From: R.P. Würtz and M. Lappe (eds.), Dynamic Perception, pp. 265–270. infix Verlag/IOS press, 2002.

Learning the Detection of Faces in Natural Images

Alexander Heinrichs[1], Christian Eckes[3], Rolf P. Würtz[1] & Christoph von der Malsburg[1,2]

[1] Institut für Neuroinformatik, Ruhr-Universität Bochum
Universitätsstr. 150, D-44780 Bochum, Germany

[2] Lab for Computational and Biological Vision, University of Southern California
Los Angeles (CA), USA

[3] Fraunhofer Institute for Media Communication,
Schloss Birlinghoven, D-53754 Sankt Augustin, Germany

Abstract. We present a two-stage face-finding system as a combination of labeled graph matching and statistical learning. The data format for both stages consists of vectors of the responses of Gabor wavelet filters. Graph matching is used to detect possible locations of faces that we call hypotheses. These typically contain many false positives. The graphs at the found locations are then reinterpreted as vectors, which can be used as input for different statistical learning methods. The methods used here are K-Nearest-Neighbour and the Support Vector Machine with the latter being more efficient.

1 Introduction

Face recognition systems based on elastic graph matching with Gabor wavelet preprocessing have shown outstanding competitive performance (see [PMRR00]) but are less suitable for detecting faces in complex scenes in the first place. Statistical investigations of patterns of faces and face-like objects are needed to overcome such limitations. Examples of successful systems following this approach include a trained multi layer perceptron performing face detection with feature vectors directly derived from grey-level images[RBK98]. Since Gabor Wavelet preprocessing and Support Vector Machines show outstanding performance in computer vision and statistical learning, it makes sense to combine both approaches in designing algorithms for face validation.

The system we propose here, is based on the system developed by Loos and Wieghardt [WL01], which we call face-finder in the following. It detects a chosen number of faces from any given image by using a simplified version of *bunch graph matching* (see [WFKvdM97] for details) and matching with a skin-color template. It delivers so called *hypotheses* of face-positions. The main idea of the new system is to combine the old version with an additional validation stage based on statistical learning to check the output of the face-finder and to improve the results.

The statistical learning methods used in this work are K-Nearest-Neighbour and linear and non-linear Support Vector Machines for classification. A detailed description of the hypothesis generation and the underlying bunch graph matching algorithm is available at [WL01]. The images used for this work are from varying sources (see [LJL98], [SH94], [TP91], [RBK98] and [BHK97] for details). Thanks to the colleagues for allowing us to use them for scientific purposes.

2 Classification of the Hypotheses

2.1 Preparation for the Classification

When the hypotheses are found, a graph is laid on every found region and the Gabor responses at the nodes are stored. The graph has the same structure as the bunch graph but just one jet stored at every node.

The samples used in statistical classifiers are vectors. Therefore, the graph structure of a hypothesis is flattened to a 640-dimensional vector. (There are 40 values in each jet and 16 nodes in one graph.) The vectors are normalized in three different ways to test the influence of the normalization on the classification. In the first dataset the vectors are normalized as a whole, in the second the single jets are first normalized and then written to a vector and in the third the vectors remain unnormalized. Normalization aims at making the representation invariant against contrast changes.

Now the hypotheses are labeled manually. The classes of faces, non-faces and unsure elements are established. The *unsure*-class contains faces which are poorly located by the face-finder due to, e.g., depth rotations not covered by the purely frontal bunch graph. Depending on the experiment, this unsure-class is treated as face or non-face or left out completely to study the influence of these hard-to-learn examples. Thus, the classification task remains a two-class-classification. In the last step of the preparation, the dataset is reordered randomly and every other value is part of the training- or test-set, respectively. Altogether a set of 3115 hypotheses was used, which was build by 855 faces, 392 unsure cases and 1868 non-faces.

2.2 Classification methods

The classification methods used in this work are K-Nearest-Neighbour, linear and non-linear Support Vector Machines for classification. As a relatively simple method, KNN does not offer optimal performance for classification, but can give an impression of the quality of the dataset and the possible performance of better classification-methods. In this work KNN has been applied with the Leave-One-Out-method, i.e., all data except one is used as training-set and the one vector is classified.

The concept of the Support Vector Machine was first introduced by Vapnik in [Vap95] and has been used in many applications with remarkable success. The method can be used for classification and regression of data but in this work, just classification is used. In [Sch97] a good introduction to the theory of Support Vector Machines is given. There are several algorithms and implementations of the SVM-method available. In this work the implementation SVM-Light by Joachims (see [Joa99] for details) is used.

k	no Unsure			Unsure as Non-faces	Unsure as Faces
	Normalized Jets	Normalized Vectors	no Normalization		
1	12.67	14.22	15.70	12.45	13.55
3	12.52	13.07	15.83	11.60	12.74
5	11.86	12.04	15.47	11.53	12.41
7	11.90	12.26	14.80	11.16	12.12
9	11.75	11.72	14.96	11.64	12.38
11	12.05	12.17	15.12	11.75	12.56
21	12.63	11.88	16.08	12.93	14.03
31	13.04	11.65	16.78	14.36	15.24
51	13.37	11.69	17.56	15.39	16.45
101	14.73	12.10	18.39	16.82	18.69

Table 1. Error-rates of the tests with the KNN-Method.

3 Results

In this paper, only few of the computed results can be presented. For a detailed view on the results, please have a look at [Hei01] (in German).

3.1 Tests with KNN

Table 1 shows the error-rate for classification with varying values of k, the number of neighbours considered. The variation of the normalization type is shown without the unsure-cases being used. One can see that for small values of k the type of normalization makes just a small difference. With a bigger k the error-rate increases for all normalization types. This effect was the same for all three methods of treatment of the unsure class, but smallest without the unsure-class. Tests were done with even bigger values of k. But as the size of the dataset has the same magnitude as k, these results depend on the chosen dataset.

Table 1 also shows the error-rates for classification with the unsure-class treated in the three different ways. Notable is the difference between the error-rate *Unsure as faces* and the other error-rates. Looking closer on the results, one can see that especially the classification of the negative examples is worse in this case. A reason could be that for the correct classification of non-faces more data is required than for the classification of faces. Most of the faces are in a relatively small area of the feature space while the non-faces are scattered around everywhere.

3.2 Tests with SVM

The following results from the experiments are just a small fraction of all results and are chosen for several reasons. They all are created with the unsure elements counted as faces and normalized jets. The normalization of jets is chosen because the tests with KNN showed that this preprocessing is more stable than the other

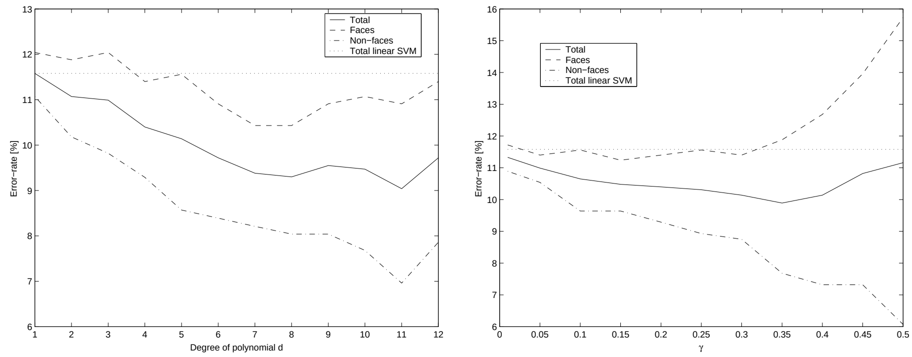

Fig. 1. Error rate vs. degree of polynomial (left) and factor γ (right)

two versions. The unsure elements are counted to the faces-class, because this fits better to human perception (see figure 2 for examples). As mentioned before, the unsure elements tend to be faces, which are not found exactly. Human observers would have no problem in classifying them as faces and so the system should learn to classify them in this way, too.

In general the performance of SVM is better than KNN regardless of the chosen parameters and kernel function. On the left, figure 1 shows the error-rate for the polynomial SVM and the linear SVM. The exact value of the error-rate for the linear SVM is 11.6%. The use of the polynomial kernel is justified, because the error-rate can be decreased to 9.0%. The best value of the degree of the polynomial is 11, which shows, that a relatively complex separation plane divides the data better. Especially the error-rate of the non-faces decreases up to this value. When the degree gets still higher, the performance decreases because of over-fitting of the separating hyper-plane.

Figure 1 also shows the classification performance for the use of a Gaussian kernel on the right. The parameter γ, which is varied here, controls the complexity of the separating plane – a high value of γ leads to a more complex plane and more support vectors. With decreasing γ the performance gets better especially for the negative examples. The performance for the faces stays the same or gets even worse. The faces lie in a relatively small area of the space and can be classified properly with few support vectors. The scattered non-faces can be classified better with more support vectors. But if the number of non-face support vectors increases, the probability of a positive example to be misclassified rises as well.

The use of a polynomial function delivers the best performance for this problem. Additional tests have shown that the performance of the system with Gaussian kernel is far more dependent on the normalization of the vectors. It is difficult to adjust the parameter γ to get a good performance for the unnormalized vectors, since these are far more scattered in the space.

Fig. 2. Examples of hypotheses, which became support vectors with the linear SVM.

4 Discussion

The classification of the hypotheses of the face-finder into faces and non-faces with the Support Vector Machine delivers satisfactory results. The delivered negative hypotheses are very similar to faces in the Gabor responses and so the shown performance is remarkably good.

Certainly the shown method can be developed further in different ways. On one hand, the performance of the hypothesis-finder might be increased by several steps. The concept of the bunch graphs might be developed further to make learning possible, while the system is working. At the current state a lot of human knowledge has to be put into the system to make it work and it is desirable to let the system start with little knowledge delivered by a human and learn independently how to increase the performance.

The performance of the SVM will increase with the amount of data available for training. Since manual labeling of these amounts of data (>10000 images) is awkward, ways have to be found for confident automatic or semi-automatic labeling. Like for the face-finder, it might be possible to start with knowledge provided by humans. This knowledge has to be increased automatically by the system to increase the performance.

As Schoelkopf has shown in [Sch97] it is possible to increase the performance on new data by retraining a SVM with just the previously misclassified data and the old support vectors. So it is possible to store only the misclassified data and the support vectors and still keep all necessary information.

For some additional insight into the form of the face class the support vectors after training have been inspected (see figure 2 for some examples). Notably many of the support vectors of the *non-face*-class still look pretty much like faces. This may indicate that the class boundaries are not very sharp or that the manual labeling has been rather conservative.

5 Acknowledgments

Funding by the German Federal Minister for Science and Education under the project LOKI (01 IN 504 E 9), the European Commission in the Research and Training Network MUHCI (HPRN-CT-2000-00111), and through the "Körber prize for European Science" awarded to C. von der Malsburg in 2000 is gratefully acknowledged.

References

[BHK97] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[Hei01] Alexander Heinrichs. Einsatz statistischer Lernmethoden zum Finden von Gesichtern in natürlichen Bildern. Internal Report IRINI 2001-07, Institut für Neuroinformatik, Ruhr-Universität Bochum, 2001.

[Joa99] T. Joachims. Making large-scale svm learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT-Press, 1999.

[LJL98] A. Loui, C. Judice, and S. Liu. An image database for benchmarking of automatic face detection and recognition algorithms. In *Proc. IEEE International Conference on Image Processing*, 1998.

[PMRR00] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 22(10):1090–1104, 2000.

[RBK98] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, January 1998.

[Sch97] B. Schölkopf. *Support Vector Learning, Dissertation*. R. Oldenbourg Verlag, Munich, Germany, 1997.

[SH94] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, 1994.

[TP91] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

[Vap95] V. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 1995.

[WFKvdM97] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[WL01] Jan Wieghardt and Hartmut S. Loos. Finding faces in cluttered still images with few examples. In Georg Dorffner, Horst Bischof, and Kurt Hornik, editors, *Artificial Neural Networks – ICANN 2001*, volume 2130 of *Lecture Notes in Computer Science*, pages 1026–1033, Vienna, Austria, 2001. Springer Verlag.