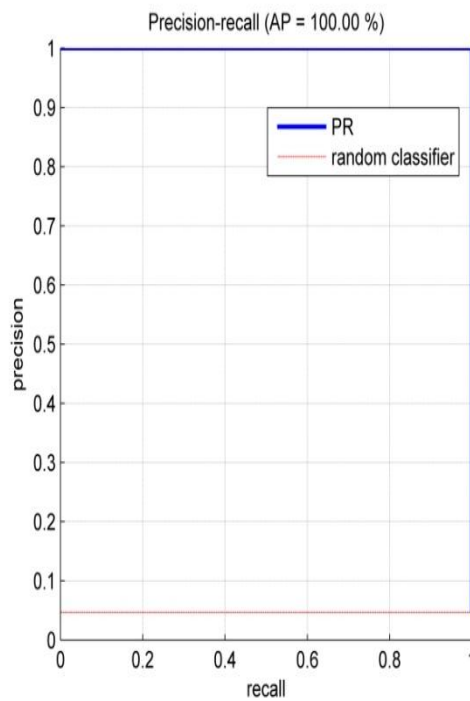


Fig 17 Precision Recall for horse class with Hellinger kernel at training

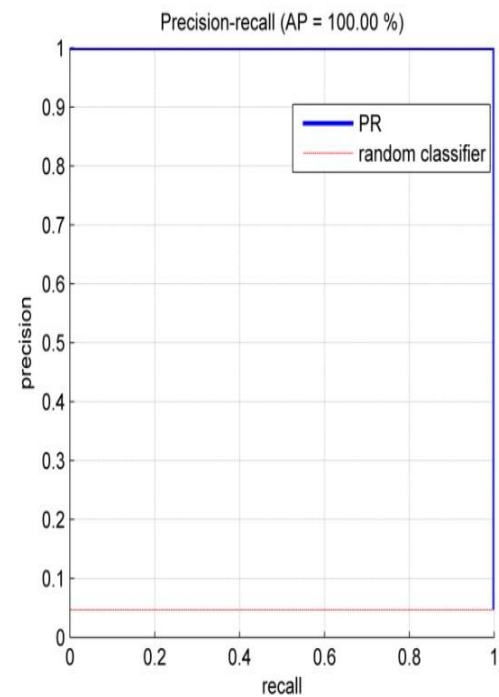


Fig 19 Precision Recall for car class with Hellinger kernel at training

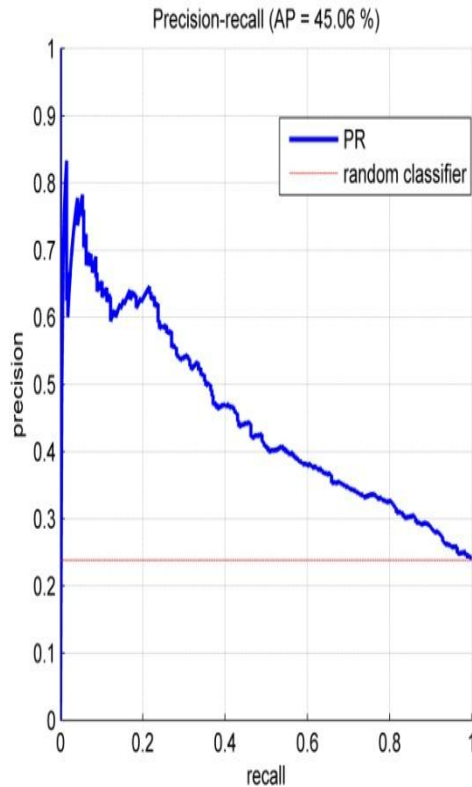


Fig 20 Precision Recall for car class with Hellinger kernel at testing

6 CONCLUSIONS

We have proposed a research paper which analyses the various cost parameter performance of SVM retrieval system of different classes. It can be seen that the linear kernel mapping and Hellinger kernel mapping differ in their results. This can be significant for research purposes to select the appropriate cost parameter and validate it. In this paper we have also presented the effects of adding more number of positive images to a database which showed enhanced performance. This paper will contribute significantly for the upcoming research works which will provide suitable pathway to select suitable cost parameters for SVM classification and increase the retrieval performance of the classification system.

7 REFERENCES

- [1] Vedaldi, A., and Zisserman, A. Efficient Additive Kernels via Explicit Feature Maps. June 2011. IEEE transactions on pattern analysis and machine intelligence, vol. xx, no. xx,.
- [2] Berkeley, U.C., Alexander C., and Malik, J. Classification Using Intersection Kernel Support Vector Machines is Efficient.
- [3] Rojas, M., Opido, I.D., Plaza, A., and Gamba, P. Comparison of Support Vector Machine-Based Processing Chains for Hyper spectral Image Classification. Satellite Data Compression, Communications, and Processing VI, edited by Bormin Huang, Antonio J. Plaza, Joan Serra-Sagristà, Chulhee Lee, Yunsong Li, Shen-En Qian, Proc. of SPIE Vol. 7810, 78100B.
- [4] Durgesh, k., and Iekha, B. Data Classification Using Support Vector Machine. Journal of Theoretical and Applied Information Technology.
- [5] Wang, Y., and Gang Hu, B. January 2002. Hierarchical Image Classification Using Support Vector Machines. ACCV2002: The 5th Asian Conference on Computer Vision, Melbourne, Australia.
- [6] Juan C.C., Cruz, A., and Fabio A.G. Histopathology Image Classification using Bag of Features and Kernel Functions. Bioingenium Research Group National University of Colombia.
- [7] Claudio, C., Ciocca, G., and Schettini, R. Image annotation using SVM.
- [8] Gidudu, A., Tshilidzi, M., and Gregg, H. Image Classification Using SVMs One-against-One Vs One-against-All.
- [9] Xiaohong Yu and Hong Liu. "Image Semantic Classification Using SVM in Image Retrieval". ISBN 978-952-5726-07-7 (Print), 978-952-5726-08-4 (CD-ROM) Proceedings of the Second Symposium International Computer Science and Computational Technology (ISCST '09) Huangshan, P. R. China, 26-28, Dec. 2009, pp. 458-461.
- [10] Boughorbel, S., Tarel, J.P., and Boujemaa, N. 2005. Generalized Histogram Intersection Kernel for Image Recognition. In ICIP, Genova, Italy.
- [11] VLFeat - An open and portable library of computer vision Algorithms