Two Kinds of Statistics for Better Face Recognition

Manuel Günther, Marco K. Müller and Rolf P. Würtz

Institut für Neuroinformatik, Ruhr-Universität, 44780 Bochum, Germany

{manuel.guenther,marco.mueller,rolf.wuertz}@ini.rub.de

Abstract. We briefly review the base techniques of elastic graph matching [1] and elastic bunch graph matching [2], which provide a method for face detection, matching, comparison, and identity decision. We then present a method that combines the advantages of Gabor-labeled graphs with maximum likelihood decision making. The improvements over pure bunch graph matching have been studied, and the method has been successfully applied to large face databases like the one for the Face Recognition Grand Challenge [3]. Finally, we describe a system based on rank order statistics that can learn invariances in a minimally supervised way from a set of examples of individual faces in several *situations* like different head pose or illumination. Recognition rates are improved significantly without explicit modeling of the image formation process [4].

Keywords: rank order coding, face recognition, pose invariance, illumination invariance, learning from examples
PACS: 89.20.Ff, 87.63.lj

INTRODUCTION

To be useful for practical applications face recognition systems must solve a variety of tasks. *Detection* procedures must select single probe faces from images, excluding the background. *Matching* or *alignment* estimates pairs of corresponding points in the probe image to be analyzed and known model faces, *comparison* calculates a similarity between probe image and models, and a *statistical decision* about the identity is made on the basis of such similarities.

ELASTIC BUNCH GRAPH MATCHING

Recognition by graph matching [1, 2] compares a given *probe* image P with *gallery images* G_g of all known persons. It first estimates the correspondences between image points on the basis of N local features $\vec{J}_n$ (Gabor jets of K dimensions) in a process called landmark finding. Then, it calculates a similarity between persons by adding (or averaging) local similarities $S_J(P, G_g, n)$ of *corresponding* features. The local similarity function is usually different from the one used for landmark finding. The recognized person is then the G_g with maximal similarity.

Bunch graphs consist of the pre-matched graphs of many persons. They must be initialized by manual annotation of facial images. In [5] we have presented a method that can automatically build good bunch graphs from only a few manually annotated images. This is also used for invariance learning in the second part of this paper.

REPRESENTATION OF MULTIPLE FACES

The multiple faces in a bunch graph can also be represented by the distribution of local features instead of the whole bunch. We are applying the maximum-likelihood (ML) classifier from [6] to the jets in a bunch graph. We estimate the ML probability without calculating a PCA, assuming independence of the components J_i. Covariances of J_i are small because the frequency sampling of the Gabor filters is chosen such that they are nearly orthogonal. The method uses ML estimates of the mean μ_i and variance σ_i^2 for each dimension $i = 1, \ldots, N$ of $\vec{J}$. Therefore, this algorithm can deal with fewer training images than dimensions ($P < N$) since the correlations between components J_i and J_j need not be estimated.

The ML probability can be either used for face detection in the one-class-fashion

$$S_{ML}\left(\vec{J}\right) = \sum_{n=1}^{N}\sum_{k=1}^{K} -\frac{\left(J_{nk}-\mu_{nk}\right)^2}{\sigma_{nk}^2} \tag{1}$$

CP1281, *ICNAAM, Numerical Analysis and Applied Mathematics, International Conference 2010*
edited by T. E. Simos, G. Psihoyios, and Ch. Tsitouras
© 2010 American Institute of Physics 978-0-7354-0831-9 /10/$30.00

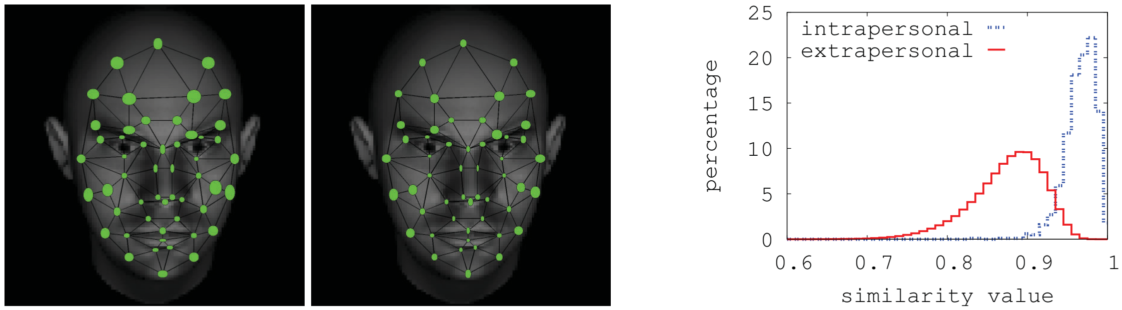


FIGURE 1. The average node positioning errors for Elastic Bunch Graph Matching (left) are much larger than for the maximum likelihood estimators (center), making the latter a much better landmark detector. The right graph shows the intrapersonal and extrapersonal distributions for artificial faces. The overlap of the distribution leads to recognition errors.

or on local image similarities s_n (n being the graph node index) to classify image differences:

$$S_{BIC}(\vec{s}) = \sum_{n=1}^{N} \left[-\frac{(s_n - \mu_{I;n})^2}{\sigma_{I;n}^2} + \frac{(s_n - \mu_{E;n})^2}{\sigma_{E;n}^2} \right]. \tag{2}$$

In the training stage means μ and variances σ are evaluated statistically on the basis of a training set $\{\vec{J}^{(p)} | p = 1, \ldots, P\}$ by assuming a Gaussian distribution. For the S_{BIC} classification of image differences the parameters of the two classes Ω_I and Ω_E have to be estimated. For this purpose two distinct sets $\{\vec{s}_I^{(p)} | p = 1, \ldots, P_I\}$ and $\{\vec{s}_E^{(p)} | p = 1, \ldots, P_E\}$ are required. Each vector $\vec{s}_I^{(p)}$ of the first set is generated as the result of comparing two images of the same person, the second set contains vectors $\vec{s}_E^{(p)}$ created from different persons. Since no prior probabilities are needed, the cardinalities P_I and P_E of the training sets can differ without any problems.

Eqs. (1) and (2) can be computed very fast because no transformation mixing input data from different images is required. Thus, time complexity $O(N)$ is reached, in comparison to $O(N \cdot M)$ when using PCA. Another nice fact is that we have not lost the benefit of the Mahalanobis distance measure, the resulting similarity value still relies on dimensionless data. Beyond that, S_{BIC} has the ability to detect outliers if both intrapersonal and extrapersonal distances are high.

We have applied the method to artificially generated face images in order to measure precisely the landmark positioning errors and to estimate the error rates to be expected (fig. 1). Note that the artificial faces are very similar because hairstyle and skin texture information is missing. For tests on large real-world databases, see [3].

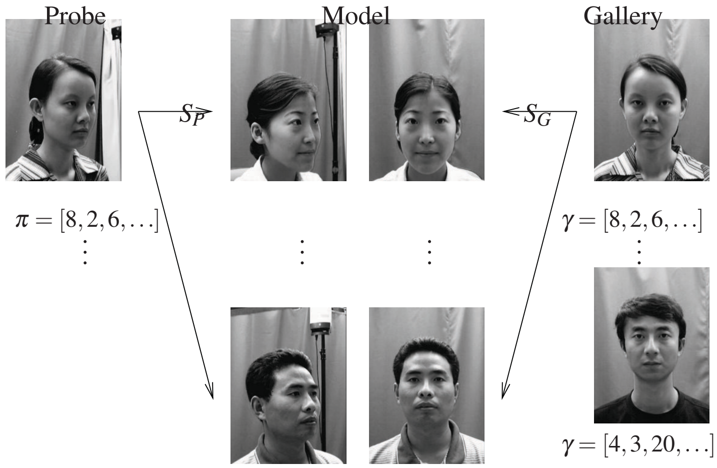
LEARNING OF INVARIANCES

Invariances for which explicit modeling is difficult, like large pose differences or illumination changes, can be handled by elastic bunch graph matching only if bunch graphs are supplied for a coarsely sampled set of variants, e.g., 10 different head poses. This is problematic from a technical point of view [7] because for a large recognition system it is infeasible to store and match all persons in all possible modalities.

From a machine learning point of view, the requirement to recognize identity independent of situation is a case of generalization. However, invariance under even a simple visual transformation such as translation in the image plane is not a generalization performed naturally by known learning mechanisms. Therefore, methods to control the generalization on the basis of examples are required.

The similarity we introduce here is a special case of *rank correlation*, one example being Spearman's rank order correlation coefficient. This sort of statistics has been used for the evaluation of biometric systems [8] and for face matching in [9]. Here, we apply it to guide generalization in a desired direction by the presentation of examples.

For the recognition of an arbitrary subject a large *gallery* database is created, which contains all known subjects in a preferred situation $v = 0$. Throughout this paper, this situation will be a frontal pose under frontal illumination. This has also been shown to be optimal in [10]. Images of the same person in different situations build a *model database*, which serves as examples for the transformation between situations (see fig. 2).



	Pose	**Illum.**
RR with given situation	99.02	89.01
Rate of correct situation estimation	99.9	92.0
RR with automatically determined situation	97.8	90.0
Best RR reported in [11]	71	51

FIGURE 2. Situation-independent recognition is mediated by a model database of some persons in all situations. Probe and gallery images are coded into rank lists π and γ by their similarities to the models. These rank lists are comparable, while the similarities are not (feature indices have been dropped for clarity). The table on the right shows the recognition results on the CAS-PEAL database.

Personal identity is coded by a similarity rank list to the models of the same situation. The rank list for a probe subject P is created as follows. First, all local similarities S^v to all model images M_m^v are calculated. For each index n and situation v a rank list r_n^v is created, which contains the rank of similarity for each model index m, so that for each pair of model images $M_m^v, M_{m'}^v$ the following holds ($r_n^v(m) \in \mathbf{N}_0$):

$$r_n^v(m) < r_n^v(m') \quad \Rightarrow \quad S_{\mathrm{loc}}(P, M_m^v, n) \geq S_{\mathrm{loc}}(T, M_{m'}^v, n)\,. \tag{3}$$

The most similar model candidate would be the one with $r_n^v(m) = 0$, the follower-up the one with $r_n^v(m) = 1$, etc. These lists now serve as a representation of a test image P. For varying P we will use the notation $r_n^v(P,m)$.

Each subject G_g in the gallery is assigned a rank list representation $\gamma_{g,n}$ by matching each of its landmarks to those of the model subjects in the preferred situation:

$$\gamma_{g,n}(m) = r_n^0(G_g, m)\,,\ m = 1\ldots N_M\,. \tag{4}$$

For recognition we first assume that a *probe* image P^v appears in the *known* situation v. The requirement to know the situation will be dropped below. This probe is also represented as a similarity rank list π_n^v for each landmark of all models in situation v:

$$\pi_n^v(m) = r_n^v(P^v, m)\,,\ m = 1\ldots N_M\,. \tag{5}$$

$$S_{\mathrm{rank}}(\pi, \gamma_g) = \frac{1}{N_M} \sum_{m=1}^{N_M} \lambda^{\pi(m)+\gamma_g(m)}\,. \tag{6}$$

So far, the feature index n has been omitted from the rank list derivations. Local features in the face graphs are numbered such that corresponding landmarks have the same value of n, across all situations. Graphs in different situations have only subsets of these landmarks as nodes, the set of node indices for each situation is denoted by $\mathscr{L}^v$.

In a realistic setting, the situation of the probe image P is, of course, unknown. It can be estimated by matching to all model images of all situations, and assigning the situation with the highest similarity.

The network was tested on the CAS-PEAL face database [12]. The landmarks are found by elastic bunch graph matching, starting from very few images, that were labeled by hand. 24 subjects have been set aside for manual labeling. From these, the basic bunch graphs have been built (12 for pose, 8 for illumination).

The remaining 1015 subjects have been split up into model sets and testing sets (500 model and 515 testing for the pose case, and 100 model and 91 testing for illumination). The table in figure 2 shows the results. The remarkable difference in recognition rate to the literature values shows that the variation due to pose and illumination changes have been successfully learned.

DISCUSSION

We have presented a face recognition system which estimates the likelihood of a facial image to belong to a personal identity. The method is not only applicable to face detection and identity decision, but can be applied to other classes as well. Recently, we have applied the maximum-likelihood classifier to the genetic syndromes studied in [13, 14, 15]. We could achieve comparable results as in [15], but with a much simpler classifier.

In the second part, we have tackled the problem of learning complicated visual invariances from example images and generalizing to a much larger group of people. The results are especially promising in the presence of illumination differences, which is a very hard problem even for human observers[16].

ACKNOWLEDGMENTS

RPW would like to thank Prof. Deuflhard for his fabulous lectures on Numerical Analysis and wish him all the best for his birthday. We gratefully acknowledge funding from the German Research Foundation (WU 314/2-2 and WU 314/5-2). Portions of the research in this paper use the CAS-PEAL face database collected under the sponsorship of the Chinese National Hi-Tech Program and IS VISION Tech. Co. Ltd. [12, 11].

REFERENCES

1. M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen, *IEEE Transactions on Computers* **42**, 300–311 (1993).
2. L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, 775–779 (1997).
3. M. Günther, and R. P. Würtz, *International Journal of Pattern Recognition and Artificial Intelligence* **23**, 433–461 (2009).
4. M. K. Müller, and R. P. Würtz, "Learning from Examples to Generalize over Pose and Illumination," in *Artificial Neural Networks – ICANN 2009*, edited by C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Springer, 2009, vol. 5769 of *LNCS*, pp. 643–652, ISBN 978-3-642-04276-8.
5. A. Heinrichs, M. K. Müller, A. H. Tewes, and R. P. Würtz, "Graphs with Principal Components of Gabor Wavelet Features for Improved Face Recognition," in *Information Optics: 5th International Workshop on Information Optics; WIO'06*, edited by G. Cristóbal, B. Javidi, and S. Vallmitjana, American Institute of Physics, 2006, pp. 243–252, ISBN 978-0-7354-0356-7, URL `http://proceedings.aip.org/proceedings/confproceed/860.jsp`.
6. B. Moghaddam, W. Wahid, and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 30–35.
7. E. Murphy-Chutorian, and M. M. Trivedi, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**, 607–626 (2009).
8. A. L. Rukhin, and A. Osmoukhina, *Journal of Statistical Planning and Inference* **131**, 1–18 (2005).
9. O. Ayinde, and Y.-H. Yang, *Pattern Recognition* **35**, 1275–1289 (2002).
10. M. K. Müller, A. Heinrichs, A. H. Tewes, A. Schäfer, and R. P. Würtz, "Similarity rank correlation for face recognition under unenrolled pose," in *Advances in Biometrics*, edited by S.-W. Lee, and S. Z. Li, LNCS, Springer, 2007, pp. 67–76, ISBN 978-3540745488.
11. W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, *IEEE Transactions on Systems, Man, and Cybernetics Part A* **38**, 149–161 (2008).
12. W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, and D. Zhao, The CAS-PEAL large-scale Chinese face database and baseline evaluations, Tech. Rep. JDL-TR-04-FR-001, Joint Research & Development Laboratory for Face Recognition, Chinese Academy of Sciences (2004).
13. H. S. Loos, D. Wieczorek, R. P. Würtz, C. von der Malsburg, and B. Horsthemke, *European Journal of Human Genetics* **11**, 555–560 (2003).
14. S. Böhringer, T. Vollmar, C. Tasse, R. Würtz, G. Gillessen-Kaesbach, B. Horsthemke, and D. Wieczorek, *European Journal of Human Genetics* **14**, 1082–1089 (2006), ISSN 1018-4813.
15. T. Vollmar, B. Maus, R. P. Würtz, G. Gillessen-Kaesbach, B. Horsthemke, D. Wieczorek, and S. Böhringer, *European Journal of Medical Genetics* **51**, 44–53 (2008).
16. P. Phillips, T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, *IEEE Trans. PAMI* **32**, 831–846 (2010).