



Image warping for face recognition: From local optimality towards global optimization

Leonid Pishchulin^{a,*}, Tobias Gass^b, Philippe Dreuw^c, Hermann Ney^c

^a MPI for Informatics – Computer Vision and Multimodal Computing Group, Campus E1 4, D-66123 Saarbrücken, Germany

^b ETH Zurich – Computer Vision Lab, Sternwartstrasse 7, CH-8092 Zurich, Switzerland

^c RWTH Aachen – Department 6, Ahornstr. 55, D-52056 Aachen, Germany

ARTICLE INFO

Available online 3 November 2011

Keywords:

Image warping
Face recognition
Energy minimization

ABSTRACT

This paper systematically analyzes the strengths and weaknesses of existing image warping algorithms on the tasks of face recognition. Image warping is used to cope with local and global image variability and in general is an NP-complete problem. Although many approximations have recently been proposed, neither thorough comparison, nor systematic analysis of methods in a common scheme has been done so far. We follow the bottom-up approach and analyze the methods with increasing degree of image structure preserved during optimization. We evaluate the presented warping approaches on four challenging face recognition tasks in highly variable domains. Our findings indicate that preserving maximum dependencies between neighboring pixels by imposing strong geometrical constraints leads to the best recognition results while making optimization efficient.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Automatic face recognition by reasoning about similarity of facial images is a hard task in computer vision. Strong local variations in expressions and illuminations, global changes in pose, temporal changes, partial occlusions, as well as affine transformations stemming from automatic face detection all contribute to rich intra-class variability which is difficult to discriminate from inter-class dissimilarity. In addition, often only a limited number of images per individual are provided as references. In the most extreme case, only one frontal image (mugshot) is available (see Fig. 2), which makes it difficult to learn a model capturing the natural variability. In this work we analyze image warping algorithms which do not build any specific facial models, but directly encode a deformation-invariant dissimilarity measure used within a nearest neighbor classification framework.

Many methods approach the image similarity problem in face recognition by extracting local features from interest points or regular grids and matching them between images. The similarity is based on the quality and number of found matches [1–5]. The main focus is put on finding an appropriate feature descriptor which is a priori invariant to certain transformations [6–9] or can be learned from suitable training data [10–12], but not on the

matching procedure. Descriptors must be chosen or trained to carry as much discriminatory information as possible making these methods prone to overfitting on a certain task. In addition, no geometrical dependencies between matches are considered, which makes these methods fast, but also disregard image structure. In contrast to related feature matching techniques, most of presented warping approaches incorporate geometric dependencies while also relying on dense local descriptors. This leads to smooth structure-preserving deformations compensating for strong appearance changes due to different poses and expressions (cf. Fig. 1) and also makes the methods robust to occlusions and illuminations.

Another group of face recognition methods try to cover global and local variability by using parametric shape models, such as elastic graph bunch matching [13], active shape [14] and active appearance models [15], or by imposing domain knowledge to infer 3D models from 2D images. For the latter, virtual pose images are generated in [16] which can be used as additional reference images and pose variability is learned from data in [17] to marginalize over poses. In contrast to these approaches, the presented warping algorithms are far less task specific as no prior knowledge on face structure is involved in the design of corresponding methods.

Recent work [18–23] has shown increased research interest in *image warping* approaches originating from one-dimensional dynamic time warping in speech recognition. Direct extension to the two-dimensional case leads to intractable models in which global optimization is NP-complete [24] due to the loopy nature of the underlying graphical model. Therefore, approximations have been proposed, which relax some of the *first-order* dependencies between single

* Corresponding author. Tel.: +49 681 932 51 208; fax: +49 681 9325 2099.

E-mail addresses: leonid@mpi-inf.mpg.de (L. Pishchulin), gasst@vision.ee.ethz.ch (T. Gass), dreuw@cs.rwth-aachen.de (P. Dreuw), ney@cs.rwth-aachen.de (H. Ney).

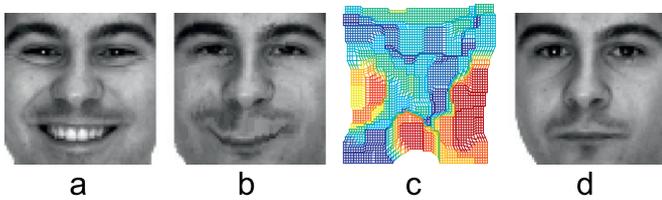


Fig. 1. Query image (a), mugshot reference image (d), and deformed reference image (b) using smooth global warping algorithm. (c) shows the deformation grid, where dark blue areas correspond to small absolute deformations and dark red corresponds to strong absolute deformations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Fig. 2. (a) Local and global face variability caused by changes in facial expressions and partial occlusions (top row), and changes in pose (bottom row). (b) Only one reference image per person is available.

positions of the image grid during optimization. Zero-order warping [25] disregards all dependencies and thus is very efficient. Pseudo-two-dimensional hidden Markov models (P2DHMM) have been used for face recognition [26,27]. They can be calculated efficiently using decoupled hidden Markov models (HMMs), but cannot find good warpings in the presence of strong non-linear variations and cannot even cope with rotations. The idea of P2DHMMs was extended to trees [28] allowing for greater flexibility at a cost of higher computational complexity. Additionally, maximum a posteriori inference (MAP) in Markov random fields (MRF) is receiving increased attention [20,18,19]. Efficient algorithms like sequential tree-reweighted message passing (TRW-S) [29] optimize all dependencies simultaneously and find good (local) optima with a huge number of labels.

Although many warping algorithms have been proposed recently, neither thorough comparison, nor systematic analysis in a common scheme has been done so far. Each new warping method is often seen as an incremental improvement over an existing approach and thus usually compared to one baseline algorithm, whereas its positioning in the global view on image warping is unclear. In this work, we systematically analyze the strengths and weaknesses of existing warping approaches by arranging them into a hierarchy of four groups of methods in a global scheme. We assign each method to a particular group respecting the dependencies which hold between single image coordinates during optimization. In our analysis, we gradually move on from simple locally optimal methods towards more complex approaches which optimize all dependencies simultaneously. We thoroughly evaluate all presented algorithms on four challenging face recognition tasks and show that preserving more dependencies described by strong geometric constraints leads to the best recognition performance.

The rest is organized as follows: We start with the problem of image warping and describe structural constraints in Section 2. Then we analyze advantages and disadvantages of existing warping approaches in Section 3. We describe some practical issues in Section 4 and perform a thorough evaluation of presented approaches in Section 5. Finally, we provide conclusion remarks.

2. Image warping

In this section, we will define the two-dimensional image warping (2DW) analogously to [30] and present structure-preserving constraints.

In 2DW, an alignment of a reference image $R \in F^{U \times V}$ to a test image $X \in F^{I \times J}$ is searched for so that the aligned or warped image $R' \in F^{I \times J}$ becomes similar to X . F is an arbitrary feature descriptor. The alignment is a pixel-to-pixel mapping $\{w_{ij}\} = \{(u_{ij}, v_{ij})\}$ of each position $(i, j) \in I \times J$ in the test to a position $(u, v) \in U \times V$ in the reference image. One is interested in an alignment $\{w_{ij}\}$ maximizing the posterior probability:

$$\{\hat{w}_{ij}\} = \arg \max_{\{w_{ij}\}} p(\{w_{ij}\} | X, R) = \arg \max_{\{w_{ij}\}} p(X, R | \{w_{ij}\}) p(\{w_{ij}\}). \quad (1)$$

Assuming a Gibbs distribution, instead of maximizing Eq. (1), we can also minimize an energy $E(X, R, \{w_{ij}\}) = -\log p(X, R | \{w_{ij}\}) p(\{w_{ij}\})$ which using first-order Markovian assumptions becomes:

$$E(X, R, \{w_{ij}\}) = \sum_{ij} \left[d(X_{ij}, R_{w_{ij}}) + \sum_{n \in \mathcal{N}(ij)} T_{n,ij}(w_n, w_{ij}) \right]. \quad (2)$$

The unary term $d(X_{ij}, R_{w_{ij}})$ is a distance between corresponding pixel descriptors and the pairwise term $T(\cdot)$ is a smoothness function in which first-order geometrical dependencies and constraints between neighboring pixels \mathcal{N} can be implemented. The optimal alignment w_{ij} is obtained through minimization of the energy $E(X, R, \{w_{ij}\})$. This optimal alignment does not change when the smoothness is only computed w.r.t. horizontal and vertical predecessors, changing Eq. (2) to

$$E(X, R, \{w_{ij}\}) = \sum_{ij} [d(X_{ij}, R_{w_{ij}}) + T_h(w_{i-1,j}, w_{ij}) + T_v(w_{i,j-1}, w_{ij})], \quad (3)$$

where $T_h(\cdot)$ and $T_v(\cdot)$ are horizontal and vertical smoothness terms. Finding the global minimum of the energy in Eq. (3) was shown to be NP-complete [24] due to cycles in the underlying graphical model representing the image lattice. Therefore, suitable approximations are necessary.

2.1. Structural constraints

It has been shown by [30] that obeying specific hard constraints is necessary for obtaining a structure-preserving alignment. To this end, constraints ensuring the monotonicity and continuity of warping have been proposed. These constraints given in Table 1 prevent large gaps and mirroring of some areas in the deformed image. They are conceptually similar to 0–1–2 HMMs in speech recognition and replace the smoothness terms T_h, T_v by constrained versions T_h^c, T_v^c which return infinity if the constraints are violated. Note that usually pairwise terms T are truncated terms, e.g. $T_\tau = \min(T, \tau)$. While this allows efficient optimization methods as for example fast distance transforms [31], it is impossible to simultaneously guarantee smooth warpings.

3. Warping algorithms

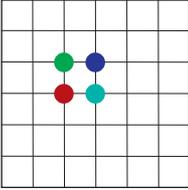
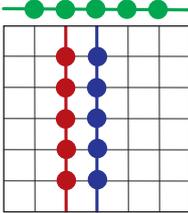
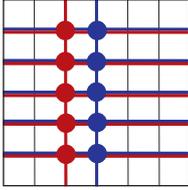
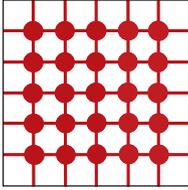
As stated before, it is NP-complete to find a global optimum of Eq. (3). Therefore, either the criterion has to be relaxed, or the

Table 1
Structural constraints as presented in [30].

Constraints	Monotonicity	Continuity
Horizontal	$0 \leq u_{ij} - u_{i-1,j} \leq 2$	$ v_{ij} - v_{i-1,j} \leq 1$
Vertical	$0 \leq v_{ij} - v_{i,j-1} \leq 2$	$ u_{ij} - u_{i,j-1} \leq 1$

Table 2

Dependency structures of warping algorithms. The lowest column shows dependencies between pixels by mutual color. In ZOW, all pixels are optimized independently. For P2D-like methods, pixel displacements are dependent within each column, and an additional HMM (green) optimizes column alignments. In TSDP, each column is optimized independently but estimates of horizontal branches are taken into account. Finally, graph-based methods try to optimize all displacements simultaneously. (For interpretation of the references to color in this table legend, the reader is referred to the web version of this article.)

Zero-order	First-order		
Point-based ZOW	Column-based P2D (+FOSE)	Tree-based (C-)TSDP	Graph based (C-)TRW-S
			

solution can only be approximated. In this section, we will review four principled approaches which are summarized in Table 2. We assess the strengths and weaknesses of all methods, and besides for the most simple one provide extensions significantly improving their performances.

3.1. Zero-order warping

In the most simple approximation no geometrical dependencies between neighboring pixels are considered, which leads to a point-based optimization. The smoothness function is replaced by a term $T_{\Delta}(ij, w_{ij})$ which penalizes the absolute deviation of each pixel ij from the position w_{ij} and restricts the maximum displacement in horizontal and vertical direction to a warping range Δ . This allows to rewrite energy function (2) as

$$E(X, R, \{w_{ij}\}) = \sum_{ij} [d(X_{ij}, R_{w_{ij}}) + T_{\Delta}(ij, w_{ij})]. \quad (4)$$

Minimization of the energy (4) can be done efficiently by optimizing the alignment of each pixel ij independently from the others. We call this approach zero-order warping (ZOW) which has been introduced, e.g. as Image Distortion Model in [32]. Despite this method being by far the fastest warping algorithm, its downside clearly is the lack of spatial information leading to unsmooth deformations. Additionally, the warp range Δ has to be chosen w.r.t. the task to disallow 'good' matches between images of different classes.

3.2. Pseudo two-dimensional warping

Another way of relaxing the original 2DW problem is to decouple the horizontal and vertical displacements. The resulting, column-based optimization is commonly denominated as Pseudo Two-Dimensional Warping (P2DW) [27,33,26]. Here, the decoupling leads to separate one-dimensional optimization problems which can be solved efficiently and optimally. An intra-column optimization finds an optimal matching between pixels in corresponding columns, while an inter-column optimization finds the column-to-column matching. This leads to the energy function (3) being transformed as follows:

$$\begin{aligned} E(X, R, \{w_{ij}\}) &= \sum_{ij} [d(X_{ij}, R_{w_{ij}}) + T_v(v_{ij}, v_{i,j-1}) + T_h(u_i, u_{i-1})] \\ &= \sum_i J \cdot T_h(u_i, u_{i-1}) + \sum_{ij} [d(X_{ij}, R_{w_{ij}}) + T_v(v_{ij}, v_{i,j-1})], \end{aligned} \quad (5)$$

where the horizontal smoothness is only preserved between entire columns by the slightly changed term T_h . The optimization then

corresponds to solving two 1D HMMs, which can be optimized globally using dynamic programming (DP) [27]. While the found optimum is global, the resulting alignment can not be guaranteed to be smooth since column-to-column matchings are optimized independently. Also, all pixels of a column must be aligned to pixels of one corresponding column, restricting the warping severely.

3.3. Extended pseudo two-dimensional warping

In [23], we proposed to permit horizontal deviations from the column centers while retaining the first-order dependencies between alignments in a strip. This results in a first-order strip extension of P2DW (P2DW-FOSE). The proposed approach allows for flexible alignments of pixels within a *strip* of width Δ of neighboring columns rather than within a single column.

Especially for large Δ , it is important to enforce structure-preserving constraints within a strip, since otherwise one facilitates matching of similar but non-corresponding pixels, degrading the discriminative power. Therefore, we model horizontal deviations from column centers while enforcing the structure-preserving constraints given in Table 1. They can easily be implemented in the smoothness penalty function T_v , by setting the penalty to infinity if the constraints are violated. Instead, we prevent the computation of all alignments by considering only those permitted by the constraints, by hard coding them in the optimization procedure.

According to the explained changes, we rewrite Eq. (5) as

$$\begin{aligned} E(X, R, \{w_{ij}\}) &= \sum_i J \cdot T_h(u_i, u_{i-1}) \\ &\quad + \sum_{ij} [d(X_{ij}, R_{w_{ij}}) + T_{cv}(w_{ij}, w_{i,j-1}) + T_{\Delta}(u_i, u_{ij})]. \end{aligned} \quad (6)$$

Here, T_{Δ} penalizes the deviations from the central column u_i of a strip, and $T_{\Delta} = \infty$ if $|u_i - u_{ij}| > \Delta$; T_{cv} is the smoothness term with continuity and monotonicity constraints. Compared to P2DW, the minimization of (6) is of slightly increased complexity, linearly depending on parameter Δ .

Fig. 3(b) (bottom row) exemplifies the advantages of the proposed approach over the original P2DW. It can clearly be seen that the deviations from columns allow to compensate for local and global misalignments, while the monotonicity and continuity constraints preserve the geometrical structure of the facial image. Both improvements lead to a visibly better quality of image warping. We accentuate that preserving structural constraints within a strip does not guarantee global smoothness, since strips are optimized independently. The latter can lead to intersecting paths in neighboring columns.

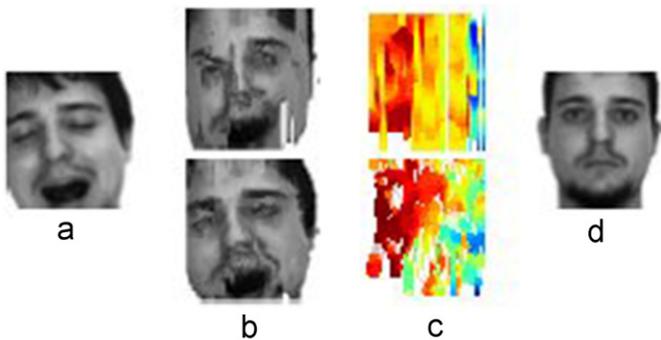


Fig. 3. The reference image (d) is warped to the query image (a) using P2DW (top row) and P2DW-FOSE approach (bottom row). The aligned reference image (b) shows vertical artifacts for P2DW while P2DW-FOSE allows for much better alignment due to flexible warping; (c) shows respective warping grids.

3.4. Tree-based optimization

Tree-serial dynamic programming (TSDP) [28] relaxes 2DW by representing the two-dimensional pixel grid as a series of individual pixel neighborhood trees. Each pixel tree i^* has its own assignment stem (column), but shares the horizontal branches with other trees. This allows to optimize each tree i^* independently:

$$E_{i^*}(X, R, \{w_{ij}\}) = \sum_j \left[\sum_i [d(X_{ij}, R_{w_{ij}}) + T_h(w_{i-1,j}, w_{ij}) + T_\Delta(ij, w_{ij})] + T_v(w_{i^*j-1}, w_{i^*j}) \right]. \quad (7)$$

The solution for Eq. (7) can be efficiently found by dynamic programming (DP) and the final global alignment is a composition of the alignments of the separate trees' stems. In [28], no hard pairwise geometrical constraints are enforced in the binary smoothness function. In order to keep the complexity of the optimization feasible, the authors penalize the absolute deviation of ij and w_{ij} by the term $T_\Delta(\cdot)$, s.t. $T_\Delta(\cdot) = 0$ if deviation $\leq \Delta$ and $T_\Delta(\cdot) = \infty$ otherwise. Therefore, we refer to this version of TSDP as TSDP- Δ .

3.5. TSDP with structural constraints

The original TSDP mainly suffers from the complexity being quadric in Δ , which restricts optimization to small absolute displacements. In [22], we presented constrained TSDP (CTSDP), which includes hard geometric constraints for a much more efficient optimization. We replace T_v, T_h in Eq. (7) with T_v^c, T_h^c , leading to increased effectiveness which allows us to discard absolute position penalty term T_Δ :

$$E_{i^*}(X, R, \{w_{ij}\}) = \sum_j \left[\sum_i [d(X_{ij}, R_{w_{ij}}) + T_h^c(w_{i-1,j}, w_{ij})] + T_v^c(w_{i^*j-1}, w_{i^*j}) \right]. \quad (8)$$

Given the hard constraints in the smoothness terms, paths in the DP recursion containing violated constraints have an infinite cost and will be discarded at the top level. Therefore, we can discard any recursion containing violated constraints or conversely, just recurse through trees with allowed label combinations. DP can easily be implemented by dedicated loops, leading to a complexity in $O(IJUV)$, while the complexity of the original TSDP- Δ is in $O(IJ\Delta^4)$. The latter rapidly outgrows the former for increasing Δ , because for each alignment w_{ij} all $(2\Delta + 1)^2$ possible alignments of neighboring positions have to be considered. We visually compare alignments resulting from TSDP- Δ with $\Delta = 17$ and CTSDP in Fig. 4, which shows that the alignment becomes much smoother using the hard structural constraints. Here, we deform the test

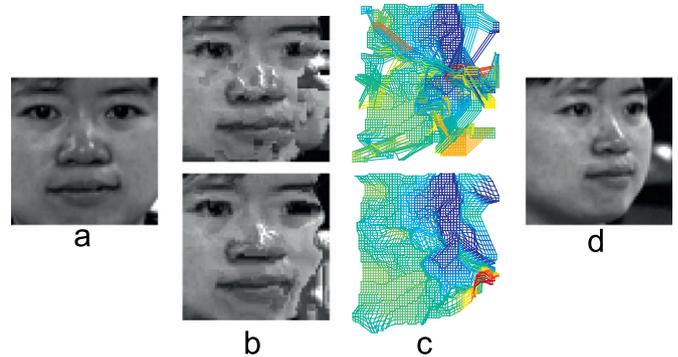


Fig. 4. Comparing TSDP without (top row) and with (bottom row) structural constraints on an example of the CMU-PIE poses database. Note that for poses, the alignment direction is reversed because of inaccurate cropping of the test images (cf. Section 5.3). The alignment using hard structural constraints is much smoother, while the original implementation contains obvious discontinuities both in the aligned reference image and the alignment grid. The alignment has been computed using SIFT features and applied to gray images. (a) Reference. (d) Test.

image (d) to best fit the reference (a) and show both the deformed test image in (b) and the deformed regular pixel grid in (c). Large artifacts are visible in the deformed test image due to huge displacement inconsistencies which are allowed in the original TSDP formulation, although penalized. In CTSDP, constraints between vertically neighboring pixels can be violated due to the independent optimization of the column trees. This is not likely though, since all column trees use the same horizontal branches.

3.6. Graph based optimization

Recent advances in Markov random field (MRF) inference makes the direct optimization of criteria Eq. (3) feasible. However, for non-convex priors, a globally optimal solution can not be guaranteed [34]. Tree-reweighted message passing algorithm (TRW-S) [29] ensures convergence to a (local) optimum and gives a lower bound on the energy which is guaranteed to monotonically increase. This can be utilized to assess the global quality of the local optimum (which is global if equal to the lower bound), and can be exploited for pruning nearest-neighbor (NN) search (cf. Section 4). TRW iteratively approximates this lower bound, which is a dual of the LP-relaxation of Eq. (3). TRW-S sequentially computes min-marginals $\Phi_{ij}(w_{ij})$, which are forced to be equal among subproblems, and performs re-parameterization by passing messages between neighboring nodes. Exploiting the structure of the subproblems, these computations can be efficiently combined.

3.7. TRW-S with structural constraints

We have shown in [19] that using hard constraints in a warping scheme approximating the optimal solution leads to smoother warpings and increased recognition performance. Here, we will briefly review the main idea.

Updating messages in TRW-S involves finding a local minimum w.r.t. a pixel alignment pair (w, w') . Eq. (9) exemplary shows the update of the forward message M^{fw} from (ij) to $(i+1, j)$ w.r.t. the label w' and consists of minimizing over all labels w of the sum of the corresponding pairwise potential, the respective backward message and the local unary potential:

$$\hat{M}_{(ij),(i+1,j)}^{fw}(w') = \min_w \{T_{(ij),(i+1,j)}^c(w, w') - M_{(ij),(i+1,j)}^{bw}(w) + d(X_{ij}, R_w)\}. \quad (9)$$

This minimum does not change when only pairs with $T^c(w, w') < \infty$ are considered, therefore these allowed pairs can be pre-computed and the minimization can be restricted to these

Table 3
Comparison of theoretical complexities of presented warping approaches.

Warping algorithm	Complexity
ZOW	$Ij(2\Delta + 1)^2$
P2DW	$3IU(1 + JV)$
P2DW-FOSE	$3IU(1 + 3\Delta JV)$
TSDP	$3Ij(2\Delta + 1)^4$
CTSDP	$3 \cdot 9IjUV$
TRW-S	$N \cdot 2Ij(UV)^2$
CTRW-S	$N \cdot 2Ij(9UV)$

pairs. According to the constraints in Table 1, the maximum number of allowed pairs is 9, leading to a speedup of the message update from $O((UV)^2)$ to $O(9 \cdot UV)$. Since the updating of messages is the main speed bottleneck of TRW-S, this provides a very significant speed-up, especially if the reference image is large.

3.8. Complexity

Different assumptions on dependencies which hold between single coordinates during image warping significantly influence the complexity of optimization. Here we compare theoretical complexities of the presented approaches. The results are listed in Table 3. Clearly, ZOW has the lowest complexity among warping methods due to independent local optimization of the displacement of each pixel. The complexity of P2DW is remarkably larger due to respecting first-order dependencies between coordinates in a column and entire columns. Additional flexibility implemented in P2DW-FOSE comes at a price of larger complexity, while the extension of pseudo two-dimensional optimization to trees further increases the complexity of warping, as then the optimization is also performed along tree branches. Comparing the complexities of CTSDP to CTRW-S, the optimization of CTSDP has an equal complexity to a single forward pass of CTRW-S. Since the latter has to perform both a forward and a backward pass, and additionally multiple iterations in order to converge, the runtime of CTSDP is at least two times faster than one single iteration of CTRW-S.

4. Implementation details

In this section we shortly describe implementation details which make the presented approaches robust to un-alignable pixels and help to reduce the overall complexity of recognition.

4.1. Occlusion handling

Facial occlusions represent a challenging problem for finding a dense alignment between local features. In general, they can be created by something as simple as sunglasses or a scarf, but also some parts of face may be invisible due to, e.g. closed eyes or rotation of the head (cf. Fig. 2). Therefore, in [22] we propose to use a local distance thresholding to deal with occlusions. Formally, the truncated distance function is expressed as $\tilde{d}_\tau(X_{ij}, R_{w_{ij}}) = \min(\tau, d(X_{ij}, R_{w_{ij}}))$, where τ is a threshold. This has several advantageous properties. First, it can be directly implemented in the local distance computation which is of minor complexity compared to the optimization. Second, it both reduces the impact of occluded pixels on the total distance and allows the optimization algorithms to more easily align non-occluded pixels while keeping influence of the occluded pixels low. It can be seen in Fig. 5 that thresholding produces a much smoother alignment and in addition local distances are influenced far less by occluded areas, since thresholded areas should be roughly the same for all reference images.

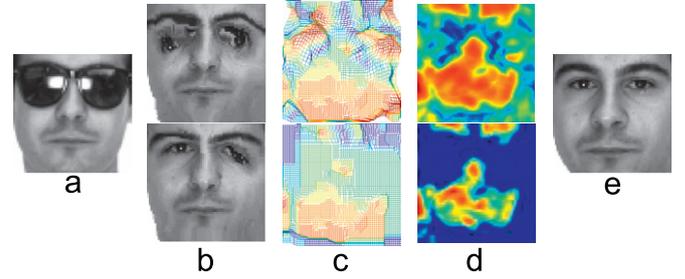


Fig. 5. 2D warping of reference to test image without (top row) and with (bottom row) occlusion handling on a sample from the AR-Face occlusion database. (a) Test. (b) Aligned reference, (c) deformation grid and (d) local similarity map between test and aligned reference image. Dark blue pixels denote low similarity and red pixels mean high similarity. The alignment computed using CTRW-S with SIFT features and applied to gray images. (e) Reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Caching

We extract a 128-dimensional SIFT [6] descriptor at each position of the regular pixel grid. As proposed by [35], we reduce the descriptor to 30 dimensions by means of PCA estimated on the respective training data and subsequently normalize each descriptor to unit length. For speedup, we cache all pairwise distances of the PCA-reduced SIFT feature descriptors during the initialization phase. We also extend the local distance d to include the local context of a pixel pair ij , w_{ij} . Assuming a context of 5×5 , the context-size normalized local distance becomes

$$d_{5 \times 5}(X_{ij}, R_{w_{ij}}) = \frac{1}{25} \sum_{\Delta_x} \sum_{\Delta_y} d(X_{i+\Delta_x, j+\Delta_y}, R_{u_{ij+\Delta_x, v_{ij+\Delta_y}}}), \quad (10)$$

with Δ_x and $\Delta_y \in -2, \dots, 2$. At image borders, the usable context and therefore the normalization term becomes smaller. Naively replacing d with $d_{5 \times 5}$ in Eq. (3) leads to a huge computational overhead, since local contexts of neighboring pixels strongly overlap and local distances are computed multiple times. We thus cache all local distances on the fly.

4.3. Pruning

As we propose in [19], optimizing the deformation between a test and a reference image can be stopped or even completely skipped if the lower bound on the energy of current comparison surpasses the lowest energy found so far. It was shown that the lower bound of TRW-S can be exploited without losing accuracy due to its guaranteed monotonicity. For all other warping methods, the sum of the lowest distances for all coordinates ij is used as a weak lower bound. This sum can be found during the distance pre-computation and thus speeds up the NN search, especially if a good warping is computed early.

5. Results

In this section, we present experimental results and show that preserving most neighborhood dependencies together with occlusion modeling helps to noticeably improve the results on four challenging face recognition tasks. First, we evaluate warping algorithms in a setting when only one training image (mugshot) is available, while test images have significant variations due to partial occlusions and changes in facial expression and head pose. Then, we show that when having more data the improved methods can cope with strong misalignments due to face registration errors which make the expression- and illumination-invariant face recognition even more challenging.



Fig. 6. Sample images from the AR Face database with automatically detected faces.

AR Face: The database [36] contains frontal facial images with different facial expressions, illuminations, and occlusions. The images correspond to 126 persons: 56 women and 70 men. Each individual participated in two sessions separated by 14 days. During each session 13 pictures per person were taken under the same conditions. Similar to [37], only a subset of 110 individuals for which all variability is available is used in our experiments.

CMU-PIE. The CMU-PIE [38] database consists of over 41 000 images of 68 individuals. Each person is imaged under 43 different illuminations, 13 poses and four various facial expressions. To evaluate the methods on 3D transformations, we use a subset of all subjects in 13 poses with neutral expressions.

Preprocessing: In the mugshot setting (AR Face Occlusions, AR Face Expressions, CMU-PIE poses) the original face images were manually aligned by eye-center locations [39], while in the setting with more training data (AR Face VJ) faces were automatically detected using publicly available OpenCV implementation of the Viola & Jones (VJ) face detector [40]. In both cases faces were cropped to 64×64 resolution. See for samples Figs. 2 and 6.

Experimental setup: As proposed by [35], we extract PCA-reduced SIFT feature descriptor [6] at each position of the regular pixel grid and then normalize each descriptor to unit length. We use a nearest-neighbor (NN) classifier for recognition directly employing the obtained energy as dissimilarity measure and the L_1 norm as local feature distance. For comparison, we use our publicly available implementation¹ of the presented warping algorithms except for CTRW-S for which we re-implement the approach of [19].

5.1. AR-Face occlusions

We evaluate presented warping approaches on partially occluded faces from the subset of the AR-Face database (cf. Fig. 2). We use the neutral non-occluded faces from session 1 as reference and all occluded faces from sessions 1 and 2 as test images leading to 440 test and 110 reference images.

We empirically found a suitable occlusion threshold ($\tau = 0.7$) value on the sunglasses occlusion subset of session 2 together with warping range $\Delta = 10$ for TSDP- Δ and ZOW, and strip width $\Delta = 5$ for P2DW-FOSE. Results for different thresholds and methods are shown in Fig. 7. Interestingly, the optimal threshold leads to a strong decrease in error. This value is very similar for all algorithms and thus most probably depends on the feature descriptor respective the distribution of the local distances. Too small thresholds lead to high error as too much discriminative information is pruned.

Comparison of the results with and without distance thresholding is provided in Table 4. It can be seen that warping algorithms with stricter dependencies and constraints suffer more from occluded parts of the face. This confirms our assumption that occluded pixels are not smoothed out by the global dependencies, but instead propagate to the entire image decreasing the overall quality of the match. In case of thresholding, all methods except ZOW produce nearly equally excellent recognition results

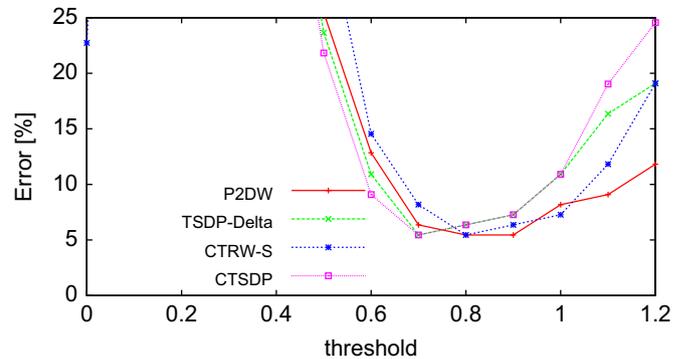


Fig. 7. Error rates of warping algorithms on the AR-Face sunglasses, session 2, with different levels of distance thresholding.

Table 4

Recognition error rates [%] on the AR-Face occlusion task using warping algorithms with and without occlusion modeling.

Model	Occlusion handling	
	No	Yes
No warping	39.22	38.10
ZOW	6.79	2.46
P2DW	7.21	1.91
P2DW-FOSE	6.00	1.48
TSDP- Δ	6.79	1.69
CTSDP	9.45	1.48
CTRW-S	8.27	1.69
SURF [1]	10.54	–
DCT [37]	3.59	–
Partial Dist. [12]	4.67 ^a	–
Stringface [41]	13.00	–
PWCM _r [11]	–	16.00
LGBPHS [8]	16.00	–

^a Used only a subset of occlusions.

with CTSDP and P2DW-FOSE achieving the best performance. Two points should be noted:

1. CTSDP which implements structural constraints performs superior to the original version with absolute constraints despite being much more general and efficient. For TSDP- Δ , the warping range $\Delta = 10$ performs best, while CTSDP models deformations of arbitrary magnitude.
2. Good performance of P2DW can be accounted to pre-aligned face images, in which virtually no global rotations are present. Additional flexibility implemented in the P2DW-FOSE helps to further reduce the recognition error, as this approach is able to cope with local rotations.

Compared to the state of the art, we conclude that all presented warping methods clearly outperform the competition. Since all competitive approaches are outperformed even by the ZOW despite using zero-order-like matching algorithms themselves, it can be concluded that the SIFT descriptor in combination with occlusion modeling already provides a significant advantage. Here, [37] uses DCT features extracted from non-overlapping blocks, LGBPHS [8] use local Gabor binary pattern histograms and [1] evaluates both SURF and SIFT features using a locally restricted matching scheme. Another two non-learning approaches, namely Stringface [41] and Partial Distance [12], employ matching procedure for recognition and hence are similar to our method: the former one is inspired by a string-based matching after representing a face as an attribute string, while the latter one uses nonmetric partial similarity measure. In opposite

¹ <http://www.hltpr.rwth-aachen.de/w2d/>

to warping approaches, the remaining two methods build a model from the training data, where SOM [42] learns a self-organizing map feature representation from data and PWCM_r [11] learns occlusion masks in order to reconstruct invisible parts from other faces where corresponding regions are not occluded. It is worth to point out that being model-free and hence very general, warping methods are able to achieve much better results, especially in comparison to the learning-based methods.

5.2. AR-Face expressions

Appearance variability due to changes in facial expressions is one of the challenging problems for face recognition algorithms. In order to achieve high recognition accuracy the presented approaches have to be able to cope with strong non-linear images deformations often occurring due to expression changes. We thus evaluate warping methods of a subset of the AR-Face database containing three different expressions taken in each of the two sessions and use the neutral expressions of Session 1 as reference images. Detailed recognition results are presented in Table 5. It can clearly be seen that both temporal and strong non-linear variation due to the *scream* expression poses the most difficult recognition task. All warping algorithms greatly outperform the un-aligned distance, while again the methods with structural constraints and additional flexibility of warping provide small but noticeable improvements with CTSDP consistently achieving the best performance. In comparison to the state of the art the presented approaches achieve much better recognition results. This is interesting since all competitive methods except for Partial Distance [12] use an essentially larger amount of training data to learn a representative subspace via Gaussian mixtures [43] and self-organizing maps [3,12]. Although the competitors perform worse than the presented warping algorithms, they are probably more efficient.

5.3. CMU-PIE poses

Pose invariant face recognition is a difficult problem due to ambiguities which contained in two-dimensional projections of three-dimensional head pose transformations. As the presented approaches do not rely on any 3D model, all variability has to be inferred in 2D space. We test the ability of warping algorithms to cope with head pose changes by evaluating them on the pose subset of the CMU-PIE database. We consider the pictures of 68 individuals with neutral facial expression shown from 13 view-points. The frontal image is used as reference and all remaining 12

poses are used as testing images. As reference (frontal) image is more accurately cropped compared to the test images containing much background (cf. Fig. 2 (bottom row)), we reverse the alignment procedure and deform the test image to the reference image, which helps to minimize the impact of background pixels, as proposed in [18]. Following the same work, we automatically generate and additionally use the left and right half of the reference face images in order to recognize near profile faces. Slightly larger threshold $\tau = 1.1$ is used for this task because of the higher background variability. Also, for ZOW and TSDP- Δ warping range Δ has to be increased to 17 leading to a prominent increase in complexity. Further increasing Δ might decrease error rate, but becomes infeasible because of the quadric complexity in TSDP- Δ method.

Detailed recognition results of warping algorithms over poses are shown in Fig. 8. It can be seen that obeying strong geometric constraints leads to a large improvement in recognition accuracy provided by CTSDP in comparison to TSDP- Δ . Also, P2DW-FOSE noticeably outperforms P2DW due to more flexible deformation. Overall, the enhanced methods outperform the weaker approaches by a large margin when compared on difficult near profile facial images (Fig. 8, left and right parts of the plot). Performance difference is formalized in Table 6. According to the results, it is clear that using structural constraints and occlusion modeling is imperative for achieving excellent recognition performance. This is supported by the fact that CTRW-S with occlusion handling

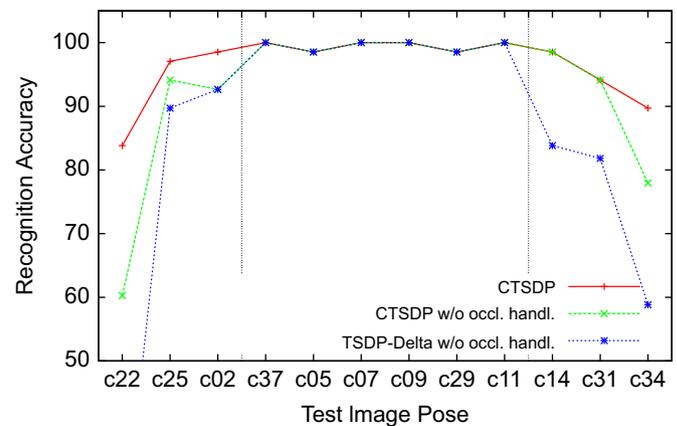


Fig. 8. Detailed plot of recognition accuracy (%) of the warping algorithms with occlusion handling across poses. Vertical grid lines divide between near profile (leftmost and rightmost) and near frontal (center division) poses as used in Table 6. It can be seen that all algorithms perform similar on near frontal poses, while the performance on near profile poses increases with thresholding and obeyed constraints in the models.

Table 5

Recognition error rates [%] on the AR-Face expressions obtained by warping algorithms with occlusion handling and comparison to the state of the art.

Model	Session 1			Session 2			Avg.
	Smile	Anger	Scream	Smile	Anger	Scream	
No warping	2.73	9.10	37.27	5.45	6.36	52.73	18.23
ZOW	0.00	0.00	3.64	0.91	1.82	17.27	3.93
P2DW	0.00	0.00	3.64	0.91	0.91	19.09	4.09
P2DW-FOSE	0.00	0.00	2.73	0.91	0.91	17.27	3.64
TSDP- Δ	0.00	0.00	3.64	0.91	1.82	17.27	3.94
CTSDP	0.00	0.00	4.55	1.82	0.91	13.64	3.49
CTRW-S	0.00	0.00	3.64	0.91	0.91	16.36	3.64
Partial Dist. [12]	0.00	3.00	7.00	12.00	14.00	37.00	12.00
Aw-SpPCA [3]	0.00	2.00	12.00	12.00	10.00	36.00	12.00
SOM [42]	0.00	2.00	12.00	12.00	10.00	36.00	12.00
SubsetModel [43]	3.00	10.00	17.00	26.00	23.00	25.00	17.00

Table 6

Average error rates (%) on CMU-PIE groups of poses.

Model	Near frontal	Near profile	Avg.
No warping	40.69	86.27	63.48
ZOW	0.49	31.61	16.05
P2DW	0.25	17.63	8.94
P2DW-FOSE	0.25	10.39	5.32
TSDP- Δ	0.98	25.36	13.17
CTSDP	0.25	7.35	3.80
CTRW-S	0.49	6.37	3.43
Hierarch. match. [18]	1.22	10.39	5.76
3D shape mod. [16]	0.00	14.40 ^a	6.55 ^a
Prob. learning [17]	7 ^b	32 ^b	19.30

^a Missing poses.

^b Estimated from graphs.

outperforms all other weaker warping methods, as well as the state-of-the-art approaches. In particular, CTSDP and CTRW-S outperform the hierarchical matching algorithm [18] that uses no occlusion modeling and decouples horizontal and vertical displacements, which allows for fast distance transforms but leads to a less tight lower bound. Zhang et al. [16] use an additional profile shot (pose 22) as reference and generate virtual intermediate poses by means of a 3D model. Since they use more training data and omit the most difficult pose in recognition, their experiments are not entirely comparable. The method of [17] uses automatically cropped images, which is an interesting task to be tackled with warping algorithms.

Comparing the runtimes for one image warping on a recent CPU, CTSDP is about ten times faster than TSDP- Δ (33 s vs. 350 s) and also 1.5 faster compared to one iteration of CTRW-S (45 s). P2DW-FOSE requires 43 s which is noticeably more than the runtime of P2DW (30 s) due to additional flexibility which comes at a price of higher complexity.

5.4. AR Face VJ

Face registration errors can have a large impact on the performance of face recognition approaches [44] due to strong misalignments of facial images. Here we show that, when using more training data, enhanced warping algorithms are not only able to cope with local variability caused by expression and illumination changes, but also robust to global changes due to registration errors. For that purpose, we use the subset of the AR Face database with different expressions and illumination conditions. Seven images of each person from the first session are used for training and the same number from the second session for testing. Simulating a real world environment we detect and crop the faces automatically, as described before.

In-plane face rotations due to noticeable misalignments of automatically detected faces can only be compensated by the approaches allowing deviations from columns during image warping. We study the extent of required flexibility by evaluating the strip width in P2DW-FOSE. The results are presented in Fig. 9 which shows the recognition error for the increasing deviation δ . It can be seen that already the smallest deviation helps to remarkably reduce the recognition error. Although the error decreases further afterwards, the return is diminishing quickly. This gives rise to two interpretations: on the one hand, it seems most important to allow (even slight) horizontal movements of individual pixels. On the other hand, big strip widths increase the chance of intersecting column paths, making the deformation less smooth.

In Table 7, we summarize our findings and compare performance of the presented methods with the results from the literature. It is clear that increasing the flexibility by strip extension in P2DW-FOSE greatly improves the accuracy compared to P2DW. Strong geometrical constraints in CTSDP and CTRW-S help to consistently improve the performance over weaker TSDP- Δ , which again supports the intuition that preserving image structure during the deformation is

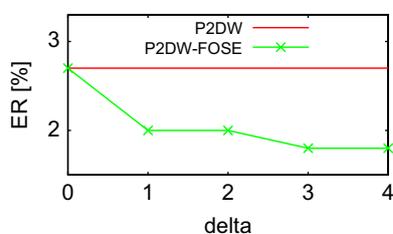


Fig. 9. Error rate on automatically detected faces for different strip widths Δ , where $\Delta = 0$ is equivalent to the P2DW.

Table 7

Recognition error rate (%) on the subset of AR Face with VJ-detected faces.

Model	ER (%)
No warping	22.3
ZOW	3.1
P2DW	2.7
P2DW-FOSE	1.8
TSDP- Δ	2.2
CTSDP	2.1
CTRW-S	2.0
SURF-Face [1]	4.1
DCT [37]	4.7 ^a
Av-SpPCA [3]	6.4 ^a

^a With manually aligned faces.

a key to achieving good recognition results in the presence of strong misalignments and non-linear image deformations.

The presented warping algorithms greatly outperform state-of-the-art feature matching approaches [1–3] which are though more efficient. Moreover, [2,3] use manually pre-register faces and thus solve much easier recognition task.

5.5. Qualitative evaluation

Experimental results suggest that preserving strong neighborhood dependencies is imperative for achieving high recognition performance. For in-depth analysis we perform a qualitative evaluation of the presented warping approaches and show computed deformations for faces with partial occlusions, variable expressions and poses. The results are shown in Table 8. For each task, we provide a test image and resulting deformations of the reference image belonging to the correct and competing class. As it can be seen, occlusion handling helps to compute relatively smooth deformations even by methods which does not preserve strong neighborhood constraints (occlusion task). Smooth warping by those methods can also be explained by a low variability in facial images which look similar apart from sunglasses. However, when variability increases (as in case of pose and expression), the visual difference in image warpings gets more remarkable. ZOW and TSDP- Δ compute unsmooth deformations of the reference image and tend to reconstruct the test image more strictly. This makes it harder to discriminate between the correct and competing classes, which negatively affects the classification performance. This is important, since in general the dissimilarity measure obtained by the warping algorithms is not optimized for discriminativeness, but tries to find the most similar transformation of the reference. Image deformations computed by P2DW and P2DW-FOSE are smooth for pose, but suffer from column artefacts clearly seen in the expression task. Expectedly, P2DW-FOSE can better reconstruct image areas having small rotations (e.g. eyes of the reference image in the pose task) compared to P2DW. Results obtained by CTSDP and CTRW-S clearly show that suitable geometric models with structure-preserving constraints are imperative to obtain a discriminative distance measure as well as visually smooth warpings.

6. Conclusion

In this work we performed a systematic analysis and thorough comparison of existing warping algorithms on the challenging tasks of face recognition in highly variable domains. For better understanding of the strengths and weaknesses of presented approaches we arrange them into a global hierarchy of groups of methods w.r.t. neighboring dependencies which hold during

Table 8
Qualitative evaluation of the presented warping approaches.

Method	Task					
	Occlusion		Expression		Pose	
Test						
Reference						
ZOW						
P2DW						
P2DW-FOSE						
TSDP- Δ						
CTSDP						
C _T TRW-S						

optimization. By gradually proceeding from very simple locally optimal methods towards more complex approaches which try to optimize all dependencies simultaneously we showed that the increasing number of neighborhood relations facilitates the computation of smoother image warpings and leads to more accurate face recognition results. Compared to other warping algorithms, it becomes clear that using weaker smoothness paradigms is less reliable when trying to cope with the variability induced by projections of 3D transformations. While results on the expressions and occlusions tasks do not vary strongly over presented warping approaches, significant differences can be observed on the CMU-PIE database, where the approaches optimizing all dependencies simultaneously perform the best while being almost as efficient as more simple heuristics.

References

- [1] P. Dreu, P. Steingrube, H. Hanselmann, H. Ney, Surf-face: face recognition under viewpoint consistency constraints, in: *BMVC*, 2009, pp. 1–11.
- [2] H.K. Ekenel, R. Stiefelhagen, Analysis of local appearance-based face recognition: effects of feature selection and feature normalization, in: *CVPRW '06*, Washington, DC, USA, 2006, p. 34.
- [3] K. Tan, S. Chen, Adaptively weighted sub-pattern PCA for face recognition, *Neurocomputing* 64 (2005) 505–511.
- [4] J. Wright, G. Hua, Implicit elastic matching with random projections for pose-variant face recognition, *IEEE CVPR 0* (2009) 1502–1509.
- [5] G. Hua, A. Akbarzadeh, A robust elastic and partial matching metric for face recognition, in: *ICCV*, 2009, pp. 2082–2089.
- [6] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2004) 91–110.
- [7] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, in: *ECCV*, 2004, pp. 469–481.

- [8] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence: a novel non-statistical model for face representation and recognition, in: ICCV, vol. 1, IEEE Computer Society, Washington, DC, USA, 2005, pp. 786–791.
- [9] H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, Surf: speeded up robust features, *Computer Vision and Image Understanding* 110 (2008) 346–359.
- [10] R. Singh, M. Vatsa, A. Noore, Face recognition with disguise and single gallery images, *Journal of Image and Vision Computing* 27 (2009) 245–257.
- [11] H. Jia, A. Martínez, Face recognition with occlusions in the training and testing sets, in: IEEE FG, 2008, pp. 1–6.
- [12] X. Tan, S. Chen, Z. Zhou, J. Liu, Face recognition under occlusions and variant expressions with partial similarity, *IEEE Transactions on Information Forensics and Security* 4 (2009) 217–230.
- [13] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997) 775–779.
- [14] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models—their training and application, *Computer Vision and Image Understanding* 61 (1995) 38–59.
- [15] G.J. Edwards, T.F. Cootes, C.J. Taylor, Face recognition using active appearance models, in: ECCV, vol. 2, Springer-Verlag, London, UK, 1998, pp. 581–595.
- [16] X. Zhang, Y. Gao, M.K.H. Leung, Recognizing rotated faces from frontal and side views: an approach toward effective use of mugshot databases, *IEEE Transactions on Information Forensics and Security* 3 (2008) 684–697.
- [17] M. Sarfraz, O. Hellwich, Probabilistic learning for fully automatic face recognition across pose, *Journal of Image and Vision Computing* 28 (2010) 744–753.
- [18] S. Arashloo, J. Kittler, Hierarchical image matching for pose-invariant face recognition, in: BMVC, 2009, pp. 1–11.
- [19] T. Gass, P. Dreuw, H. Ney, Constrained energy minimisation for matching-based image recognition, in: ICPR, Istanbul, Turkey, 2010, pp. 3304–3307.
- [20] S. Liao, A. Chung, A novel Markov random field based deformable model for face recognition, in: IEEE CVPR, 2010, pp. 2675–2682.
- [21] G. Oxholm, K. Nishino, Membrane nonrigid image registration, in: ECCV, 2010, pp. 763–776.
- [22] T. Gass, L. Pishchulin, P. Dreuw, H. Ney, Warp that smile on your face: optimal and smooth deformations for face recognition, in: IEEE Automatic Face and Gesture Recognition 2011 (FG), IEEE, Santa Barbara, USA, 2011, pp. 456–463.
- [23] L. Pishchulin, T. Gass, P. Dreuw, H. Ney, The fast and the flexible: Extended pseudo two-dimensional warping for face recognition, in: Iberian Conference on Pattern Recognition and Image Analysis, Gran Canaria, Spain, 2011, pp. 1–8.
- [24] D. Keysers, W. Unger, Elastic image matching is NP-complete, *Pattern Recognition Letters* 24 (2003) 445–453.
- [25] S.J. Smith, M.O. Bourgoin, K. Sims, H.L. Voorhees, Handwritten character classification using nearest neighbor in large databases, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (1994) 915–919.
- [26] F. Samaria, Face Recognition Using Hidden Markov Models, Ph.D. Thesis, Cambridge University, 1994.
- [27] S. Eickeler, S. Miller, G. Rigoll, High performance face recognition using pseudo 2-d hidden Markov models, in: European Control Conference (ECC), 1999, pp. 1–6.
- [28] V. Mottl, A. Kopylov, A. Kostin, A. Yermakov, J. Kittler, Elastic transformation of the image pixel grid for similarity based face identification, in: ICPR, IEEE Computer Society, 2002, p. 30549.
- [29] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006) 1568–1583.
- [30] S. Uchida, H. Sakoe, A monotonic and continuous two-dimensional warping based on dynamic programming, in: ICPR, 1998, pp. 521–524.
- [31] P. Felzenszwalb, D. Huttenlocher, Distance Transforms of Sampled Functions, Cornell Computing and Information Science Technical Report, 2004.
- [32] D. Keysers, T. Deselaers, C. Gollan, H. Ney, Deformation models for image recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (2007) 1422–1435.
- [33] S.S. Kuo, O.E. Agazzi, Keyword spotting in poorly printed documents using pseudo 2-d hidden Markov models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (1994) 842–848.
- [34] H. Ishikawa, Exact optimization for Markov random fields with convex priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2003) 1333–1336.
- [35] Y. Ke, R. Sukthankar, PCA-sift: a more distinctive representation for local image descriptors, in: IEEE CVPR, vol. 2, 2004, pp. 506–513.
- [36] A. Martínez, R. Benavente, The AR Face Database, Technical Report, CVC Technical Report, 1998.
- [37] H.K. Ekenel, R. Stiefelhagen, Why is facial occlusion a challenging problem? in: ICB, 2009, pp. 299–308.
- [38] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression (PIE) database, in: IEEE AFGR, 2002, pp. 46–51.
- [39] R. Gross, 2001, <<http://ralphgross.com/FaceLabels>>.
- [40] P. Viola, M. Jones, Robust real-time face detection, *International Journal of Computer Vision* 57 (2004) 137–154.
- [41] W. Chen, Y. Gao, Recognizing partially occluded faces from a single sample per class using string-based matching, in: ECCV, vol. 3, 2010, pp. 496–509.
- [42] X. Tan, S. Chen, Z. Zhou, F. Zhang, Recognizing partially occluded, expression variant faces from single training image per person with som and soft KNN ensemble, *IEEE Transactions on Neural Networks* 16 (2005) 875–886.
- [43] A. Martínez, Y. Zhang, Subset modeling of face localization error, occlusion, and expression, in: Face Processing: Advanced Modeling and Methods, 2005, pp. 577–615.
- [44] E. Rentzeperis, A. Stergiou, A. Pnevmatikakis, L. Polymenakos, Impact of face registration errors on recognition, in: AIAI, 2006, pp. 187–194.

Leonid Pishchulin received his master degree in Computer Science in 2010 from RWTH Aachen University, Germany. Since June 2010 he is a Ph.D. student in Computer Vision and Multimodal Computing group at Max Planck Institute for Informatics, Germany. His research interests are people detection, pose estimation and face recognition.

Tobias Gass is a Ph.D. student at the Computer Vision Laboratory at ETH Zurich. Before that he was a researcher at the Human Language Processing Group of the Computer Science Department of the RWTH Aachen University in Aachen, Germany. He received his diploma in computer science from RWTH Aachen University in 2009. His research interests cover 3D image registration and segmentation, image recognition and machine learning.

Philippe Dreuw is currently working as Ph.D. research assistant at the Computer Science Department of RWTH Aachen University. In 2003 he joined the Human Language Technology and Pattern Recognition Group headed by Prof. Dr.-Ing. Hermann Ney. His research interests cover sign language recognition, face recognition, object tracking, and off-line handwriting recognition.

Hermann Ney is a full professor of computer science at RWTH Aachen University, Germany. Before, he headed the speech recognition group at Philips Research. His main research interests lie in the area of statistical methods for pattern recognition and human language technology and their specific applications to speech recognition, machine translation and image object recognition. In particular, he has worked on dynamic programming for continuous speech recognition, language modeling and phrase-based approaches to machine translation. He has authored and co-authored more than 450 papers in journals, books, conferences and workshops. From 1997 to 2000, he was a member of the Speech Technical Committee of the IEEE Signal Processing Society. In 2006, he was the recipient of the Technical Achievement Award of the IEEE Signal Processing Society.