A modified concept of PCA to reduce the classification error using kernel SVM classifier

Dr. Neeraj Bhargava
Associate Professor
School of eng.&sys.science
MDSU, Ajmer
dr.neerajbhargava@yahoo.co.in

Abhishek Kumar
Comp. Science Dept
School of engg.&sys.science
MDSU, Ajmer
abhishekkmr812@gmail.com

Devesh Kumar
M.Tech Student
Comp. Science Engineering
Govt. Engg. College, Ajmer
devsharma1507@gmail.com

Meenakshi
Assistant Professor
Computer Science Dept.
Bhagwant University, Ajmer
Meenakshi.smiley@gmail.com

Abstract- This paper focuses on the mathematical technique PCA with the drawback of its mixing of data pixel. We have extracted principal directions of the covariance ellipse as done in PCA, but we will not blindly take the Eigen vectors corresponding to k largest values. Instead, we transform the data vectors into the new n– dimensional (n is dimension of old input space) vector space spanned by the Eigen vectors of the covariance matrix of the input data and then take one attribute at a time to perform classification. Then, we choose attributes corresponding to k - largest accuracy measures this approach of Eigen vector selection shall prove to be more effective than the traditional one with the improved approach to the conventional method in which we have considered the feature vector corresponding to the k- minimum error in order to reduce the dimensionality. The work has been implemented on more than 545 images of men and women to give efficient consequences in context of advance approach of PCA.

Index term: Principal Component Analysis, Support Vector Machine classifier classification error Euclidian distance kenel svm , matlab tool

1. INTRODUCTION

Principal component analysis (PCA) was given to the world in 1901 by Karl Pearson. PCA is a Dimension reduction procedure and useful for compact relevant information when obtained data have some redundancy. This will result into reduction of variables into smaller number of variables which are called Principal Components which will account for the most of the variance in the observed variable PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences of the face images. Each face image in the database set can be represented exactly in terms of a linear combination of the Eigen faces. This technique allows the system to represent the necessary information for comparing the faces using the little information once the mathematical representation accomplished which it is need to have a lot of faces to be store. On the other hand, it suffers a bit from the fact that facial image have to be normalized that meaning they all have to be the same size and the eyes, nose, and mouth in the sample images must be lined up before the PCA applied. Image Processing Toolbox will provide a set of algorithms and graphical tools for image processing analysis, algorithm development, and visualization, pattern recognition and many other fields. As far as Image Processing Toolbox is concerned it supports scientists and engineers in areas such as remote sensing, biometric surveillance, gene expression, semiconductor testing, image sensor design, and color science. It also facilitates the learning and teaching of image processing techniques. The Principal Component Analysis is one of the most efficient and accurate techniques that have been used in gender recognition. PCA is using mathematical rigor under the massive title of factor analysis. The PCA algorithm reduce the large dimensionality of the data we are having to the smaller dimensionality of independent feature

space which are needed to describe the data. The jobs of PCA are to do data compression, prediction, redundancy removal, feature extraction, etc. As PCA is a mathematical technique which can perform something in the linear domain, applications having linear models are suitable, such as image processing, signal processing, system and communications control theory, etc. Face recognition has many applicable areas. Moreover, it can be categorized into face identification, face classification, or sex determination. The most useful applications contain crowd Surveillance, video content indexing, personal identification (ex. driver's license), mug shots matching, entrance security, etc. We will extract principal directions of the covariance ellipse as done in PCA, but we will not blindly take the Eigen vectors corresponding to k largest values. Instead, we transform the data vectors into the new n – dimensional (n is dimension of old input space) vector space spanned by the Eigen vectors of the covariance matrix of the input data and then take one attribute at a time to perform classification. Then, we choose attributes corresponding to k - largest accuracy measures.

2. PRINCIPAL COMPONENT ANALYSIS

In this paper, we implemented a face recognition system using the Principal Component Analysis (PCA) algorithm. Automatic face recognition systems try to find the identity of a

given face image according to their memory. The memory of a face recognizer is generally simulated by a training set. In this paper, our training set consists of the features extracted from known face images of different persons. Thus, the task of the face recognizer is to find the most similar feature vector among the training set to the feature vector of a given test image. Here, we want to recognize the identity of a person where an image of that person (test image) is given to the system. We have used advance approach of PCA as a feature extraction algorithm in this paper as an extension of our work in [1,7]. In the training phase, we should extract feature vectors for each image in the training set. Let Ω_A be a training image of person A which has a pixel resolution of *M* *N* (*M* rows, *N* columns). In order to extract PCA features of Ω_A We will first convert the image into a pixel vector ©*A* by concatenating each of the *M* rows into a single vector. The length (or, dimensionality) of the *A* vector will be *M* *N*. In this project, you will use the PCA algorithm as dimensionality reduction technique which transforms the vector ©*A* to a vector $\mathbf{w}^a$, *which* has a dimensionality *d* where *d* > > *M** *N*. For each training image ai, you should calculate and store these feature vectors!*i*. In the recognition phase (or, testing phase), you will be given a test image Ω*j* of a known person. Let αj be the identity (name) of this person. As in the training phase, you should compute the feature vector of this person using PCA and obtain!*j* . In order to identify, $\mathbf{a}^j$ you should compute the similarities between $\mathbf{w}^j$ and all of the feature vectors' $\mathbf{w}^i$ s in the training set. The similarity between feature vectors can be computed using Euclidean distance. The identity of the most similar $\mathbf{w}^i$ will be the output of our face recognizer. If *i* = *j*, it means that we have correctly identified the person *j*, otherwise if i ≠ *j*, it means that we have misclassified the person *j*. of computational resources. Thus, to sum up, the jobs which the PCA technique can do are prediction, redundancy removal, feature extraction and data compression.

A. Mathematical approach of the PCA

We are considering 2d image for working so first of all it will get converted in to 1 d vector by concatenating rows and columns T Suppose we have M vectors each of size N (rows * columns) representing a set of sampled images. Let 'p_j' represent the values of the pixels.

$$X_i = [\, p_1, p_2, p_3, p_4, \ldots, p_N \,] \; ; \; i = 1,2,3,4,\ldots M. \qquad (1)$$

then the images are mean centered when we subtract the mean image from each image vector. Let us suppose m as the mean image:

$$m = (1/M) * (\Sigma X_i) \qquad (2)$$

Let Wi be the mean centered image:

$$W_i = X_i - m \qquad (3)$$

Ultimately we have to find the values of ei's which have the largest possible projection onto each of the wi's. The purpose is to get M orthogonal vectors ei for which the quantity

$$\Lambda_i = (1/M)\, \Sigma\, (e_i{}^T * w_n)2 \qquad (4)$$

is normalized with the orthogonality constraint:

$$e_i{}^T * e_k = \delta_{lk} \qquad (5)$$

The values of e_i's and λ_i's are calculated from the Eigen vectors and the Eigen values of the covariance matrix:

$$C = W * W_T \qquad (6)$$

W is a matrix formed by the column vectors w_i places side by side. The size of the covariance matrix is enormous (N*N). It is not possible to solve for eigenvectors directly. In mathematics, there are areas where one needs to find the numbers λ and the vectors v that satisfy the equation where A is the square matrix:

$$Av = \lambda v \qquad (7)$$

Any λ satisfying the above equation is the Eigen value of A. The vector v is called the eigenvector of A. The Eigen values and eigenvectors are obtained by solving the equation:

$$[A - \lambda I] = 0 \qquad (8)$$

3.PROBLEMS WITH CONVENTIONAL PCA

In traditional PCA the larger eigenvectors corresponding with the larger Eigen values are selected as principal components, which make the variance of data minimum. However, there are still some questions of PCA should be addressed as the following. Firstly, The idea of PCA is correct from image compression point of view; keeping the largest nonzero principal components means that we keep most of the energy (information) of that image by projecting into lower dimension subspace. However, with a view for pattern classifications this argument may not be true. The main reason is that, in pattern classification we would like to find a set of projection vectors that can provide the highest discrimination between different classes. Thus, choosing the largest principal components as the bases for dimensionality reduction may not be optimal. Secondly, PCA is unsupervised and statistical type algorithm to extract features in input data since it does not use the class information of input data. In specific case, the extracted features may not be fitted for classifications. Therefore, the principal components are not always useful for classification since they are not the most discriminating feature. Some of the small principal components may have better performance of classification than the selected large principal components AS we have discussed in the earlier segment emphasizing that PCA is no doubt a revolutionary tool as far as dimension reduction is concerned but when it comes in pattern recognition sometimes it mixes the data pixel, for example it could be problematic when there is moustache and without moustache image of man then it could be confusion in getting the right density of the particular area of the face [5], hence k nearest neighbor concept at that particular time could be altered with considering some assumptions to get overcome this particular problem, As discussed above that Eigen value and Eigen vector are so basics and parametric that it is impossible to bring about any commutation in that so we have consider PCA to get altered a bit with respect to the error minimization as compared to the traditional approach this is why it is titled as " Error Minimization In Gender Recognitions With Improved PCA.

Our Proposed Solution-In addition to the traditional PCA which sometimes mixes the data pixel that will further increase the classification error as final consequence. The concept of minimization of approximation error in traditional

PCA [4] is get applicable in my work. When the data points get extracted from the image it should be hold their optimal position so that better classification result get occur, we have commuted some concept in traditional PCA in a way that we have considered k- minimum feature vector extracted from the image so what We have done basically is have choose a feature vector projected along k- minimum error, instead of considering the k- largest Eigen value corresponding feature vector.

Now the main and the basic thing in doing the above alteration in conventional PCA is to rely upon some tool or method, which will calculate the error with great degree of accuracy so we have chosen [1,7]. We have chosen the linear as it provides almost the same accuracy as the radial basis but the complexity and calculation is less in linear . The main reason to choose, with linear trick is not to face the extensive calculation with the application of radial basis and polynomial. So it will prove to be risky, costly and time consuming

Fig. 1 Mean Image for Men and Women created from the respective database images

4. SUPPORT VECTOR MACHINE

We have used SVM for calculating the classification error once the classification error calculated easily the vector can be projected by taking the minimum error Eigen value. The basic notion of SVM is to map a linear non-separable input vector in to some higher dimensional space so that the more optimal hyper plane can be found with minimum classification error, so what I have used in my work is linear SVM just to calculate the classification error so the ultimate aim to minimize the network error comparatively to the conventional PCA. We have used SVM to calculate the classification error on the basis of which the vector would be projected along the k minimum error value. SVM is a revolutionary procedure for data classification. Even though it's considered that Neural Networks are easier to use than this, however, sometimes unsatisfactory results are obtained. As far as a classification task is concerned it usually involves with training and testing datasets which consist of some data instances. Each instance in the training set contains one target values and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes Classification in SVM is an example of Supervised Learning. Known labels help indicate whether the system is performing in a right way or not. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly. A step in SVM classification involves identification as which are intimately connected to the known classes. This is called feature selection or feature extraction. Feature selection and SVM classification together have a use even when prediction of unknown samples is not necessary. They can be used to identify key sets which are involved in whatever processes distinguish the classes

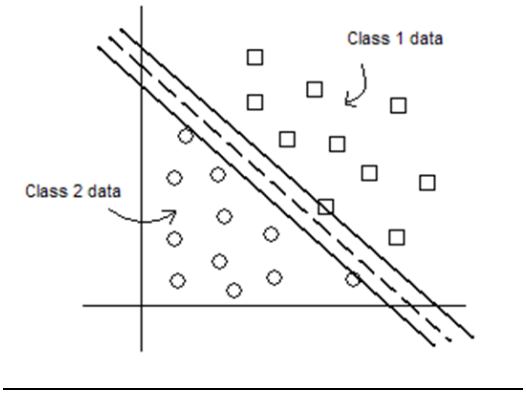

Fig.2 Distributions of support vectors

5. PROPOSED ALGORITHM WITH ALTERATION IN CONVENTIONAL PCA

Iterative_PCA$_{\text{min error}}$ (D^M, e_t)

1. Calculate covariance of $D^M \rightarrow D^{mx}$
2. Perform PCA on D^m to obtain $D^M{}_P$ with PCA-scores arranged in ascending order of the most significant Eigen value projected along k- minimum classification error.
3. Get the pc-score with most significant Eigen value=$e_{\text{min error}}$ with count being k (say).
4. Z score normalizes the k pc- score.
5. M_i= MEAN OF Z normalized pc- scores.
6. M_i=zero vector=0 m_{serror}=most significant unit Eigen-vector with respect to k- minimum error.

While d (M_i, m_{serror})> d (M_{i+1}, m_{serror}) $m_{serror} \leftarrow$ most Significant unit Eigen vector e^t=(e^t - Δe^t).

7. $D^m \leftarrow D^m$ – (data records corresponding to k – minimum error- pc score).
8. If D^m is not empty repeat steps 1 to 6 with new D^M and e^t.

The algorithm we have contemplated is based on traditional PCA algorithm considering the assumption that data points are independent of each other otherwise it would be capacious and complex calculations, so what alteration I have done is basically is to project the feature vector projected corresponding to k- minimum error unlike the traditional PCA consider the feature vector projected along k- largest Eigen value , the motivation of this work is not to recognize the face basically because lots of work has already been done in this respect but to minimize the overall classification error as my proposed method is completely based on the projection of vector along minimum classification error.

In the conventional PCA we consider the vector projected along k largest Eigen value like-

λ_{max1} , λ_{max2}, λ_{max3} ....., $\lambda_{max\ n}$

Fig. 3 Reconstructed Image using Men's database

But the altered approach has something else to take in to consideration and that is classification error, that's why I have termed my approach as advance approach of PCA .In my method we have considered the Eigen value corresponding to k-minimum error like-

$\lambda_{\text{min error1}}$, $\lambda_{\text{min error2}}$, $\lambda_{\text{min error3}}$ ..... $\lambda_{\text{min error n}}$.

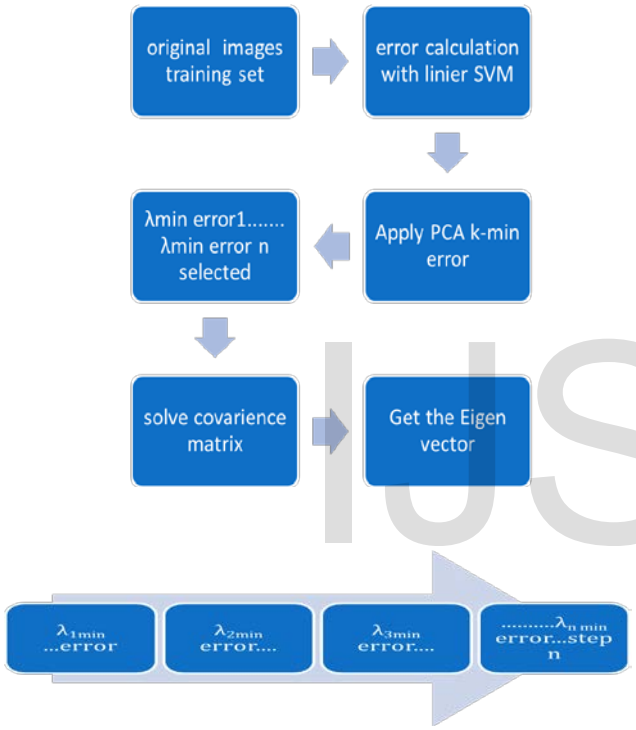

Fig4. New Approach of $PCA_{\text{MIN ERROR}}$

6. Implementation and experimental results

We have considered 20 images at one time to proceed further as far as supportive evidence to our approach is concerned We have tried out to get the result for the reliability of the method We have proposed. So in order to get the efficient result We have taken 545 images once worked on each one altogether at one time and calculated the classification error of the vector corresponding to k- largest value and every time we have got the result biased toward our proposed method in the context of classification error[1,7]. What We have observed that in conventional approach λ is not increasing or decreasing in particular pattern. If there would be the outcome like "greater the Eigen value corresponding to the smallest classification error then our proposed method would certainly be diminished reliability wise but we have to be continuous with the assumption that all data points are independent with each other. This very implementation provides a reliable platform to advance approach of PCA [1] for Gender recognition. We have made that assumption with a view to ignore any complexity and capacious calculation. The result of the implementation is given below, which is illustrating clearly that the classification error corresponding to the Eigen value can be increase or decrease in random pattern i.e. the conventional PCA is not completely reliable as far as classification error is concerned, the point to give this proof through the experimental results is just to provide a reliable platform to our proposed concept as the output generated in the result is showing classification error

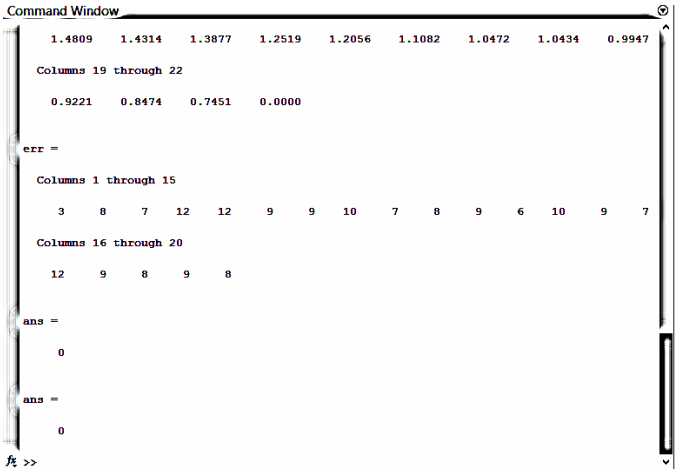

Fig. 5. Supporting evidence of drawbacks of PCA

So the above result that our new approach of PCA has given is the lesser classification error .We have implemented our work to get the right classification error percentage by taking 545 images of males and females and then by calculating the classification error of each value of λ going through with column wise. The final result we have got that is 50% with my approach as compare to the classical technique which has given 51% of classification error. We have gone through with bigger image data base from IIT Kanpur in order to get better accuracy. Primarily we have been working on the reduction of classification error in dimension reduction by PCA. We are using it in speech recognition or face recognition or elsewhere. We have used the notion that PCA mixes data pixel sometimes which further results in misclassification. We have proposed this very work in more hypothetical manner previously [1, 7].The future work will conspire more strengthen concept than the work recognized so far, we will consider more images and lots of other data to work on further with an aim to reduce the dimension in such a way that classification error would be minimized efficiently .The algorithm proposed in the work with an attempt to make efficient use of classical approach of PCA with an additional advance approach to consider the Eigen value corresponding to the k minimum error.

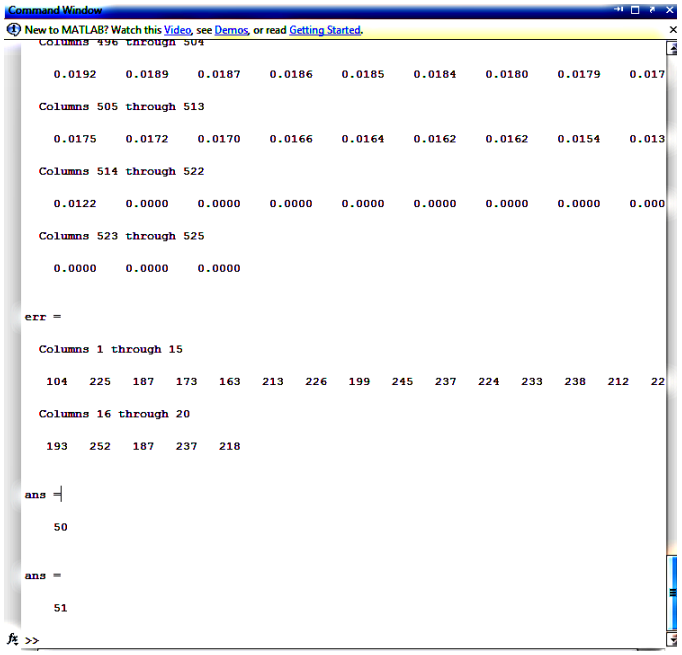

Fig. 6.Final Result for Calculation of classification Error

7. CONCLUSION AND FUTURE WORK

After doing a thoroughly review of PCA application in the domain of face or gender recognition along with the different algorithm scope in this field such as back propagation, support vector machine, graph matching. Although my main emphasis is on the improvement of conventional PCA in the direction of classification error minimization by calculating the generalized error in the Eigen value or feature vector going further to be used, while doing the review regarding the gender recognition with PCA what we have got that conventional PCA has scope in it to get improved in regarding the classification error. The paper [5,7] has given a healthy and important clue in this direction. The main drawback of PCA is that it mixes data points at some phase of classification which generally mislead the classification procedure as this process has lots of scope to ignore the error so in order to overcome this very drawback and with a aim to minimize the classification error. In the proposed algorithm, We are approaching with altered concept of PCA analysis with some concept of kernel SVM which is a training method basically used for error minimization,

There is lots of other approach that have been proposed but to deal with PCA approach is much more efficient as it is simple for use and implementation very robust in nature, but there is a problem with PCA as it sometimes mixes the data points [5] together which can further proved to be great problem for classification.

In the proposed algorithm, We have approached with adaptive concept of PCA analysis with some application notion of kernel SVM which is a training method basically used for error minimization, There is lots of other approach that have been proposed but to deal with PCA approach is much more efficient as it is simple for use and implementation very robust in nature, but there is a problem with PCA as it sometimes mixes the data points [5] together which can further be proved as great problem for classification. The existing work [1,7] shows the application of conventional PCA,along with the advance approach of PCA, in which the feature vector considered corresponding to k largest value but what my work primarily all about is to minimize the network error rather than to recognize the face. With the improved approach to the conventional method in which We have considered the feature vector corresponding to the k- minimum error in order to reduce the dimensionality, we have use kernel SVM further to calculate the error and in order to achieve better accuracy comparatively. In order to have better classification, low classification error is very important in training sample, so kernel SVM will play an important role in error calculation. Unlike conventional PCA we are considering minimum error Eigen value and then we are using Euclidean distance method in order to get the similarity between the faces in training set and test set. We are planning in future to include more images and make our technique to go through artificial neural network, SVM, in order to check out its reliability and flexibility and heterogeneous compatibility.

REFERENCES.

[1] Abhishek Kumar,Deepak Gupta,Karan Rawat “An Advance Approach of PCA For Gender Recognition”ICICES IEEE ISBN-978-1-4673-5786-9 ,Feburary 2013

[2] Jain, A. K. Ross, A. Prabhakar, S.,"An Introduction to Biometric Recognition IEEE Transactions on circuits and systems for video technology, 2004, Vol. 14; No. 1, pp. 4-20, 2004.

[3] Tan X, Chen S, Zhou Z, Zhang F (2006) Face recognition from a single image per person: a survey. Pattern Recogn 39(9):1725–1745

[4] Abhilash Alexander Miranda, Yann-Ael Le Borgne, Gianluca Bontemp September 12, 2007 New Routes from Minimal Approximation Error to Principal Components

[5] Hans-Peter Kriegel, Peer Kr¨oger, Erich Schubert, Arthur Zimek In Proc. 20th Int. Conf. on Scientific and Statistical Database Management (SSDBM), Hong Kong, China, *2008*

[6]Dr. H.B. Kekre et. al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 959-964

[7]Abhishek Kumar,Deepak Gupta,Karan Rawat “ An Advanced Approach of Face alignment for Gender recognition Using PCA” SPRINGER LINK (AISC SERIES) SOCPROS ISSN NO-2194-5357 December 2012

[8] Nello Cristianini and John Shawe-Taylor, “An Introduction to Support Vector Machines and Other Kernel-based Learning Methods”, Cambridge University Press, 2000

[9] A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, M.A. Sid-Ahmed Evaluation of face recognition technique using PCA, wavelets and Pattern Recognition, Volume 44, Issues 10–11 elsevier2011

[10] Wankou Yang Laplacian bidirectional PCA for face recognition, Neurocomputing, Volume 74,elsevier2010

[11]] H. C¸ evikalp, M. Neamtu, M.Wilkes, and A. Barkana. Discriminative common vectors for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1):4–13, 2005

[12] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. IEEE MTransactions on Pattern Analysis and Machine Intelligence, 23(6):681–685, 2001

[13] S. Duffner and C. Garcia. A connexionist approach for robust and precise facial feature detection in complex scenes. In Fourth International Symposium on Image and Signal Processing and Analysis (ISPA), pages 316–321, Zagreb, Croatia, September 2005

[14] S. Duffner and C. Garcia. A neural scheme for robust detection of transparent logos in TV programs. In International Conference on Artificial Neural Networks (ICANN), volume 2, pages 14–23, Athens, Greece, September 2006

[15] S. Duffner and C. Garcia. Face recognition using non-linear image reconstruction. In International Conference on Advanced Video and Signal- Based Surveillance (AVSS), London, UK, September 2007

[16] S. Duffner and C. Garcia. An online backpropagation algorithm with validation error-based adaptive learning rate. In International Conference on Artificial Neural Networks (ICANN), volume 1, pages 249–258, Porto, Portugal, September 20