Comparing Facial Line Drawings with Gray-Level Images: A Case Study on PHANTOMAS

Wolfgang Konen

ZN GmbH, Bochum,
Germany
WWW: http://www.zn.ruhr-uni-bochum.de

Abstract. We report on an application of neural face recognition algorithms to a task with relevance to forensic investigations: The software tool PHANTOMAS (**phantom** **a**utomatic **s**earch) allows to compare facial line drawings (the German "Phantomzeichnung") with gray-level images of faces. In addition to normal (textual) database search actions, this software tool allows picture-to-picture searches. We present first results on the evaluation of a benchmark on this task. The ranking quality of PHANTOMAS allows to spot the true match belonging to a certain drawing on the average within the upper 2.7% of the database ($N = 103$). It is shown that this is comparable to the human performance on the same data material. Computation times make it feasible to search online in large databases ($N \approx 10000$). – With the same algorithm it is also possible to classify complex characteristica in faces (gender classification of line drawings better than 90%).

1 Introduction

Recognition of faces is a remarkable example for the ability of humans to perform complex visual recognition tasks. Even more striking is the fact that we can reliably identify persons from simple line drawings which have, on the gray-level basis, little in common with the original picture.

To build automated systems with similar capabilities is not only of research interest, but also has an important application in forensic investigations: There one often has the line drawing or a sketch as starting point, but the actual identification task requires that the witness should look through thousands of images – often too much to complete the search successfully. Needed is a tool for automated sorting the database into similar and definitely unsimilar faces, allowing the witness to focus his/her attention on the important cases.

Many algorithms for face recognition have been described in the literature, see e.g. [1, 2, 3]. Most successful algorithms use principles from neural information processing, for example neural filters (e. g. Gabor wavelets) for the early visual processing or neural-net based learning of higher-level invariants [4].

Our algorithm is based on von der Malsburg's Elastic Graph Matching Algorithm [5] and has been successfully applied for the access control system ZN-Face [6] where it has shown to provide robust and secure verification behavior.

1.1 Face recognition by graph matching

The basic face recognition algorithm is described in [6] and is an extension of the Elastic Graph Matching Algorithm [5]; the reader is referred to these articles for a comprehensive description. Here we just mention that faces are stored as flexible graphs or grids (see Fig. 1) with characteristic visual features (Gabor features) attached to the nodes of the graph (*labeled graphs*).

A **B**

Fig. 1. (A) *Labeled graphs are the data representation in* PHANTOMAS. **(B)** *Such graphs can be shifted, scaled and deformed efficiently in the image domain.*

The Gabor features are based on the wavelet transform

$$G_{\mathbf{k},\sigma}(\mathbf{x}) = C_{\mathbf{k},\sigma} \exp(\frac{k^2 x^2}{2\sigma^2}) \exp(i\mathbf{k}\mathbf{x}), \qquad (1)$$

and have shown to provide a robust information coding for object recognition [5] (invariance against intensity or contrast changes in the image). Furthermore, Gabor features are less affected by changes in head posture, size and facial expression than raw grey level features.

The matching of a stored graph against an image consists of an optimization of the graph location in the image ("Global Move", GM). We have shown previously [6], that within a range of $\pm 20\%$ variation in scale and aspect ratio it is possible to combine the GM-displacement with simultaneous optimization in scale and aspect ratio. This works without getting stuck in possible local minima and without transforming the Gabor features. We will refer to this joint optimization procedure as 'GM' in the following. [1]

In summary, the face recognition algorithm consists of five basic steps:

(i) Choose a search template and the desired search fundus (which may be the whole database or certain subparts of it),
(ii) Convolve the image with each Gabor Wavelet used in the labeled graph representation,
(iii) With a representative set of graphs from the database we execute the GM. From the best GM, we keep the graph extracted from the search template as "optimally positioned graph" (OPG).
(iv) Match all graphs in the search fundus against the OPG (60 msec/match).
(v) Rank the results according to matching similarity.

2 Features of the PHANTOMAS system

The PHANTOMAS system hardware consists of a PC and a color scanner as input device for database or search template images. As an option, a color camera with framegrabber may be used for direct digital input.

The software for PHANTOMAS was developped using IDL (Interactive Data Language). This interpreter-type language with many graphic visualization and GUI-capabilities enables platform independence and large flexibility during the development process. The price to pay for this is a somewhat reduced performance in (some) computing intensive task. Therefore, the final product version will have the search kernel routines implemented in C++.

[1] The whole optimization procedure and its application to verification or database searches has been patented [7].

	$\mathcal{O}(1)$	$\mathcal{O}(N)$	Total
N=100	31 sec	6 sec	37 sec
N=10 000 (extrapol.)	31 sec	600 sec	10 min 51

Table 1. Performance times for the PHANTOMAS search as a function of the persons N in the search fundus

Computational cost: Concerning the number N of persons in the search fundus, there is only one step in the algorithm, namely step (iv), which is linearly dependent on N (complexity $\mathcal{O}(N)$), while the other steps are of complexity $\mathcal{O}(1)$.

The performance times for the unoptimized IDL-code on a Sun Sparc 10/40 are given in Tab. 1. From this we can extrapolate the performance for a ($N = 10000$)-search task to be less than 11 min. Using a Pentium P130 and optimized search kernel routines we expect a speedup of at least a factor of 2.

The PHANTOMAS software is equipped with a comfortable graphical user interface (GUI) where nearly all actions can be invoked by simple mouse clicks. Among the features of the GUI are the following topics:

- The results of search projects can be presented in many different ways, e.g. (a) with different sortings (rank of similarity, name, randomly), (b) combined with additional matching constraints, (c) with and without descriptive data.
- Interactive recursive search: Given the performance times in the order of 10 min(Tab. 1), it is feasible to perform multiple searchs in one session. Example: The result of the first search project may bring up an image which even better matches the witness' memory. In that case a second search can be started with this new image as search template and so on.
- Additional textual search criteria may be defined (e.g. age, male/female, bearded/nonbearded, ...) to constrain the search fundus and speed up the search process.
- Zooming-in on pictures of special interest.

3 Evaluation

3.1 The benchmark

The benchmark undertaken with the system consisted of a rather tough task, namely the compa-

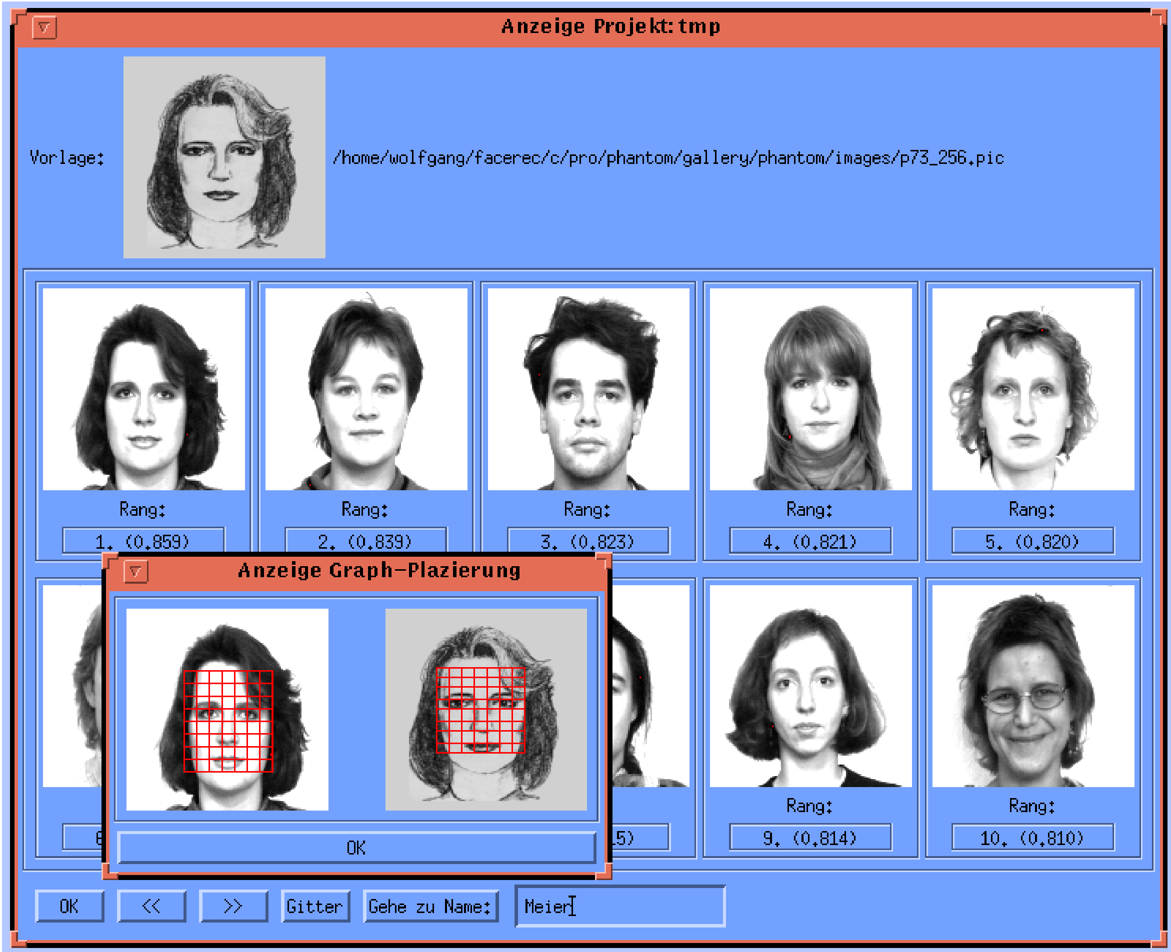

Fig. 2. The result of a PHANTOMAS search on a line drawing (upper left) where the true match is found at place 1 of the ranking list. The small inset shows how the graphs are positioned on the faces in order to account for the different size and aspect ratio.

rision of facial line drawings with original photos. The database of original photos is a slight extension of [5] (103 persons, 33 females, 70 males). The pictures were taken in front of an unstructured background and with mostly frontal illumination. However, some pictures had considerable side illumination (see e.g. picture # 1 in Fig. 3). From 13 of the database pictures, a professional forensic artist produced the "phantom drawings". In these drawings, slight deviations from the original (e. g. moustache, different hair-style) can occur. Some examples are shown in Fig. 3. In most cases the aspect ratios in the drawings and the photos were somewhat different (see inset in Fig. 2).

3.2 Automated similarity ranking

Fig. 2 shows an example result for the top-ranking persons as found by PHANTOMAS. Although the gender of the line drawing was not specified as search constraint, PHANTOMAS has picked up dominantly females in the first places. The summary of the ranking results is given in Tab. 2.

As a comparision we show in Tab. 2 also the result of a "naive" algorithm where one tries to rank the line drawings based on the normalized image cross correlation (NC1). Clearly, this approach does not lead to satisfactory results. Even if we use partial information from the graph matching algorithm and warp the input image according to scale and aspect ratio as found by PHANTOMAS before calculating the normalized correlation coefficient, the results improve only

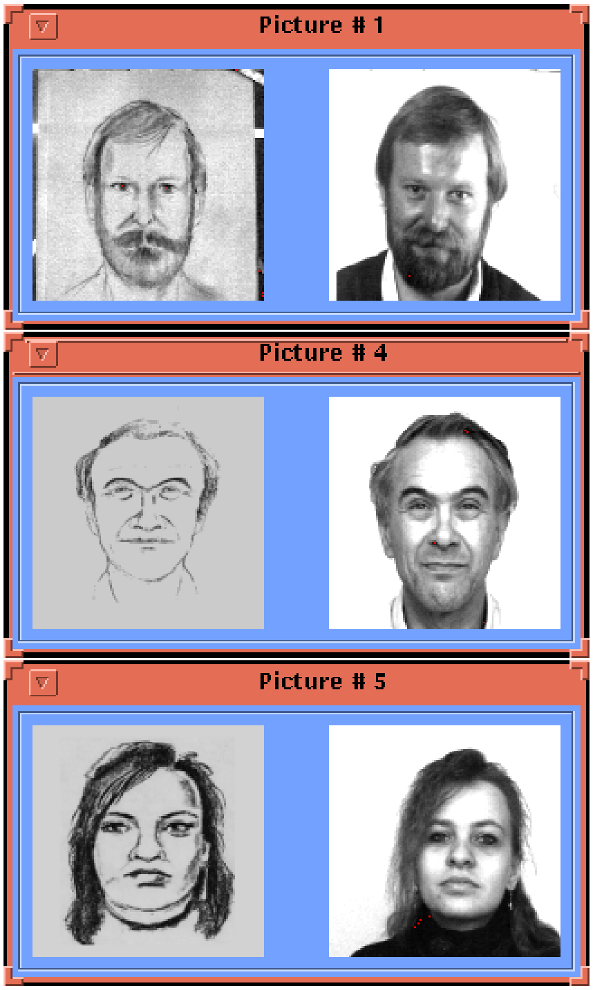

Fig. 3. Three examples of the facial line drawings together with the original image

picture	gender	PH	NC1	NC2	subjects				
					A	B	C	D	E
# 1	m	6	75	50	2	n/f	2	n/f	2
# 2	m	3	100	65	1	1	1	10	n/f
# 3	m	2	3	1	1	1	1	n/f	2
# 4	m	1	7	29	1	1	1	5	1
# 5	f	1	1	14	2	1	1	2	1
# 6	m	1	2	1	1	1	1	3	1
# 7	m	1	6	2	1	1	1	1	1
# 8	m	1	18	4	1	1	1	n/f	1
# 9	m	1	3	1	1	1	1	1	1
# 10	f	13	17	1	n/f	3	n/f	n/f	1
# 11	m	3	25	28	1	1	1	1	1
# 12	m	1	1	2	1	1	1	n/f	1
# 13	f	1	1	1	3	n/f	1	1	1
	Median:	1	4.5	2	1	1	1	4	1
	Mean:	2.7	19.9	14.6	-	-	-	-	-
	Mean w/o # 10:	1.8	20.2	16.4	1.3	-	1.1	-	-
	$\langle\tilde{R}\rangle$:	2.7	19.9	14.6	2.1	3.6	1.8	9.9	2.8

Table 2. Performance evaluation: For all 13 facial line drawings a ranking of the 103 database images was performed. Shown is the place R of the true match in the ranking list (1=best match). **PH**: PHANTOMAS, **NC1**: Normalized cross correlation without warping, **NC2**: Normalized correlation with warped images, **A-E**: five different human subjects (n/f: true match was not found).

slightly (NC2).

3.3 Comparision with human performance

In order to get a rough feeling for the difficulty of the ranking task, we performed with 5 naive human subjects a similar task. This was not meant to be a comprehensive psychophysical experiment, but only to give an indicator on how well humans can spot the right match. The subjects were asked to rank to each line drawing the 10 most similar photos they found in the database. They were allowed to pick less than 10 persons only if they were sure, that no one of the non-considered persons was the one actually portrayed by the forensic artist. [2]

As we see from Tab. 2, none of the subjects was able to find all true matches. There is no clear prescription on how to calculate the mean ranking, if some true matches were not found by the subject. We have chosen the following prescription: Since the gender of the line drawings was always correctly classified, even in the 'n/f'-cases, a hypothetical forced choice in the 'n/f'-cases would produce on the average at least the same ranking as random choice ($p = 1/2$) on the same-gender persons:

$$\tilde{R} = \begin{cases} pN_{same-gender} & \text{if true match n/f} \\ R & \text{else} \end{cases} \qquad (2)$$

We have chosen $p = 1/3$ when calculating the expectation value $\left\langle \tilde{R} \right\rangle$ in Tab. 2.

3.4 Gender Classification

[2] It does not make sense to force the subjects to rank *all* 103 database members, since the results would be in most cases non-reproducible.

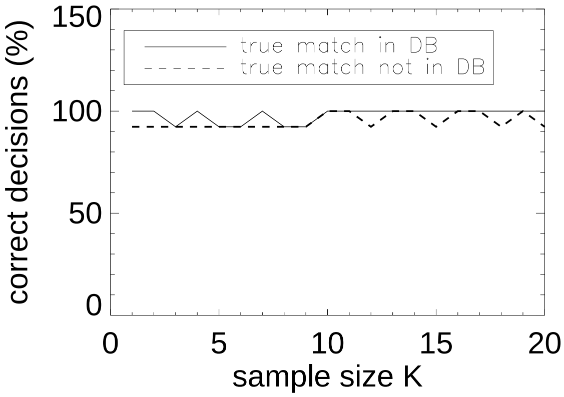

Fig. 4. Gender classification of facial line drawings based on the PHANTOMAS ranking. The “oscillations” between 92% and 100% originate from exactly one of the line drawings, this one being for PHANTOMAS “on the edge” between male and female.

An interesting feature of the presented face recognition algorithm is its ability to pick up complex facial characteristics (e.g. “is of Basedow-type”) which cannot be contributed to a single facial feature alone.

As an example we consider here the male/female distinction. The gender classification has been previously described in the context of Elastic Graph Matching by Wiskott et al. [8] with good results. We show here that with the PHANTOMAS software and a very simple prescription the composite characteristicum “gender” can be reliably identified even for line drawings:

The search template of a certain search project is classified “female”, if the number of females within the first K persons from the ranking list is greater than the random-choice expectation value E_F.

If there are N_F females in a database of N persons in total, then the probability for having r females in a sample of K persons is given by the hypergeometric distribution $p(r, K, N)$. The expectation value is

$$E_F = \sum_{r=0}^{K} r\ p(r, K, N) = K \frac{N_F}{N} \qquad (3)$$

The results in Fig. 4 show that the correct gender decision is more or less independent from the sample size K and works even if the true match is not in the database. The correct-decision rate of better than 90% is comparable to results reported in other works [8, 4].

The same mechanism should also work for other complex characteristica X as long as the database provides representative material for the both cases “belongs to class X” and “does not belong to class X”. This could provide a mechanism for an objective labeling of facial pictures with respect to those characteristica.

4 Discussion

It has been shown that Elastic Graph Matching can successfully recognize faces from facial line drawings. In all cases the true match was within the upper rankings.

For comparision the ranking task was also performed by human subjects. The line drawings difficult for PHANTOMAS (especially # 1 and # 10 in Tab. 2) were also difficult for the human subjects. The overall performance of PHANTOMAS and human subjects seems to be comparable although an “average ranking” is difficult to extract from the subjects' results due to the missing values (each subject had at least one “not found”-case). We developped with $\left\langle \tilde{R} \right\rangle$ a simple model for the average ranking in the case of missing values. Tab. 2 shows that PHANTOMAS' $\left\langle \tilde{R} \right\rangle$ is close to the median of the human subjects' $\left\langle \tilde{R} \right\rangle$.

Why are Gabor wavelets so successful for the recognition of line drawings? We have no clear answer to this question, but one reason might be that dominant features of line drawings are the orientation of bars and step edges, and the Gabor code is also dominated by orientation features.

It has been shown that a similar ranking quality can not be achieved by “naive” approaches (e.g. cross correlation) on the same data material.

The results presented in this case study have of course to be extended to larger databases ($N \approx$ 10000 or larger). We are currently designing a pilot project for a law enforcement institution to evaluate a 10 000-person database. A simple extension of the current performance to larger databases leads to the expectation that the average rating performance should be at the 2.7%-level, i.e. instead of 10 000 only 270 images would need to be examined manually.

With representative data material for complex facial characteristics, further research will go into automated labeling of complex face types. The aim is to provide an objective basis for the the assignment of textual labels to faces which in turn will make the textual description of faces more reliable.

References

1. V. Bruce and M. Burton. *Processing Images of Faces*. ABLEX Publishing-Corporation, Norwood, NJ, 1992.
2. M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of CVPR'91*. IEEE Press, 1991.
3. M. Bichsel, editor. *Proceedings of the International workshop on automatic face- and gesture-recognition (IWAFGR)*, MultiMedia Laboratory, University of Zurich, Switzerland, 1995.
4. R. Brunelli and T. Poggio. Caricatural effects in automated face perception. *Biological Cybernetics*, 69:235–241, 1993.
5. M. Lades, J. Vorbrüggen, J. Buhmann, J. Lange, C.v.d. Malsburg, R.P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transaction on Computers*, 42:300–311, 1993.
6. W. Konen and E. Schulze-Krüger. ZN-Face: A system for access control using automated face recognition. In M. Bichsel, editor, *Int. Workshop on Face and Gesture Recognition*, 1995.
7. W. Konen, J.C. Vorbrüggen, and R.P. Würtz. Deutsches Patent Nr. 44 06 020: Verfahren zur automatisierten Erkennung von Objekten. Erteilt am 29.06.1995.
8. L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition and gender determination. In M. Bichsel, editor, *Int. Workshop on Face and Gesture Recognition*, 1995.