# How to Train a Classifier Based on the Huge Face Database?

Jie Chen[1], Ruiping Wang[2], Shengye Yan[2],
Shiguang Shan[2], Xilin Chen[1,2], and Wen Gao[1,2]

[1] School of Computer Science and Technology,
Harbin Institute of Technology, 150001, China
[2] ICT-ISVISION Joint R&D Lab for Face Recognition,
Institute of Computing Technology, Chinese of Academy of Sciences,
Beijing, 100080, China
{jchen, rpwang, syyan, sgshan, xlchen, wgao}@jdl.ac.cn

**Abstract.** The development of web and digital camera nowadays has made it easier to collect more than hundreds of thousands of examples. How to train a face detector based on the collected enormous face database? This paper presents a manifold-based method to subsample. That is, we learn the manifold from the collected face database and then subsample training set by the estimated geodesic distance which is calculated during the manifold learning. Using the subsampled training set based on the manifold, we train an AdaBoost-based face detector. The trained detector is tested on the MIT+CMU frontal face test set. The experimental results show that the proposed method is effective and efficient to train a classifier confronted with the huge database.

## 1 Introduction

Over the past ten years, face detection has been thoroughly studied in computer vision research for its wide potential applications, such as video surveillance, human computer interface, face recognition, and face image database management etc. Face detection is to determine whether there are any faces within a given image, and return the location and extent of each face in the image if one or more faces are present [31]. Recently, the emphasis has been laid on the data-driven learning-based techniques, such as [7, 13, 14, 15, 19, 20, 21, 22, 30]. All of these schemes can be found in the recent survey by Yang [31]. After the survey, the methods based on boosting are much researched. Viola described a rapid object detection scheme based on a boosted cascade of simple features. It brought together new algorithms, representations and insights, which could broaden the applications in computer vision and image processing [23]. And the algorithm has been further developed by other researchers [11, 12, 28].

The performance of these learning-based methods highly depends on the training set, and they suffer from a common problem of data collection for training. It makes easier to collect more than hundreds of thousands of examples with the development of web and digital camera nowadays. How to train a classifier based on the collected immense face database? This paper will give a solution.

W. Zhao, S. Gong, and X. Tang (Eds.): AMFG 2005, LNCS 3723, pp. 84 – 95, 2005.

In nature, how to train a classifier based on the collected immense face database is a problem of data mining. In this paper we will use the knowledge of the manifold to subsample a small subset from the collected huge face database. Manifold can help us to transform the data to a low-dimensional space with little loss of information, which can enable us to visualize data, perform classification and cluster more efficiently. Recently, some representative techniques, including isometric feature mapping (ISOMAP) [25], local linear embedding (LLE) [17], and Laplacian Eigenmap [1], have been proposed. The ISOMAP algorithm is intuitive, well understood and produces reasonable mapping results [9, 10, 29]. Also, it is supported theoretically, such as its convergence proof [2] and it can recover the co-ordinates [4]. There is also a continuum extension of ISOMAP [32]. A mixture of Gaussians is applied to model a manifold and recover the global co-ordinates by combining the co-ordinates from different Gaussian components [3, 18, 24, 26], or by other methods [27]. To estimate the intrinsic dimensionality, different algorithms also have been considered in manifold learning [8, 16].

The main contributions of this paper are:

1. Subsample a small but efficient and representative subset from the collected huge face database based on the manifold learning to train a classifier.
2. Discuss the effect of outliers on the trained classifier.
3. The performance is instable to train a detector based on the random subsampling face set from a huge database. However, a detector trained based on the subsampled face set by the data manifold is not only stable and but also can improve the detector performance.
4. When we prepare the training set, we should collect more samples along those dimensionalities with larger variances to get a nearly uniformed distribution in the manifold, for example, left-right pose of face more than up-down pose.

The rest of this paper is organized as follows: After a review of ISOMAP, the proposed subsampling method based on the manifold learning is described in section 2. Experimental results are presented in section 3, followed by discussion in section 4.

## 2 Subsampling Based on ISOMAP

As discussed in [25], for two arbitrary points on a nonlinear manifold (for example, in the "Swiss roll" manifold), their Euclidean distance in the high dimensional input space may not accurately reflect their intrinsic similarity, as measured by geodesic distance along the manifold. Therefore, we use the geodesic distance for subsampling and the geodesic distance can be calculated as in ISOMAP. That is to say, the smaller the geodesic distance between two points is, the more their intrinsic similarity is. When the distance is smaller than a given threshold, one point is deleted as shown in Fig 2.

### 2.1 ISOMAP Algorithm

The goal of learning the data manifold is to show high-dimensional data in its intrinsic low-dimensional structures and use easily measured local metric information to learn the underlying global geometry of a data set [25].

In the ISOMAP algorithm, firstly, distances between neighboring data points are calculated. The neighborhood can be *knn* -neighborhood. ISOMAP supposes that the data set $X$ lie on a manifold of dimension $d$ and tries to find the global co-ordinates of those points on the manifold. And then an undirected neighborhood graph is constructed.

Secondly, for each pair of non-neighboring data points, ISOMAP finds the shortest path through neighborhood graph, subject to the constraint that the path must hop from neighbor to neighbor. The length of this path (we call it "the estimated geodesic distance") is an approximation to the true distance between its end points (we call it "geodesic distance"), as measured within the underlying manifold. That is to say after embedding the high-dimensional data manifold into low-dimensional structures, we can use straight lines in the embedding to approximate the true geodesic path.

Finally, the classical multidimensional scaling is used to construct low-dimensional embedding.

### 2.2 The Residual Variance of ISOMAP

The residual variance ( *RVar* ) of ISOMAP denotes the difference between the Euclidean distance in the $d$-dimensional Euclidean space and the true geodesic path [25]. The less the value of *RVar* is, the more approximate between them. The intrinsic dimensionality of the data can be estimated by looking for the "elbow" at which this curve ceases to decrease significantly with added dimensions, i.e., the inflexion of the curve. The relationship between the ISOMAP embedding dimensionality and the residual variance of the 698 face image of [25] is shown in Fig. 1.

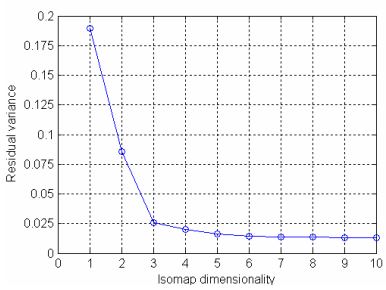


**Fig. 1.** The residual variance of ISOMAP embedding on the 698 face database of [25]

As discussed in [25], each coordinate axis of the embedding correlates highly with one degree of freedom underlying the original data: left-right pose corresponding to the first degree of freedom, up-down pose corresponding to the second one and lighting direction to the third one. That is to say the scatter of face images in left-right pose is the biggest while the scatter in up-down pose is the smallest among these three factors. We can conclude that, in order to select representative example set, we should pay more attention to the left-right pose variations than the up-down pose.

### 2.3 Subsampling Algorithm

As discussed in [25], during the manifold learning, we can get the estimated geodesic distance in the high-dimensional space between pairs of the data points. And then they can be used directly to sample by deleting some examples from the database. And the remained examples can still keep the data's intrinsic geometric structure basically. By this means, we can get a small representative subset of the huge data.

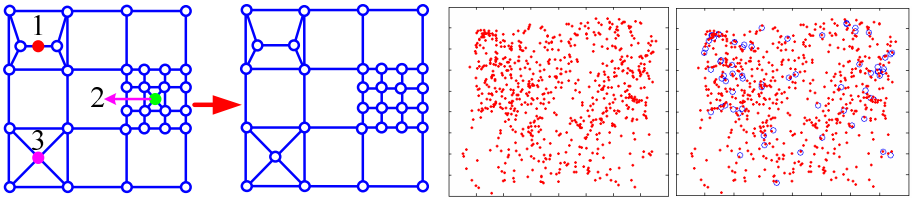


**Fig. 2.** Subsampling based on the manifold learning. (a) The schematic of subsampling based on the estimated geodesic distance; (b) the manifold of 698 faces; (c) the results of subsampling based on the estimated geodesic distance.

The scheme is demonstrated in Fig. 2 (a). We sort all of the estimated geodesic distances between pairs of points in the high-dimensional space in increasing order. If one of the estimated geodesic distances between an example and others is smaller than a given threshold, it will be deleted while others will be reserved. For example, as shown in Fig. 2 (a), the data point 1 and the data point 2 will be deleted during subsampling. As to the data point 3, it is preserved since the estimated geodesic distance between it and others are larger than the given threshold. From the figure of top right in Fig. 2 (a), the remained examples can still maintain the data's intrinsic geometric structure basically.

As demonstrated in Fig. 2 (b), it is a two-dimensional projection of 698 raw face images where the three-dimensional embedding of data's intrinsic geometric structure is learned by ISOMAP ($K$=6) [25]. Fig. 2 (c) is the results of subsampling where some data points (in circle) are deleted and the remained data points are still in solid dots.

If we want to subsample 90% examples from a whole set, what we need to do is to delete its 10% examples since their corresponding estimated geodesic distances to others are in the front of the sorted distance sequence.

## 3 Experiments

### 3.1 The AdaBoost-Based Classifier

A large number of experimental studies have shown that classifier combination can significantly exploit the discrimination ability comparing with individual feature and classifier. Boosting is one of the common used methods for combining classifiers. AdaBoost, a version of the boosting algorithm, has been used in face detection and is

capable of processing images extremely rapidly while achieving high detection rates [23]. Therefore, we use the AdaBoost algorithm to train a classifier. A final strong classifier is formed by combining a number of weak classifiers, which is described in Fig. 3. For the details of the AdaBoost based classifier, please refer to [23].

- Given example set ***S*** and their initial weights $\omega_1$;
- Do for $t$=1,……,T:
  1. Normalize the weights $\omega_t$;
  2. For each feature, $j$, train a classifier $h_j$ with respect to the weighted samples;
  3. Calculate error $\varepsilon_t$, choose the classifier $h_t$ with the lowest error and compute the value $\alpha_t$;
  4. Update weights $\omega_{t+1}$;
- Get the final strong classifier: $h(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$.

**Fig. 3.** The AdaBoost algorithm for classifier learning

### 3.2 Detector Based on the MIT Face Database

The data set is consisted of a training set of 6,977 images (2,429 faces and 4,548 non-faces) and the test set is composed of 24,045 images (472 faces and 23,573 non-faces). All of these images are $19\times19$ grayscale and they are available on the CBCL webpage [33].

We let $K$=6 when ISOMAP learns the manifold of 2,429 faces in MIT database. By the manifold learning as discussed in [25], we can get the estimated geodesic distance in the high-dimensional space between pairs. And then they can be used directly to sample by deleting some examples from the database.

Note that all of these examples are performed by histogram equalization before the manifold learning. It is because all examples to train a classifier are needed histogram equalization which maps the intensity values to expand the range of intensities.

In order to study the relationship between the distribution of the training set and the trained detector, we subsample the MIT face database by 90%, 80% and 70% (named as ISO90, ISO80, ISO70 later) as discussed in section 2.3. Subsampling by 90% is to say we reserve 90% examples of the database and the same meaning of 80% and 70%. Note that ISO70 is a subset of ISO80 and ISO80 is a subset of ISO90 in fact.

The three subsampled face sets together with the non-face are used to train three classifiers based on the AdaBoost as demonstrated in [23]. And then they are tested on the test set of MIT database. The ROC curves of these three classifiers are shown in Fig. 4. From these ROC curves, one can conclude that all of these three detectors base on ISO90, ISO80 and ISO70 get the comparable performance to the detector

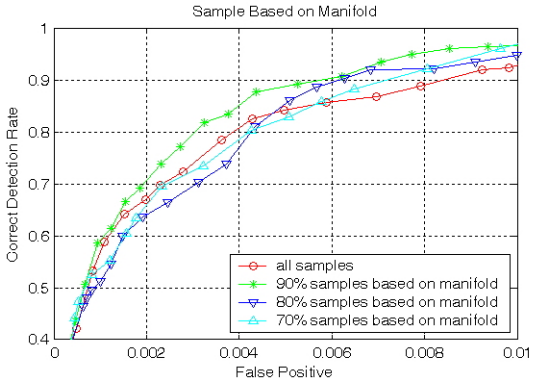


**Fig. 4.** The ROC curves on the MIT test set using the subsampling face example sets based on the manifold learning and the whole set as training set for a fixed classifier

based on the whole set. It demonstrates that it is reasonable to subsample based on the manifold, i.e. the subsampled subsets ISO90, ISO80 and ISO70 can still maintain the.data's intrinsic geometric structure basically. Further, the detector trained by ISO90 is the best of all and improves the performance of the detector distinctly compared with the detector even by the entire face examples in MIT database. Even the detector trained on ISO70 is a little better than the detector trained on the entire examples. Some possible reasons: the first one is the examples subsampled based on the manifold distribute evenly in the example space and have no examples congregating compared with the whole set; the second is that the outliers in the whole set deteriorate its performance which is to be discussed later.

However, random subsampling from the MIT database is not so lucky. We choose four subsets randomly-subsampled from the MIT database and each subset has the same number of examples as in ISO90. After trained on these four sets, they are also tested on the same test set. The ROC curves are shown in Fig. 5. In this figure, we plot the resulting ROC curves of detectors on the whole set, ISO90, and two randomly chosen subsets. Herein, the curve "*90% examples based on the random subsampling n1*" and the curve "*90% examples based on the random subsampling n2*" represent the best and the worst results of these four random sampling cases. From these ROC curves, one can conclude that the detector based on ISO90 is still the best of all and the results based on random subsampling is much instable. We also think that the evenly-distributed examples and no outliers contribute to this kind of results.

During the ISOMAP learning, we get 30 outliers. Using the examples by ISO90 plus the 30 outliers, we train a classifier also based on AdaBoost. Evaluated on the MIT test set, some resulting ROC curves are shown in Fig. 6. One can find that the detector, based on the ISO90 plus the 30 outliers, will deteriorate its performance. It also denotes that the effects of the evenly-distributed examples on the trained detector are more important than that of the outliers. Integrating these two factors, the detector based on the ISO90 is much better than the detector on the total face set.

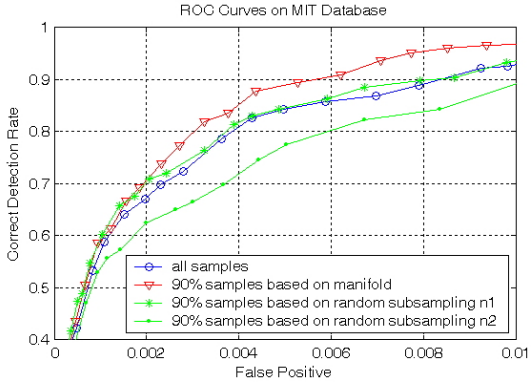


**Fig. 5.** The ROC curves on the MIT test set using the subsampling face example sets based on the manifold learning, two random sampling sets and the whole set as training set for a fixed classifier

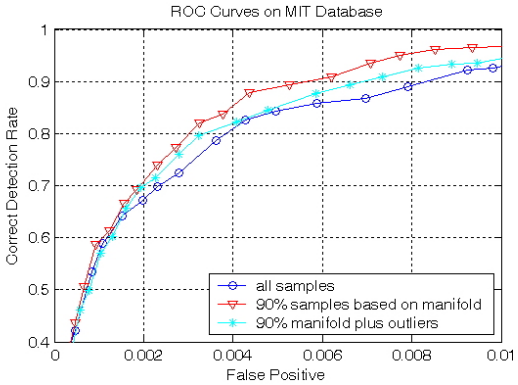


**Fig. 6.** The ROC curves on the MIT test set using the subsampling face example sets based on the manifold learning, the subsampling sets based on the manifold embedding plus outliers and the whole set as training set for a fixed classifier

### 3.3 Detector Based on the Huge Database

To compare the performance difference on different training sets further, we also use another three different face training sets. The face-image database consists of 100,000 faces (collected form web, video and digital camera), which cover wide variations in poses, facial expressions and also in lighting conditions. To make the detection method robust to affine transform, the images are often rotated, translated and scaled [6]. Therefore, we randomly rotate these samples up to $\pm 15^{\circ}$, translate up to half a pixel, and scale up to ±10%. After these preprocessing, we get 1,200,000 face images which constitute the whole set. The first group is composed of 15,000 face images which are sampled by the ISOMAP (called ISO15000, here). The second or third group is also composed of 15,000 face images which are random subsampling examples (named Rand1-15000 and Rand2-15000, respectively).

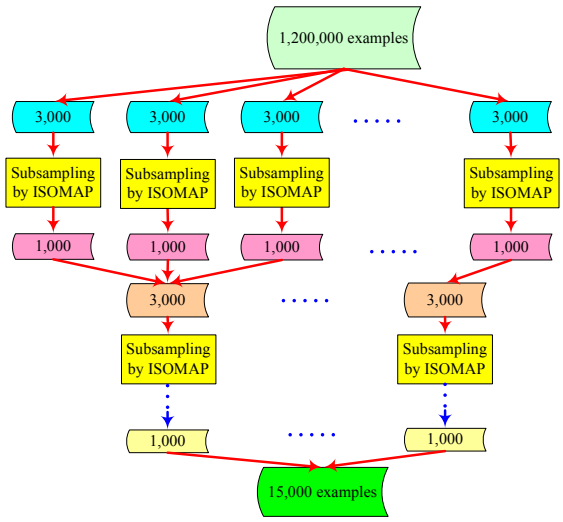


**Fig. 7.** Subsampling procedure by ISOMAP to get 15,000 examples from 3,000,000 examples

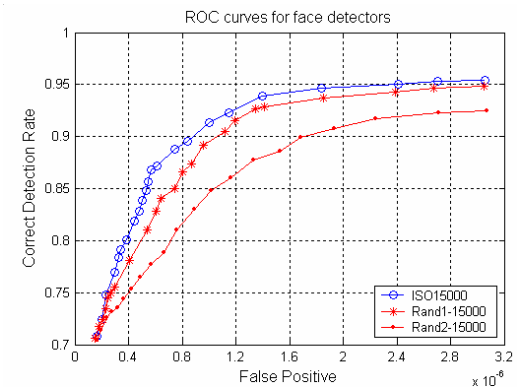


**Fig. 8.** The ROC curves for the trained detectors, based on the sampled training set by the ISOMAP and the random subsampling training set, tested on the MIT+CMU frontal face test set

It is hard to learn the manifold from 1,200,000 examples by the ISOMAP because it needs to calculate the eigenvalues and eigenvectors of a 1,200,000×1,200,000 matrix. In order to avoid this problem, as demonstrated in Fig. 7, we divide the whole set into 400 subsets and each subset has 3,000 examples. We get 1,000 examples by the proposed method from each subset and then incorporate every three subsampled sets into one subset. With the same procedure, we can get the total 15,000 examples after incorporating all subsampled examples into one set.

In order to avoid destroying the intrinsic structure of the manifold when the whole examples are divided into 400 subsets, we divide the samples in the similar distribution into the same subset. That is to say, the examples vary in poses, facial

**Table 1.** The detection rates comparison of our system and others

| Methods | Detection rate (%) | False alarms |
|---|---|---|
| Fröba [5] | 89.7 | 22 |
| Li [11] | 90.2 | 31 |
| Rowley [19] | 86.0 | 31 |
| Schneiderman [21] | 94.4 | 65 |
| Viola [23] | 89.7 | 31 |
| Xiao [28] | 88.2 | 26 |
| Our method | 91.28 | 18 |


**Fig. 9.** Output of our face detector on a number of test images from the MIT+CMU frontal face test set and other web images

expressions or lighting conditions are fallen into the different subsets respectively. And this criterion is also applied to incorporate the subsampled sets.

The non-face class is initially represented by 15,000 non-face images. Each single classifier is then trained using a bootstrap approach similar to that described in [22] to increase the number of negative examples in the non-face set. The bootstrap is carried out several times on a set of 13,272 images containing no faces.

The resulting detectors, trained on the three different sets, are evaluated on the MIT+CMU frontal face test set which consists of 130 images showing 507 upright faces [19]. The detection performances on this set are compared in Fig. 8. From these ROC curves one can conclude that the detector based on ISO15000 is the best of all and the results based on random subsampling is also much instable. During the

ISOMAP learning, we get 2,639 outliers. We think that the evenly-distributed examples and no outliers contribute this kind of results, again.

In table 1, the experimental results of our method is compared with the results reported on the same test set — MIT+CMU frontal face test set. We get the detection rate of 91.28% and 18 false alarms with the detector trained on the set ISO15000.

Herein, the results of Fröba [5] and Xiao [28] are read from the ROC curves given in their paper, which might result in a little difference with their real results. In this table, all of the algorithms in [5], [11], [23], [28] are based on boosting, [19] based on neural network, and [21] on Bayes. From the experimental results in table 1, one can conclude that our system outperforms the results achieved by Fröba [5], Li [11], Rowley [19], Viola [23] and Xiao [28]. Although the accuracy is lower than that of Schneiderman [21], our system is approximately 15 times faster. Furthermore, our system has less false detects than that of Schneiderman [21].

However, different criteria (e.g. training time, the number of training examples involved, cropping training set with different subjective criteria, execution time, and the number of scanned windows in detection) can be used to favor one over another, which will make it difficult to evaluate the performance of different methods even though they use the same benchmark data sets [31]. Some results of this detector are shown in Fig. 9.

## 4 Conclusion

In this paper, we present a novel manifold-based method to subsample a small but efficient and representative training subset from the collected enormous face database. After calculating the geodesic distance by learning the manifold from the collected face database, we subsample the training set in the high dimensional space. An AdaBoost-based face detector is trained on the subsampled training set, and then we test it on the MIT+CMU frontal face test set. Compared with the detector using random subsampling examples, the detector trained by the proposed method is more stable and achieve better face detection performance. We conclude that the evenly-distributed examples, due to the training set subsampled based on the manifold learning, and no outliers, discarded during the manifold learning, contribute to the improved performance.

## Acknowledgements

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, “100 Talents Program” of CAS, ShangHai Municipal Sciences and Technology Committee (No.03DZ15013), and ISVISION Technologies Co., Ltd.

## References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Inform. Proc. Systems 14*, pp.585-591. MIT Press, 2002.

[2] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. *Technical report, Department of Psychology, Stanford University*, 2000.

[3] M. Brand. Charting a manifold. In *Advances in Neural Information Proc. Systems 15*, pp. 961-968. MIT Press, 2003.
[4] D. L. Donoho and C. Grimes. When does ISOMAP recover natural parameterization of families of articulated images? *Technical Report 2002-27, Depart. of Statistics, Stanford Universit*y, 2002.
[5] B. Froba and A. Ernst, "Fast Frontal-View Face Detection Using a Multi-Path Decision Tree," *In Proceedings of Audio and Video based Biometric Person Authentication*, pp. 921-928, 2003.
[6] B. Heisele, T. Poggio, and M. Pontil. Face Detection in Still Gray Images. *CBCL Paper #187*. MIT, Cambridge, MA, 2000.
[7] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Machine Intell.*, pp.696–706, 2002.
[8] D. R. Hundley and M. J. Kirby. Estimation of topological dimension. In *Proc. SIAM International Conference on Data Mining*, 2003. http://www.siam.org/meetings/sdm03/proceedings/ sdm03_18.pdf.
[9] O. C. Jenkins and M. J Mataric. Automated derivation of behavior vocabularies for autonomous humanoid motion. In *Proc. of the Second Int'l Joint Conference on Autonomous Agents and Multiagent Systems*, Melbourne, Australia, July 2003.
[10] M. H. Law, N. Zhang, A. K. Jain. Nonlinear Manifold Learning for Data Stream. In *Proc. of SIAM Data Mining*, pp. 33-44, Florida, 2004.
[11] S. Z. Li, L. Zhu, Z.Q. Zhang, A. Blake, H. J. Zhang, and H. Shum. Statistical Learning of Multi-View Face Detection. *In Proc. of the 7th European Conference on Computer Vision*. 2002.
[12] C. Liu, H. Y. Shum. Kullback-Leibler Boosting. *Proceedings of the 2003 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'03)*. 2003.
[13] C. J. Liu. A Bayesian Discriminating Features Method for Face Detection, *IEEE Trans. Pattern Anal. and Machine Intel.*, pp. 725-740. 2003.
[14] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 130–136. 1997,
[15] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proc. 6th Int. Conf. Computer Vision*, pp.555–562. 1998,
[16] K. Pettis, T. Bailey, A. K. Jain, and R. Dubes. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Trans. of Pattern Analysis and Machine Intel.*, pp.25-36, 1979.
[17] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: pp.2323-2326, 2000.
[18] S. T. Roweis, L. K. Saul, and G. E. Hinton. Global coordination of local linear models. In *Advances in Neural Information Processing Systems 14*, pp. 889-896. MIT Press, 2002.
[19] H. A. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection. *IEEE Tr. Pattern Analysis and Machine Intel.* pp. 23-38. 1998.
[20] H. A. Rowley, S. Baluja, and T. Kanade. Rotation Invariant Neural Network-Based Face Detection. *Conf. Computer Vision and Pattern Rec.*, pp. 38-44. 1998.
[21] H. Schneiderman and T. Kanade. A Statistical Method for 3D Object Detection Applied to Faces. *Comp. Vision and Pattern Recog.*, pp. 746-751. 2000.
[22] K. K. Sung, and T. Poggio. Example-Based Learning for View-Based Human Face Detection. *IEEE Trans. on PAM*. pp. 39-51. 1998.
[23] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. *Conf. Comp. Vision and Pattern Recog.*, pp. 511-518. 2001.

[24] Y. W. Teh and S. T. Roweis. Automatic alignment of local representations. In *Advances in Neural Information Processing Systems 15*, pp. 841-848. MIT Press, 2003.
[25] B. J. Tenenbaum, V. Silva, and J. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, Volume 290, pp.2319-2323, 2000
[26] J. J. Verbeek, N. Vlassis, and B. Krose. Coordinating principal component analyzers. In *Proc. of International Conf. on Artificial Neural Networks*, pp. 914-919, Spain, 2002.
[27] J.J. Verbeek, N. Vlassis, and B. Krose. Fast nonlinear dimensionality reduction with topology preserving networks. In *Proc. 10th European Symposium on Artificial Neural Networks*, pp.193-198, 2002.
[28] R. Xiao, M. J. Li, H. J. Zhang. Robust Multipose Face Detection in Images, *IEEE Trans on Circuits and Systems for Video Technology*, Vol.14, No.1 pp. 31-41. 2004,
[29] M.-H. Yang. Face recognition using extended ISOMAP. In *International Conf. on Image Processing*, pp.117-120, 2002.
[30] M. H. Yang, D. Roth, and N. Ahuja. A SNoW-Based Face Detector. *Advances in Neural Information Processing Systems 12*, MIT Press, pp. 855-861. 2000.
[31] M. H. Yang, D. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. *IEEE Tr. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58. 2002.
[32] H. Zha and Z. Zhang. Isometric embedding and continuum ISOMAP. In *International Conference on Machine Learning*, 2003.
http://www.hpl.hp.com/conferences/icml2003/ papers / 8.pdf.
[33] http://www.ai.mit.edu/projects/cbcl/software- dataset / index.html.