View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems | School of Information Systems

10-2009

Unsupervised Face Alignment by Robust Nonrigid Mapping

Jianke ZHU
ETH Zurich, Switzerland

Luc VAN GOOL
KU Leuven

Steven C. H. HOI
Singapore Management University, CHHOI@smu.edu.sg

DOI: https://doi.org/10.1109/ICCV.2009.5459325

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons

Citation

ZHU, Jianke; VAN GOOL, Luc; and HOI, Steven C. H.. Unsupervised Face Alignment by Robust Nonrigid Mapping. (2009). *IEEE 12th International Conference on Computer Vision ICCV 2009: Kyoto, Japan, 29 September - 2 October 2009*. 1265-1272. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/2371

This Conference Proceeding Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Unsupervised Face Alignment by Robust Nonrigid Mapping

Jianke Zhu[1]
[1]Computer Vision Lab
ETH Zurich, Switzerland
{zhu,vangool}@vision.ethz.ch

Luc Van Gool[1,2]
[2]PSI-VISICS, ESAT
KU Leuven, Belgium
Luc.Vangool@esat.kuleuven.be

Steven C.H. Hoi[3]
[3]School of Comp. Eng.
NTU, Singapore
chhoi@ntu.edu.sg

Abstract

We propose a novel approach to unsupervised facial image alignment. Differently from previous approaches, that are confined to affine transformations on either the entire face or separate patches, we extract a nonrigid mapping between facial images. Based on a regularized face model, we frame unsupervised face alignment into the Lucas-Kanade image registration approach. We propose a robust optimization scheme to handle appearance variations. The method is fully automatic and can cope with pose variations and expressions, all in an unsupervised manner. Experiments on a large set of images showed that the approach is effective.

1. Introduction

Image alignment or registration is essential to many vision tasks. These include depth extraction, mosaicking, and motion analysis, but also object tracking [25] and recognition [15]. In particular, face recognition [1, 2] can benefit by first bringing the faces into a canonical pose.

Face alignment is the task of finding a transformation between two facial images so that they can be matched as good as possible. Facial appearance changes with different poses and expressions. In addition, there are large differences among individuals.

Recently, unsupervised image alignment techniques, such as Learned-Miller's congealing [11] and least squares congealing [7], emerged as a new promising direction toward face alignment. Unlike typical supervised image alignment methods [6, 23, 10, 19] that often require manual labeling, the unsupervised approaches only assume that the parameterization of the alignment is known and the input images have similar appearance, making them more flexible and practical for some real applications.

Least squares congealing [7] treats the entire facial region at once. It assumes an affine transformation suffices for the alignment. Such approach can only deal with frontal faces with a neutral expression. Moreover, the method has a high computational cost and can therefore only handle few images at a time. To better deal with pose variations, one can split the face into several patches and apply image alignment to each patch separately. Such approach [1] improves recognition performance under pose variations, but discards the consistency between neighboring patches and still assumes a rigid transformation per patch. Splitting a face into disconnected patches also precludes the synthesis of photo-realistic, rectified facial images.

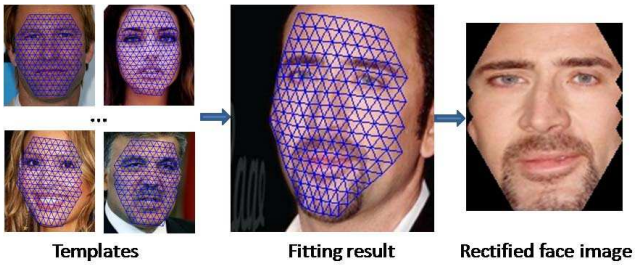


Figure 1. Illustration of our unsupervised face alignment with a nonrigid mesh model. We first fit a set of template images to the input image by a robust deformable Lucas-Kanade fitting scheme. Then, the input face image is rectified into the canonical frontal view.

In this paper, we address these current limitations by allowing for *nonrigid* transformations. Inspired by Zhu et al.'s nonrigid shape recovery [25], the mapping between two facial images is parameterized directly by the vertex coordinates. We employ a nonrigid face model to triangulate a template image and to preserve the regularity of deformations. The related optimization problem can be efficiently solved by the Lucas-Kanade algorithm for deformable registration [19]. Moreover, we propose a robust optimization scheme to account for outliers. Furthermore, to incorporate more appearance information from different individuals, we propose a joint face alignment scheme can employ multiple templates simultaneously. Once the nonrigid mapping is found, we can rectify the face image into a canonical frame. The rectified face images can be directly used for recognition and face swapping applications. Fig. 1 illustrates the main ideas of the proposed unsupervised face alignment approach.

In summary, the main contributions of this paper are: (1)

a novel unsupervised face alignment method, which allows for non-rigid mapping; (2) a robust and efficient optimization scheme to make the solution reliable against outliers; (3) a joint face alignment scheme that allows to use multiple templates to deal with large appearance changes; (4) experiments on large sets of images, comparing against state-of-the-art face alignment techniques, showing good qualitative and quantitative performance.

The rest of this paper is organized as follows. Section 2 goes somewhat deeper into previous methods for face/image alignment. Section 3 proposes our novel approach to tackle the unsupervised, deformable face alignment problem, and presents the optimization scheme. Section 4 describes the experimental results. Section 5 concludes the paper and suggests some future work.

2. Related Work

Face alignment received quite a bit of attention already. A traditional show-stopper has been the amount of user interaction needed. Some methods [6, 23, 10, 19] depend on statistical shape models built from a set of representative examples, in which landmark points have to be manually labelled. Blanz and Vetter [5] proposed a sophisticated 3D model learnt from 3D scans. The alignment is computationally rather intensive and often requires a good initialization.

To align facial images without any supervision, earlier methods have focused on estimating a set of aligned basis images to account for spatial variations. Most of this work is closely related to subspace learning again, such as principal component analysis and transformed component analysis [9]. Recently, a new strand of fully unsupervised face alignment from exemplars has been introduced. The seminal congealing method [15, 11] employed a sum of entropies cost function and a sequential algorithm to find some transformation parameters. Later, Cox et al. [7] extended this approach by introducing a sum-of-squared error cost function. The alignment was formulated under the Lucas-Kanade framework [17], and the optimization can be iteratively solved by the Gauss-Newton method. However, their approach only handles frontal face alignment, using 2D affine transformations. Only a small set of facial images with neutral expressions and uniform lighting were evaluated.

Our work is also related to viewpoint invariant face recognition. Kanade and Yamada [14] developed a probabilistic model of how face appearance changes with viewpoint, in which the facial images are separated into a set of independent patches. Later, Lucey and Chen [18] proposed a joint distribution model of individual patches to deal with the misalignment problem. Ashraf et al. [1] added a stack flow algorithm to find the spatial deformations through the transformation between the separate patches. Although gaining great improvements on face recognition under viewpoint changes, this method assumes the face is aligned and the head pose is known. In addition, the input facial images must be grouped by their poses at the start.

Unlike existing approaches, our unsupervised face alignment method can effectively handle various challenging conditions, including pose and moderate lighting variations as well as differences in identity and expression. This is because we can find the nonrigid mapping between two facial images. To the best of our knowledge, this is the first unsupervised alignment scheme that finds such nonrigid mapping.

3. Unsupervised Face Alignment

In this section, we present the proposed framework for unsupervised face alignment. Firstly, we formulate the face alignment problem as a Deformable Lucas-Kanade (DLK) fitting task, employing a 2D nonrigid face model directly parameterized by the mesh vertex coordinates (section 3.1). Secondly, we adapt it further to our face problem (3.2), through an alignment algorithm which can handle outliers effectively (3.2.1) and an efficient dual inverse compositional method to address lighting variations (3.2.2). Thirdly, we suggest a joint face alignment approach that employs multiple template images to improve the performance by incorporating more information (3.3).

3.1. Problem Formulation

The key idea of alignment is to iteratively align a single image with respect to another by minimizing the sum of squared differences between these two images. Let I and T denote an input image and a template image, respectively. The goal is to find parameters $\mathbf{p}$ that minimize the following energy function:

$$E(\mathbf{p}) = \sum_{\mathbf{x}} \Big(I(W(\mathbf{x};\mathbf{p})) - T(\mathbf{x}) \Big)^2 \qquad (1)$$

where $W(\mathbf{x};\mathbf{p})$ is some parametric warping function for the pixel coordinates $\mathbf{x}$. As already discussed, a global or even patch-wise affine warp does not suffice for faces. Instead, we overlay a triangular mesh (see Fig. 2(a)) and directly parameterize the transformation by the mesh vertex coordinates $\mathbf{s} = \begin{bmatrix} x_1 & \dots & x_N & y_1 & \dots & y_N \end{bmatrix}^\top$.

This brings the formulation close to that of general image registration approaches, such as the Lucas-Kanade algorithm [17]. Cox et al. [7] already proposed such unsupervised face alignment formulation. With its $2N$ free variables, the problem is ill-posed. To overcome this challenge, we adopt the algorithm proposed by Zhu et al. [25], who successfully tackle this optimization problem by introducing a regularization term. This term represents the mesh deformation, and is composed of the sum of the squared

(a) Model mesh (b) Template (c) Initialization (d) Sampled face (e) Fitting results (f) Rectified face (g) Average

Figure 2. Example of image-to-image face alignment. (a) the model mesh is automatically generated from a frontal face image based on detected eye corners and the distance between two eyes. (b) the template image is mapped onto the reference frame. (c) the model mesh is aligned to the input image helped by the detected eye corners. (e) the resulting mesh is overlaid on the input image. (f) We rectify the face image by mapping it onto the reference frame. We can now compare the differences between the initialization (d) and the rectified image (f). The average face of the template (b) and the rectified face (f) is plotted in (g).

second-order derivatives of the mesh vertex coordinates. Specifically, let $\lambda_{\mathbf{s}}$ be a regularization coefficient, we employ the inverse compositional method [2] to formulate the problem as follows:

$$E(\mathbf{s}) = \sum_{\mathbf{x}} \Big(T(\mathbf{x}; \Delta\mathbf{s}) - I(W(\mathbf{x}; \mathbf{s})) \Big)^2 + \lambda_{\mathbf{s}} \mathbf{s}^\top K \mathbf{s} \quad (2)$$

where $\Delta\mathbf{s}$ is defined as the increments to the mesh vertices and the matrix $K \in \mathbb{R}^{2N \times 2N}$ is a sparse matrix determined by the neighborhood structure of the mesh model [21, 24, 25].

The nonlinear optimization problem in Eqn. 2 can be linearized by taking the first order Taylor expansion:

$$E(\mathbf{s}) \approx \sum_{\mathbf{x}} \left(T + A\Delta\mathbf{s} - I(\mathbf{s}) \right)^2 + \lambda_{\mathbf{s}} (\mathbf{s} + \Delta\mathbf{s})^\top K (\mathbf{s} + \Delta\mathbf{s})$$

where $A = \nabla T \frac{\partial W}{\partial \mathbf{s}}$ is the steepest descent image, ∇T is defined as the gradient of T, and $\frac{\partial W}{\partial \mathbf{s}}$ is the Jacobian of the warp parameters evaluated at $\mathbf{s}$. Let $H \in \mathbb{R}^{2N \times 2N}$ denote the Hessian matrix:

$$H = \sum_{\mathbf{x}} A^\top A + \lambda_{\mathbf{s}} K \quad (3)$$

Thus, the solution of the above equation can be computed below:

$$\Delta\mathbf{s} = H^{-1} \left(\sum_{\mathbf{x}} A^\top (I(\mathbf{s}) - T) - \lambda_{\mathbf{s}} K \mathbf{s} \right) \quad (4)$$

The solution to the deformable Lucas-Kanade algorithm in Eqn. 2 can be found through the Gauss-Newton optimization method [2]. In particular, we iteratively calculate $\Delta\mathbf{s}$ and update the estimated mesh vertices by $\mathbf{s} \leftarrow \mathbf{s} - \Delta\mathbf{s}$.

3.2. Image-to-Image Face Alignment

The above algorithm was originally designed for deformable object tracking, but still falls short of successfully aligning faces across different individuals with various poses and expressions. As a solution, we propose a robust fitting scheme and a dual inverse compositional algorithm. They address practical issues such as outliers and lighting changes. Facial feature detectors are employed to locate the eye corners in order to facilitate a fully automatic approach. More specifically, we initially position and scale the template mesh based on the eye coordinates and the inter-ocular distance. The whole process for the proposed image-to-image face alignment is illustrated in Fig. 2.

3.2.1 Robust Deformable Lucas-Kanade Algorithm

In order to effectively handle outliers in intensity, we apply a robust estimator $\rho(u)$ to the image differences u in Eqn. 2. Thus, we derive the energy function for the robust deformable Lucas-Kanade algorithm as follows:

$$E(\mathbf{s}) = \sum_{\mathbf{x}} \rho \Big(\big(T - I(\mathbf{s}) \big)^2 \Big) + \lambda_{\mathbf{s}} \mathbf{s}^\top K \mathbf{s} \quad (5)$$

where $\rho(u)$ can be chosen among a variety of robust estimators, such as the German-McClure function [2], exponential function [7], and Huber loss function [13]. To facilitate the first order Taylor expansion [2], we choose the modified Huber loss function:

$$\rho(t) = \begin{cases} t & 0 \le t \le \sigma^2 \\ \sigma(2\sqrt{t} - \sigma) & t > \sigma^2 \end{cases} \quad (6)$$

where the threshold parameter σ determines the switch from quadratic to linear.

As described in [2], we perform a Taylor expansion of the energy function in Eqn. (5). Let ρ' denote the first order derivative of the robust estimator, we can obtain the following update equation:

$$\Delta\mathbf{s} = H_R^{-1} \left(\sum_{\mathbf{x}} \rho'(u^2) A^\top \big(I(\mathbf{s}) - T \big) - \lambda_{\mathbf{s}} K \mathbf{s} \right) \quad (7)$$

Then, the Hessian matrix H_R for the robust deformable Lucas-Kanade algorithm is

$$H_R = \sum_{\mathbf{x}} \rho'(u^2) A^\top A + \lambda_{\mathbf{s}} K \quad (8)$$

3.2.2 Lighting Variations

Global lighting variations are an important issue for faces captured in an uncontrolled environment. We employ a global gain and bias transformation $l(I)$ of the image to account for uniform lighting changes:

$$l(I) = a \cdot I + b \cdot \mathbf{1}$$

where a is the gain factor, and b is the bias. We put the gain and bias into a vector $\mathbf{g} = \left[\begin{array}{cc} a & b \end{array}\right]^\top$. In our optimization functional, we add an extra term to regularize the lighting parameters, and arrive at the following energy function:

$$E(\mathbf{s}) = \sum_{\mathbf{x}} \Big(T - l\big(I(\mathbf{s})\big)\Big)^2 + \lambda_{\mathbf{s}} \mathbf{s}^\top K \mathbf{s} + \lambda_{\mathbf{g}} \|\mathbf{g} - \bar{\mathbf{g}}\| \quad (9)$$

where $\lambda_{\mathbf{g}}$ is a regularization coefficient, and the expectation $\bar{\mathbf{g}} = \left[\begin{array}{cc} 1 & 0 \end{array}\right]^\top$. The above optimization problem can be efficiently solved by a dual inverse compositional algorithm [3], which takes advantage of a constant Hessian matrix. With the step descent matrix for lighting parameters $A_{\mathbf{g}} = \left[\begin{array}{cc} T & \mathbf{1} \end{array}\right]^\top$, the update function becomes

$$\Delta \mathbf{s} = H_L^{-1} \left(\sum_{\mathbf{x}} \left[\begin{array}{c} A \\ A_{\mathbf{g}} \end{array}\right]^\top (I(\mathbf{s}) - T) - \lambda_{\mathbf{s}} K \mathbf{s} - \lambda_{\mathbf{g}} (\mathbf{g} - \bar{\mathbf{g}}) \right)$$

where the Hessian matrix accounting for lighting variations is computed as

$$H_L = \sum_{\mathbf{x}} \left[\begin{array}{c} A \\ A_{\mathbf{g}} \end{array}\right]^\top \left[\begin{array}{c} A \\ A_{\mathbf{g}} \end{array}\right] + \left[\begin{array}{cc} \lambda_{\mathbf{s}} K & 0 \\ 0 & \lambda_{\mathbf{g}} \cdot \mathbf{1} \end{array}\right] \quad (10)$$

Note that we also impose a robust estimator on the lighting models.

3.3. Joint Face Alignment

Face alignment with a single template image may suffer from appearance variations across different individuals with various poses and expressions. To increase the chances of homing in on the solution, we introduce multiple template images into the energy function. These jointly drive forward the evolution of the mesh on the input – i.e. target – face. Therefore, we have the following energy function for joint face alignment:

$$E(\mathbf{s}) = \sum_i \sum_{\mathbf{x}} \rho\left(\big(T_i - I(\mathbf{s})\big)^2\right) + \lambda_{\mathbf{s}} \mathbf{s}^\top K \mathbf{s} \quad (11)$$

In general, there are two ways to expand the above equation and solve the corresponding optimization problem. One strategy is simply to treat the image averaged over all templates as a single template, and then solve the optimization problem by the previous image-to-image alignment method. This approach is an efficient and straightforward implementation. Unfortunately, it smoothes out the texture variations and therefore may lead to poor results. An alternative and more effective strategy is to employ the steepest descent matrices computed from each of the template images [1, 7]. This strategy no longer loses texture detail and has been applied in the experiments. Similar to [1, 7] we obtain the following update equation:

$$\Delta \mathbf{s} = H_M^{-1} \left(\sum_i \sum_{\mathbf{x}} \rho'(u^2) A_i^\top \big(I(\mathbf{s}) - T_i\big) - \lambda_{\mathbf{s}} K \mathbf{s} \right) \quad (12)$$

where $A_i = \nabla T_i \frac{\partial W}{\partial \mathbf{s}}$ is the steepest descent image, and $H_M \in \mathbb{R}^{2N \times 2N}$ is the Hessian matrix that is computed as follows:

$$H_M = \sum_i \sum_{\mathbf{x}} \rho'(u^2) A_i^\top A_i + \lambda_{\mathbf{s}} K \quad (13)$$

Remark. Regarding the issue of **building the template images**, we use a bootstrap procedure. We start from a single template image with a frontal face, where we build the reference mesh based on detected eye corners and the distance between two eyes. For a first set of additional images, the mesh is aligned with the image-to-image face alignment. From then on, the template images are formed by the aligned images and corresponding meshes with the lowest sum of squared texture differences.

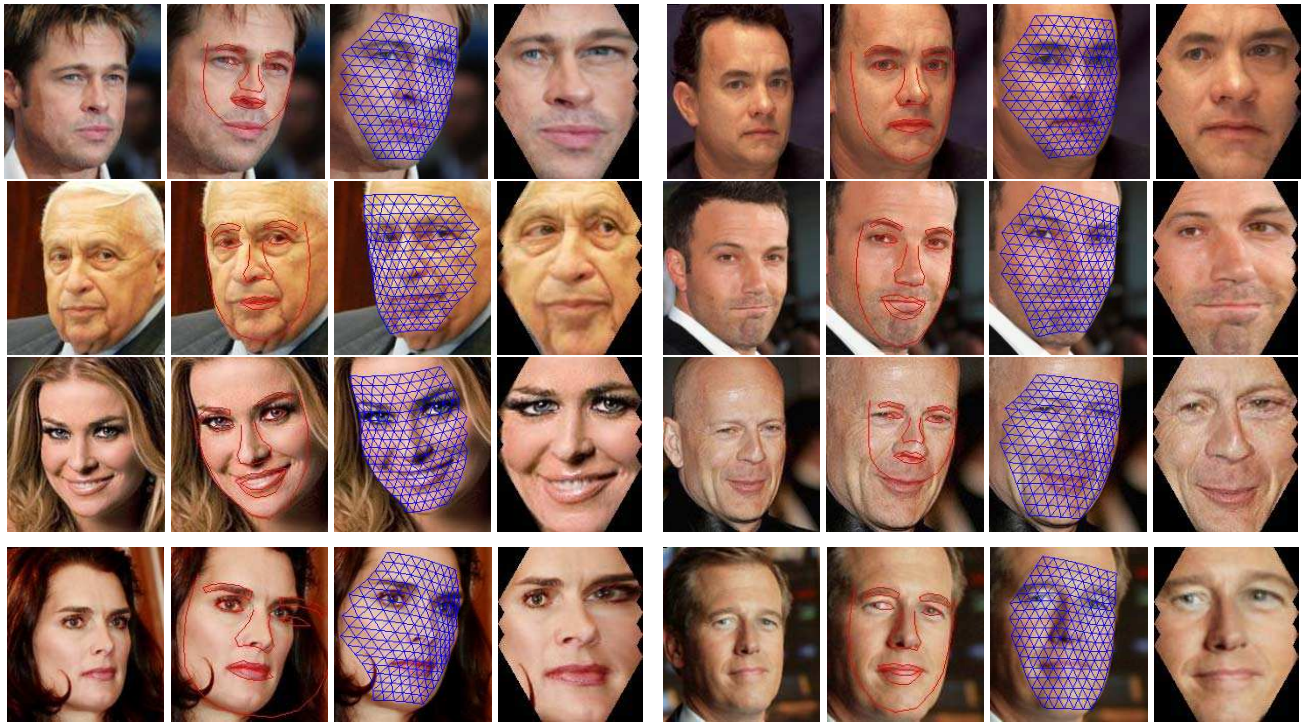
4. Experimental Results

In this section, we give details of our experimental implementation and discuss the results of face alignment performance tests. We examine how effective the proposed approach is in aligning face images that were obtained via internet crawling. These have often been captured in uncontrolled environments. In addition, we demonstrate promising results when applying our technique for face swapping, which is a key component for a face de-recognition system.

4.1. Experimental Testbed

To conduct comprehensive evaluations, we have collected face images belonging to 101 different individuals [1] using Google image search. To facilitate an automatic approach, we cascade a face detector [22] with a facial feature detector [8] to locate faces and eye corners in the input images. We remove face images of very small sizes, and eliminate duplicate face images by comparing grid color moments for each pair of images. Finally, we formed our testbed of 3467 images in total, in which each individual has $17 \sim 62$ images. Besides the collected face images, we also evaluated our method on the LFW dataset [12].

[1] http://www.vision.ee.ethz.ch/ zhuji/facealign



(a) Congealing (b) CMU's method (c) JDLK (d) Rectified face (e) Congealing (f) CMU's method (g) JDLK (h) Rectified face

Figure 3. Comparison of face alignment for various head poses. We show the results of the congealing algorithm, CMU's online system and the proposed joint deformable Lucas-Kanade approach respectively. Also, we rectify the faces on canonical view by recovered meshes.

4.2. Comparison Schemes and Setup

For simplicity, our proposed image-to-image face alignment method is denoted by "DLK" for short, and the proposed joint face alignment approach is denoted by "JDLK", resp. To evaluate them, we conduct an extensive comparison with several state-of-the-art approaches, incl. the congealing algorithm [11], CMU's face alignment method [10], and the direct face alignment method [8].

Specifically, to make a fair comparison, we adopt the authors' own implementation of the congealing algorithm with the default parameter settings. For CMU's method, as no public code was available, we compare the same set of images with the results returned from CMU's online face alignment system [2], which is based on a state-of-the-art supervised face alignment method proposed in [10]. Finally, for the direct face alignment method [8], we simply align the face images to the predefined template using the detected eye corners from the facial feature detector [8].

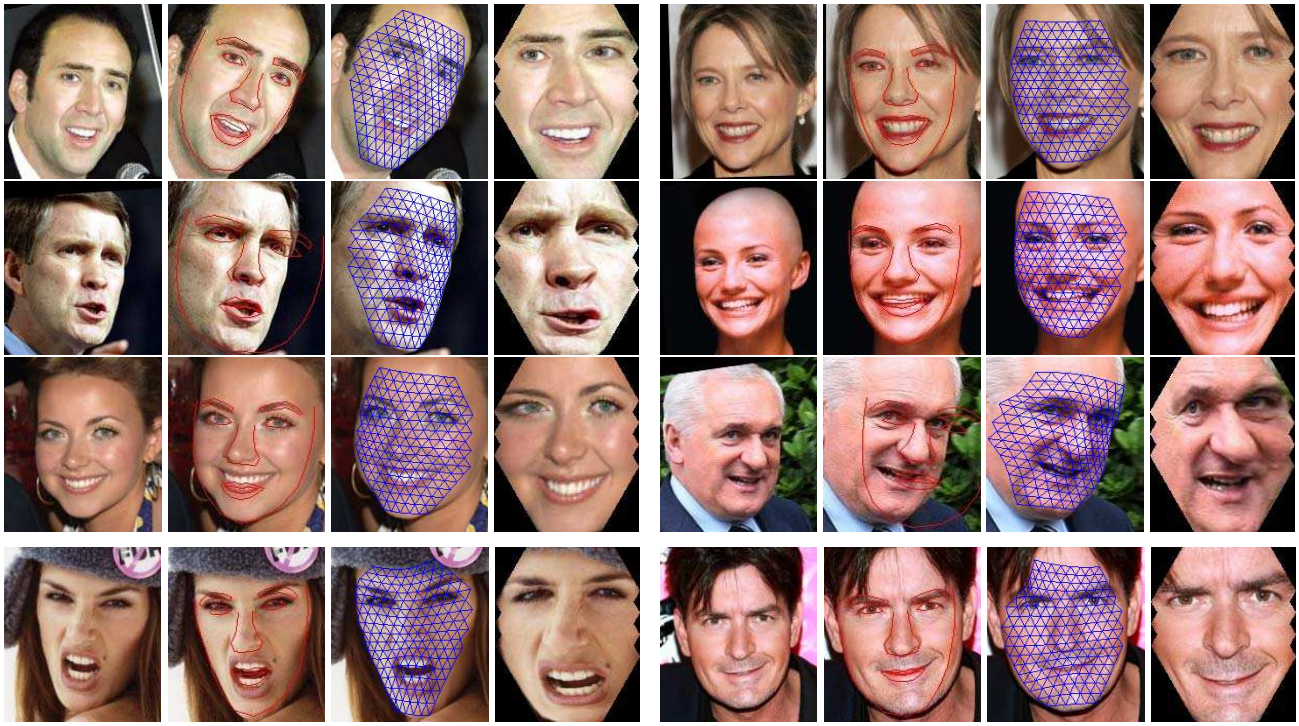
For the proposed joint image alignment approach, we employ 20 templates, which are built using the method described in Section 3.3. The regularization coefficient $\lambda_{\mathbf{s}}$ is empirically set to 10^6 for DLK and 10^7 for JDLK, and $\lambda_{\mathbf{g}}$ is set to 10^3. With a software rendering engine, the proposed DLK method is able to process around 6 images per second, and 0.5 images per second for JDLK. All of our experiments were carried out on a PC with Intel Core-2 Duo 2.0GHz processor and 2GB RAM.

[2] http://facealignment.ius.cs.cmu.edu/alignment/webdemo.html

4.3. Evaluation on Face Alignment

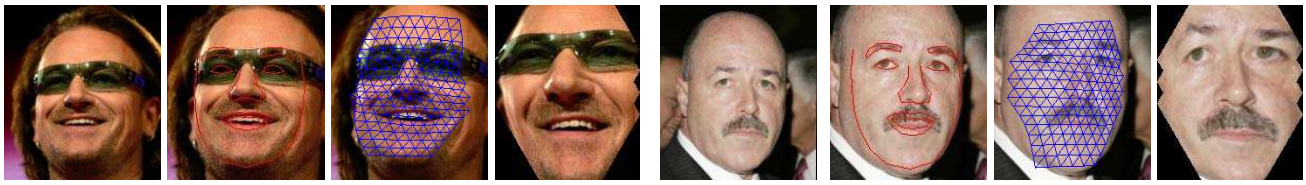
We now evaluate the performance of face alignment under challenging conditions, including pose variations, various expressions, with glasses and beards.

Pose Variations. In Fig. 3(c) we overlaid the estimated mesh onto the slightly zoomed in faces as they appear in the original images (for reasons of clarity), which demonstrates the robustness of the proposed method in handling various poses, such as out-of-plane rotations, pitch, and yaw. The resulting mesh tends to deform to the geometry of the target face. Also, we plot the rectified faces by rendering the input face image onto the reference frame using the recovered mesh. Every face component comes out to be accurately registered in the canonical view, which is crucial for a face recognition task. The congealing algorithm can roughly correct the in-plane rotations of face images. However, it performs quite poorly on estimating the facial region size of these non-frontal images. This is mainly due to its rigid transformation assumption that does not account for the out-of-plane rotations. CMU's face alignment system was mainly trained from frontal face images, which leads



(a) Congealing (b) CMU's method (c) JDLK (d) Rectified face (e) Congealing (f) CMU's method (g) JDLK (h) Rectified face

Figure 4. Comparison on face images with various expressions. We show the results of the congealing algorithm, CMU's online system, and the proposed joint deformable Lucas-Kanade approach.



(a) Congealing (b) CMU's method (c) JDLK (d) Rectified face (e) Congealing (f) CMU's method (g) JDLK (h) Rectified face

Figure 5. Comparison on face images with glasses and moustache.

to some poor results on non-frontal faces as shown in the figures. Despite the promising results shown in Fig. 3, our method still has some limitations. For example, when the facial feature locator [8] fails for faces with heavily eye-occluded eyes, then the poor initialization may lead further steps of the method astray. Besides, in some cases of faces with large out-of-plane rotation, large artifacts were observed in the rectified images due to the lacking of appropriate textures.

Expressions. We also evaluated our methods for face images with various expressions. As shown in Fig. 4, both the congealing algorithm and the proposed approach are robust to such expressions. Since it is hard to model the large variations of mouth shapes through a statistical model, large errors occur in the mouth regions in CMU's method.

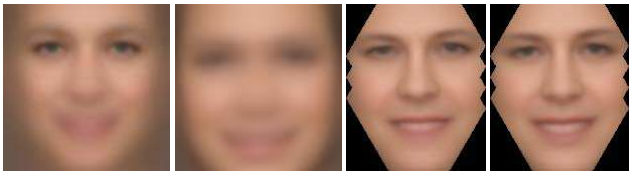
Glasses and Beards. We also consider faces with glasses, moustaches and beards, as shown in Fig. 5. The proposed method can handle these partially occluded faces. The robust estimator component has proven to be of key importance here.

Finally, Fig. 6 shows the mean faces, averaged over all aligned facial images. From the results, we found that the mean faces using both image-to-image alignment and joint face alignment are generally much clearer than those of the other two approaches. This means that the proposed methods can more effectively remove the spatial variations across pose variations and individuals. On the other hand, we found that the eyes are accurately aligned in the direct alignment method, while for the congealing algorithm, large spatial variations still exist in the alignment results.

4.4. Evaluation on Face Recognition Performance

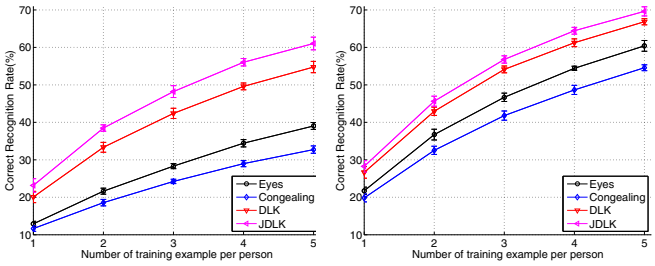
4.4.1 Google Image Dataset

To further examine the performance of the proposed unsupervised face alignment techniques, we compare our method with the congealing algorithm and the direct align-



(a) Align by eyes (b) Congealing (c) DLK (d) JDLK

Figure 6. Average of aligned facial images.



(a) Intensity (b) Gabor wavelets feature

Figure 7. Correct recognition rate on face images from the Internet (3467 images from 101 persons).

ment by eye corners for face recognition tasks. We partition the collected face dataset into a training set and a test set.

The number of labeled examples for each person is gradually increased from 1 to 5. A variation of 10-fold cross validation was performed in the experiments, in which the labeled data were randomly selected. We adopt the correct recognition rate as the performance metric, and employ a linear Support Vector Machine (SVM) as the classifier. The penalty parameter C for SVMs is set to 10 for all experiments. Besides the evaluation on the raw intensity, we also extract Gabor wavelet features [16] for evaluation.

Fig. 7 summarizes the experimental results of recognition rates under different settings. We can make several observations from the results. First of all, the proposed two face alignment methods, DLK and JDLK, significantly outperform the congealing algorithm and the direct alignment by eye corners. Specifically, the correct recognition rate achieved by the proposed JDLK is almost double of those achieved by the two reference methods when using one training example for each person. In addition, we found that the recognition performance of all methods is greatly improved by using Gabor wavelet features. They boost the performance around 22% over the raw intensity for the proposed JDLK method on the tough task with a single training example per person. Compared to our technique, it is difficult for some patch-based methods, such as [1], to take advantage of such effective features. Finally, we found that the joint image alignment method, JDLK, achieves considerably better results than DLK, the image-to-image alignment approach. This indicates that the alignment accuracy can be effectively improved by exploring more information from multiple template images.

4.4.2 LFW Dataset

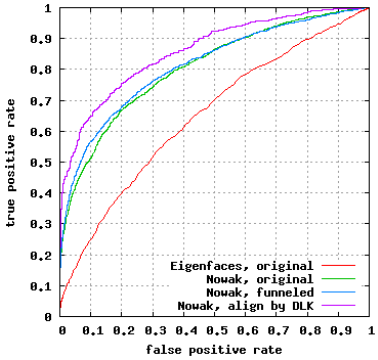


Figure 8. ROC curves on LFW dataset.

We also evaluate the proposed method on the LFW dataset [12] using image-restricted training. To facilitate the mutually exclusive requirement, we only evaluate the proposed DLK method. Since we concentrate on face alignment rather than the pairwise image matching problem, we only compare with the reported results using the congealing method [11]. Similarly, we directly feed the rectified images into the binary version of Nowak's program [20] with the same settings. Fig. 8 shows the ROC curves with several baselines [12]. Although Nowak's method has already handled the misalignment issue by sampling corresponding patches, the rectified images by the proposed DLK can further improve the performance considerably.

4.5. Application to Face Swapping

As an additional application, the alignment techniques are used to swap faces, i.e. to overlay the inner part of one face onto the image of another. This is relevant for movies, where an actor's face should replace that of a stuntman. Also, as a protection of privacy, people's faces could be replaced by some standard, average face (think of Google's problems with privacy and the visually disturbing smoothing of faces applied now). Fig. 9 shows two examples of face swapping.

Comparing with recent work in computer graphics [4], our approach more fully exploits the appearance information by a novel nonrigid image mapping approach. Our technique is able to offer more accurate registration results and flexility in selecting the candidates than the direct face alignment method using a mere five facial points, proposed in [4].

5. Conclusions

This paper proposed a novel unsupervised face alignment method, which is different from previous unsupervised alignment approaches, e.g. in not assuming a rigid affine transformation for alignment. The proposed deformable Lucas-Kanade algorithm registers the face images

(a) Original (b) Example 1 (c) Example 2

Figure 9. Examples of automatic face swapping.

across different individuals, pose variations, and various expressions. Moreover, we present a joint image alignment method that incorporates information from multiple templates. We have conducted extensive evaluations on a face image dataset collected from the Internet. The encouraging experimental results showed that our method performs better than alternative approaches, especially on difficult cases.

Despite these promising results, some limitations and future directions should be addressed. Currently, our method was only tested on face images that have similar appearances. Besides, we have yet to carefully address the lighting issue in the joint face alignment scheme. For future work, we will address these issues and extend our technique to other objects by adopting appropriate metrics, such as mutual information.

Acknowledgments

The work was fully supported by the EC project Hermes (Cognitive systems unit). The authors also thank the reviewers for their helpful comments.

References

[1] A. B. Ashraf, S. Lucey, and T. Chen. Learning patch correspondences for improved viewpoint invariant face recognition. In *CVPR*, June 2008.

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework: Part 1, 2&3. Technical report, Robotics Institute, February 2004.

[3] A. Bartoli. Groupwise geometric and photometric direct image registration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(12):2098–2108, 2008.

[4] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar. Face swapping: automatically replacing faces in photographs. *ACM Trans. Graph.*, 27(3):1–8, 2008.

[5] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[6] T. Cootes, G. Edwards, and C. Taylo. Active appearance models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):681–685, June 2001.

[7] M. Cox, S. Lucey, S. Sridharan, and J. Cohn. Least squares congealing for unsupervised alignment of images. In *CVPR*, June 2008.

[8] M. Everingham, J. Sivic, and A. Zisserman. Hello! my name is... buffy – automatic naming of characters in tv video. In *BMVC*, 2006.

[9] B. J. Frey and N. Jojic. Transformed component analysis: Joint estimation of spatial transformations and image components. In *in ICCV*, pages 1190–1196, 1999.

[10] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In *ECCV*, pages 413–426, 2008.

[11] G. Huang, V. Jain, and E. Learned Miller. Unsupervised joint alignment of complex images. In *Proc. Int'l Conf. Computer Vision*, 2007.

[12] G. B. Huang, M. Ramesh, T. Berg, , and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, October 2007.

[13] P. J. Huber. Robust statistics: a review. *Annals of Statistics*, 43:1041, 1972.

[14] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach toward pose-invariant face recognition. In *CIRA'2003*, pages 954 – 959, July 2003.

[15] E. G. Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(2):236–250, 2006.

[16] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fld model for face recognition. *IEEE Trans. Image Processing*, 11:467–476, 2002.

[17] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81*, pages 674–679, 1981.

[18] S. Lucey and T. Chen. A viewpoint invariant, sparsely registered, patch based, face verifier. *Int. J. Comput. Vision*, 80(1):58–71, 2008.

[19] I. Matthews and S. Baker. *Active Appearance Models Revisited*, 60(1):135 – 164, November 2004.

[20] E. Nowak and F. Jurie. Learning visual similarity measures for comparing never seen objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, jun 2007.

[21] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration, and realistic augmentation. *Int'l J. Computer Vision*, 76(2):109–122, 2008.

[22] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, 2004.

[23] J. Zhu, S. C. Hoi, and M. R. Lyu. Real-time non-rigid shape recovery via active appearance models for augmented reality. In *Proc. European Conf. Computer Vision*, pages 186–197, 2006.

[24] J. Zhu and M. R. Lyu. Progressive finite newton approach to real-time nonrigid surface detection. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2007.

[25] J. Zhu, M. R. Lyu, and T. S. Huang. A fast 2d shape recovery approach by fusing features and appearance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(7):1210–1224, 2009.