Non-Rigid Object Alignment with a Mismatch Template Based on Exhaustive Local Search

Yang Wang, Simon Lucey, Jeffrey Cohn
The Robotics Institute, Carnegie Mellon University
Pittsburgh, PA 15213, USA
{wangy,slucey,jeffcohn}@cs.cmu.edu

Abstract

Non-rigid object alignment is especially challenging when only a single appearance template is available and target and template images fail to match. Two sources of discrepancy between target and template are changes in illumination and non-rigid motion. Because most existing methods rely on a holistic representation for the alignment process, they require multiple training images to capture appearance variance. We developed a patch-based method that requires only a single appearance template of the object. Specifically, we fit the patch-based face model to an unseen image using an exhaustive local search and constrain the local warp updates within a global warping space. Our approach is not limited to intensity values or gradients, and therefore offers a natural framework to integrate multiple local features, such as filter responses, to increase robustness to large initialization error, illumination changes and non-rigid deformations. This approach was evaluated experimentally on more than 100 subjects for multiple illumination conditions and facial expressions. In all the experiments, our patch-based method outperforms the holistic gradient descent method in terms of accuracy and robustness of feature alignment and image registration.

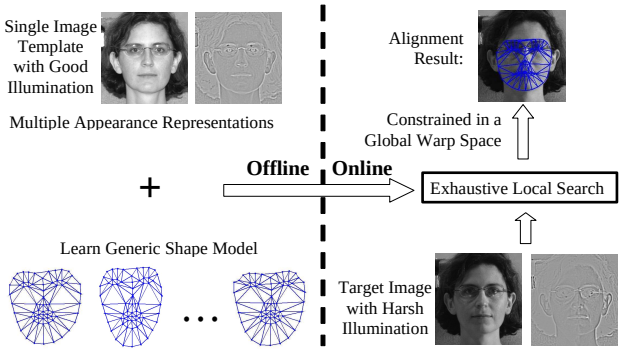


Figure 1. Illustration of the problem studied in this paper: to align a non-rigid object using a mismatched appearance template. The proposed method uses a generic (i.e. subject independent) shape model and a single appearance template of the target subject, while the test image is taken under different illumination conditions, e.g., a side-lit image. The image fitting process uses an exhaustive local search and constrains all local warp updates within a global warping space. Furthermore, our method is able to handle multiple representations and therefore offers a natural framework to integrate local features, such as filter responses, to improve the accuracy and robustness of the alignment in presence of large initialization errors, changing illumination conditions and non-rigid deformations.

1. Introduction

The accurate registration of objects is important in many computer vision applications, such as feature detection, object tracking and recognition. Despite its difficulty, great progress has been made in the last couple of decades [1, 3, 20, 19, 5, 11, 15, 8, 9]. In literature, most existing methods employ a generative model combined with the image templates of the object appearance, such as Active Appearance Model (AAM) [1, 3, 19]. By synthesizing a good appearance template of the target object, one can achieve high registration accuracy using generative appearance models. However, their performance will decrease when the object has a large appearance variation from the training set, such as the changes caused by different illumination conditions and non-rigid motion. To address this problem, many attempts have been made to handle appearance variation [2, 12, 4, 1]. These methods, however, require multiple training images to capture the appearance variance. In this paper, we present a new approach to align unseen images based on an exhaustive local search, which uses only a single appearance template of the target object as illustrated in Fig.1. Instead of relying on a holistic representation for the alignment process, our method fits the model to an unseen image based on an exhaus-

tive local search and constrains all the local warp updates within a global warping space. Furthermore, our method is able to handle multiple representations and therefore offers a natural framework to integrate local features, such as filter responses, to improve the robustness to large initialization errors, changing illumination conditions and non-rigid deformations. The performance of our framework is demonstrated through various experimental results, including the improvement to the accuracy and robustness of feature alignment and image registration.

Our approach is in principle similar to previous patch-based methods, such as Constrained Local Model (CLM) [5] and 3D based face alignment [11], but differs in a number of aspects: (1) Our method uses only a single appearance template of the target subject, which does not match the source images to be aligned because of the changes stemming from different illumination conditions and non-rigid motion; (2) Multiple or alternative local feature representations are incorporated in our proposed framework to improve the accuracy and robustness of object alignment; (3) We expand previous work to study the differences between holistic gradient descent methods and patch-based correlation methods and attempt to explain why exhaustive local search can achieve better performance.

2. Background

In the section, we will briefly review two important techniques used in image alignment and registration, i.e., the gradient descent alignment and the generative approach. After that, we will discuss the differences between the holistic gradient descent and the patch-based correlation techniques in the next section, which will lead to our proposed approach based on an exhaustive local search.

2.1. Gradient Descent Alignment

Derived from the Lucas-Kanade algorithm [14], many methods have been proposed to do image alignment based on the gradient descent technique [1, 3]. In this section, we briefly review the gradient descent alignment approach for clarity and notation. Given a template $T(\mathbf{z})$ and a source image $Y(\mathbf{z}')$, we attempt to find the best alignment between them, where $\mathbf{z}'$ is the warped image pixel position of $\mathbf{z}$, i.e.,

$$\mathbf{z}' = \mathcal{W}(\mathbf{z}; \mathbf{p}), \tag{1}$$

with warp parameters $\mathbf{p}$. For clarity purposes, $\mathbf{z}$ is represented in a concatenation of individual pixel 2D coordinates $\mathbf{x}_i = (x_i, y_i)$, i.e.,

$$\mathbf{z} = [x_1, y_1, \ldots, x_N, y_N]^T \tag{2}$$

and similarly

$$T(\mathbf{z}) = [T(\mathbf{x}_1), \ldots, T(\mathbf{x}_N)]^T \tag{3}$$

For our work, we defined the warp function $\mathcal{W}(\mathbf{z}; \mathbf{p})$ as

$$\mathcal{W}(\mathbf{z}; \mathbf{p}) = \mathbf{J}\mathbf{p} + \mathbf{z}_0 \tag{4}$$

We refer to this linear model as a point distribution model (PDM)[3], where procrustes analysis is applied to all shape training observations in order to estimate a similarity normalized base shape template $\mathbf{z}_0$[3, 16]. Principle component analysis (PCA) was then employed to obtain shape eigenvectors $\mathbf{J}$ that preserved 95% of the similarity normalized shape variation in the train set. The first 4 eigenvectors of $\mathbf{J}$ were forced to correspond to similarity (i.e., translation, scale and rotation) variation.

The outline of the canonical gradient-descent fitting algorithm is listed as follows:

1. Warp the source image Y with $\mathbf{z}' = \mathcal{W}(\mathbf{z}; \mathbf{p})$ to get the $N \times 1$ vector $Y(\mathbf{z}')$
2. Compute the error image vector $Y(\mathbf{z}') - T(\mathbf{z})$
3. Estimate the warp update,
$$\triangle\mathbf{p} = \mathbf{R}[Y(\mathbf{z}') - T(\mathbf{z})] \tag{5}$$
where $\mathbf{R}$ is the $P \times N$ update matrix, P is the number of warp parameters and N is the number of pixels.
4. Update the warp, in particular, using the *inverse compositional update*,
$$\mathbf{z}' = \mathcal{W}(\mathbf{z}; \mathbf{p}) \leftarrow \mathcal{W}(\mathbf{z}; \mathbf{p}) \circ \mathcal{W}(\mathbf{z}; \triangle\mathbf{p})^{-1} \tag{6}$$
5. Iterate the above steps $1 - 4$, until it converges, i.e., $|\triangle\mathbf{p}| \leq \epsilon$.

2.2. Generative Approach

In order to constrain the local gradient to model a holistic warp update $\triangle\mathbf{p}$, a common approach is to form a linear generative model of the warp variation and then solve for $\triangle\mathbf{p}$ using a least-squares criteria [1]. More specifically, if we apply a first order Taylor series expansion on $T(\mathbf{z})$, we will have

$$Y(\mathbf{z}') \approx T(\mathbf{z}) + \mathbf{K}^T \triangle\mathbf{p} \tag{7}$$

and

$$\mathbf{K} = \frac{\partial \mathcal{W}(\mathbf{z}; 0)}{\partial \mathbf{p}} \frac{\partial T(\mathbf{z})}{\partial \mathbf{z}} \tag{8}$$

Therefore, we can solve for $\triangle\mathbf{p}$ using the following equation:

$$\triangle\mathbf{p} = (\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{K}[Y(\mathbf{z}') - T(\mathbf{z})] \tag{9}$$

which leads to the final update matrix $\mathbf{R}$ in Eqn.(5):

$$\mathbf{R} = \mathbf{K}^+ \leftarrow (\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{K} \tag{10}$$

3. Limitations of Holistic Gradient Descent

Although holistic gradient descent methods can achieve good performance on image alignment [1, 3], it is difficult to handle unseen appearance variation, such as the changes caused by different illumination conditions and non-rigid motion. Instead, working on the patch level will offer us more flexibilities on the local regions [5, 11]: 1) Since lighting in smaller image regions is more homogeneous, we can expect the overall estimation error to be smaller than a single holistic approximation; 2) The variance of texture within a patch is considerably smaller than that of the whole face; and 3) Because the holistic warp often employs a complicated piece-wise affine warp based on pixels[1], the appearance variance can not be separated from the shape variance and the shape alignment error will introduce non-linear noise to the appearance estimation. As a result, their methods suffer from unseen appearance variation, such as the changes caused by different illumination conditions and non-rigid deformations.

An example is shown in Fig.2 to illustrate the differences between the holistic and patch-based methods for non-rigid image alignment. As we can see, the patch-based approach has an inherent advantage over the holistic warp in that it only performs a similarity transform on the source image from which patch regions are then extracted. Even when noise is present in the alignment the patch-based warped image still looks like a face. The holistic warp, however, employs a complicated piece-wise affine warp based on the pixels which can lead to strange looking images when noise is present in the alignment. In all our alignment experiments we found a patch-based warp to outperform a holistic warp.

Furthermore, another major limitation of gradient descent methods is that it is difficult to integrate multiple representations, such as edge information which however is an important cue to align images[4, 17]. A toy example of $1D$ gradient descent alignment is illustrated in Fig.3, where Fig.3(a) shows the source 1D signal, and Fig.3(b) the computed 1D gradient. Fig.3(c) is the template 1D signal, after shifting the second signal in Fig.3(a) to the right. The estimated warp update using gradient descent, Eqn.(5), is shown in Fig.3(d). As we can see, because there are no gradient values at the edge positions in both the source and template signals, the warp update $\triangle \mathbf{p} = 0$ from Eqn.(5) and the model will not move towards the right direction. However, if we use the correlation based exhaustive search we can find the local update correctly.

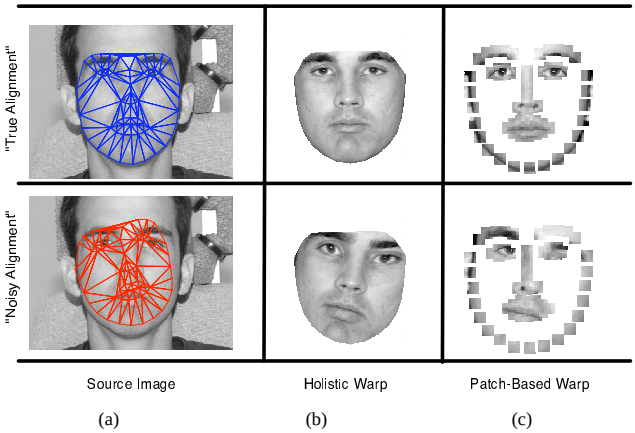


Figure 2. Comparison between holistic and patch-based warp alignment methods: The source images with different alignment are shown in column (a). The resulting warps from the holistic and patch-based methods are shown in column (b) and (c), respectively. The top row shows the ground truth alignment, where both the holistic and patch-based warps give faithful approximation of the original image. However, when the alignment is perturbed by noise, as shown in the bottom row, the holistic warp gives a strange result (b), which is not close to the original image. The patch-based warp, which performs only the inverse similarity transform, generates a reasonable result (c).

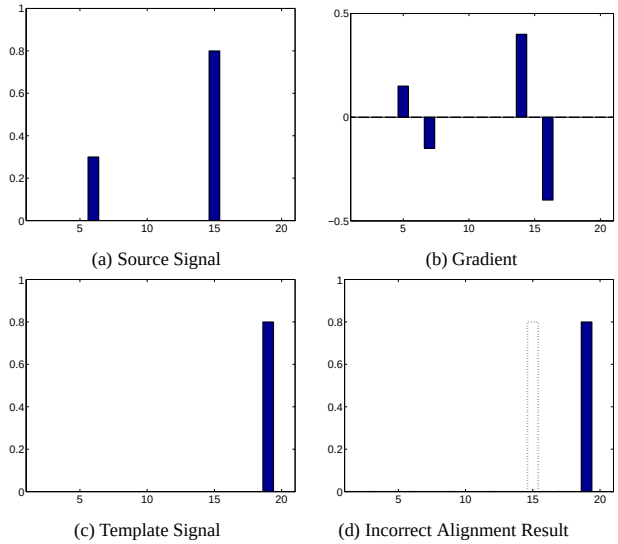


Figure 3. A toy example of $1D$ gradient descent alignment: (a) The source 1D signal. (b) The computed 1D gradient. (c) The template 1D signal, after shifting the second signal in (a) to the right. (d) The result using gradient descent. It is misaligned with the ground truth data shown in the dashed bar. Because there are no gradient values at all the edge locations, the estimated local displacement is 0 from Eqn.(5), while it can be recovered correctly by local search using signal correlation.

4. Proposed Method

In this section, we propose a new approach to achieve robust image alignment using a patch-based exhaustive local search. Rather than generating the holistic appearance template of the target object, we find the optimal local updates to fit an unseen image using an exhaustive local search and constrain them within a global warping space. By doing so, we obviate the computation of the image gradients,

which are sensitive to image noise and discontinuities, and can integrate multiple cues, e.g. local features, which are important to object alignment.

Patch-based methods have been previously used in image alignment, such as Constrained Local Models (CLM) [5] and 3D face alignment [11]. However, to the authors' best knowledge there has not been a complete study to compare the differences between the holistic and the patched-based warp update methods. Furthermore, different from previous methods [5, 11] our proposed framework allows multiple feature representations, such as edge information, which are less sensitive to lighting conditions than raw intensity [4, 17].

4.1. Patch-Based Approach

As described in Section 3, there are certain limitations associated with optimizing for the holistic warp update $\triangle\mathbf{p}$ directly. Instead, we propose an approach that exhaustively searches within N local neighborhoods to find the N best translation updates $\triangle\mathbf{x}$, and the constrain all the local updates using the Jacobian matrix

$$\mathbf{J} = \frac{\partial \mathcal{W}(\mathbf{z}; 0)}{\partial \mathbf{p}} \tag{11}$$

Following the same convention as in Active Appearance Model (AAM) [1, 3], the 2D shape of a face is defined by a set of 2D points, $\mathbf{x}_i = (x_i, y_i)$, concatenated into a vector $\mathbf{z} = [x_1, y_1, \ldots, x_n, y_n]^T$. Once we obtain the local pixel updates $\triangle\mathbf{z}$, the global warp update can be estimated by a weighted least-squares optimization,

$$\triangle\mathbf{p} = (\mathbf{J}\mathbf{W}\mathbf{J}^T)^{-1}\mathbf{J}\mathbf{W}\triangle\mathbf{z} \tag{12}$$

where the weighting matrix $\mathbf{W}$ is defined as a diagonal matrix,

$$\mathbf{W} = diag\{w_{x_1}, w_{y_1}, \ldots, w_{x_n}, w_{y_n}\} \tag{13}$$

In order to find the optimal local update, we can perform an exhaustive normalized cross-correlation search within a local neighboring region[1]. However, because the patch-wise correlation is sensitive to the image scale and rotation, it is important to normalize the image to be aligned. For these purposes, we remove the similarity transformation between the source and the template images.

4.2. Local Feature Representation

Although image intensities have been used directly in most existing methods to build the appearance model [1, 3, 5, 11], they might suffer from changes in illumination conditions and non-rigid deformations. Instead, a more robust representation can be achieved through the employment of a set of local kernel functions, such as using Gabor wavelets [13, 10], Gaussian filters and the derivatives [7], and Filtered Component Analysis (FCA) [6]. Furthermore, Cootes and Taylor [4] found that edge structure can be used to improve model matching. This was additionally confirmed by Stegmann and Larsen [17] who demonstrated that hue and edges have good performance on localizing faces.

One advantage of our proposed method is that it offers a natural framework to integrate multiple feature representations, such as edges, which can aid in gaining invariance to unseen variation (e.g. illumination) [4]. In our implementation, we tried both Laplacian and Gabor filters to extract local features.

4.3. Algorithm Outline

Our proposed approach for non-rigid face image alignment includes the following steps:

1. *Initialization:* Given a source image, we obtain an estimate of the similarity transformation (i.e. the scale, translation and rotation) between the source image and our template. A common method for obtaining this initial estimate is through an exhaustive search rigid face detector such as the one proposed by Viola and Jones [18]. After that, the mean face shape $\mathbf{z_{init}}$ with the initial similarity transformation is defined as the initial face shape.

2. *Similarity Normalization:* To remove the similarity transformation between the source image, Y, and template images, T, by applying the inverse estimated similarity transformation on the input image.

3. *Local Warp Update Search:* For each patch template at each mesh feature point $\mathbf{x}_i = (x_i, y_i)$, find the best match at the location $\mathbf{x}'_i = (x'_i, y'_i)$ using the normalized cross-correlation (NCC). The local patch displacement is computed as $\triangle\mathbf{z} = \mathbf{z}' - \mathbf{z}$.

4. *Compute Global Warp Update:* The global warp parameter update $\triangle\mathbf{p}$ is computed by Eqn.(12).

Steps 2-4 are repeated until convergence.

Although the weighting matrix in Eqn.(12) could be simply an identity matrix, a better choice can be used based on the maximum score of the normalized cross-correlation S_{ncc}. This was done to improve the robustness to the textureless regions and outliers:

$$W_i = \frac{1}{1 + e^{-max_i(S_{ncc})}} \tag{14}$$

5. Experiments

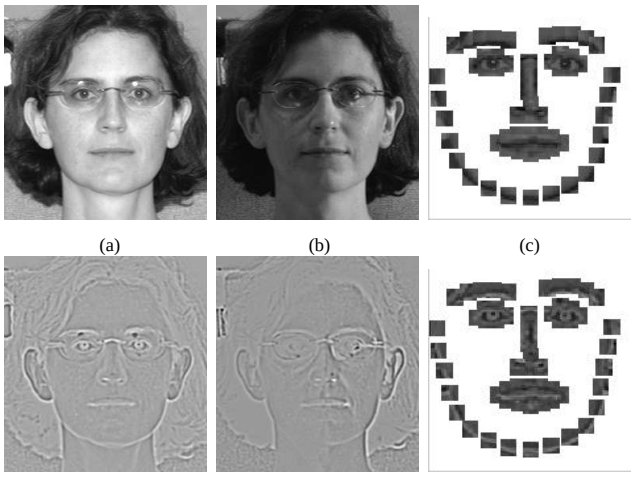
In this section we present the experimental results of three different approaches previously described in this paper, namely: (1) the holistic gradient descent method, (2)

[1] typically a 10×10 window

the patch-based correlation method, and (3) the patch-based correlation method using local filters. In order to compare their performance, we employed a dataset that includes more than 100 subjects with mismatched illumination conditions for training and testing. The size of the template face region is 110×110 pixels with the inter-ocular distance of 45 pixels.

Two main aspects are tested in our experiments: (1) robustness to noise and (2) the ability to handle unseen appearance variation. To test the robustness of each method we randomly initialized the warp parameters in the following manner. We selected the center of the left eye, the center of the right eye and the tip of the nose in the base template, and then perturbed these points with a vector generated from white Gaussian noise. Two different magnitudes of the Gaussian noise are used: (a) 10 pixel and (b) 5 pixel root mean squared point error (RMS-PE) from the ground-truth coordinates. In our experiments, 5 random initial warps, generated by 5 different initial Gaussian noise with the same magnitude, were created for each test image in the evaluation set. In order to test the ability of each method to handle unseen appearance variation, we use a front-lit image of a target subject as the template (Fig.4(a)) and test the method on a side-lit image of the same subject (Fig.4(b)).

All our experiments were conducted using a generic (i.e. subject independent) point distribution model (PDM) estimated from 100 subjects which differed to those used in testing. Some example images in our experiments including their patch representations are shown in Fig.4.



(a) (b) (c) (d) (e) (f)

Figure 4. Example template and test images. (a) A front-lit image as the appearance template. (b) A side-lit image used for testing. (d,e) The filtered versions of (a,b), where a 9×9 Laplacian filter is applied. (c,f) The patch representations of (a,d), respectively. A 9×9 normalized patch is extracted at each feature point of the shape model.

In order to better understand the differences between the holistic and the patched-based warp update methods, we conducted several comparison experiments between the holistic gradient descent method and the patch-based correlation method. For our patch-based method the patch size is fixed to be 9×9 pixels in Fig.5. As we can see the patch-based normalized cross-correlation method outperforms the holistic gradient descent method with different initialization errors. Moreover, the convergence rate of the patch-based normalized cross-correlation method is much higher than the holistic gradient descent method, even if we allow more iterations for the fitting process (Fig.5(c,d)). For a fair comparison we measured alignment error in the following way. Given the ground-truth shape of the source image we apply a similarity transform that minimizes the alignment error with the base template shape. We then apply this same similarity transform to the estimated shape of the source image and compute the RMS-PE between all the mesh model points[2].

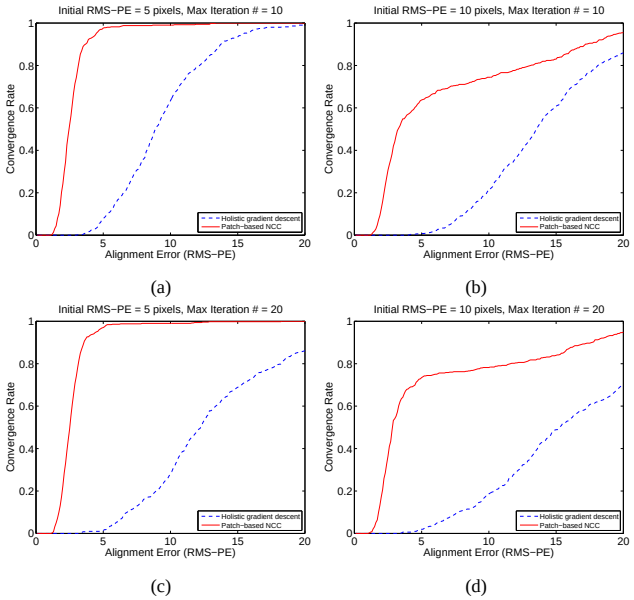


Figure 5. Comparison between holistic and patch-based methods under different illumination conditions. Different initial root mean squared point errors (RMS-PEs) are tested: 5 pixels in (a,c) and 10 pixels in (b,d). The number of iterations for the fitting process is 10 in (a,b) and 20 in (c,d). The size of the face template is 110×110 pixels and the patch size is 9×9 pixels. The horizontal axis shows the resulting alignment error (RMS-PE) in pixels and the vertical axis shows the convergence rate in relation to the alignment error. As we can see, the patch-based normalized cross-correlation method outperforms the holistic gradient descent method for both 5 and 10 pixel initialization errors, even if we allow more iterations for the fitting process as shown in (c,d).

As described in Section 4.2, one advantage of our pro-

[2]In our implementation, we used a 68 point shape model.

posed method is that it offers a natural framework to integrate multiple feature representations. These multiple feature representations are less sensitive to varying illumination conditions than raw image intensity. In our experiments we used different filters, such as Laplacian filters and Gabor filters, to extract local features. In Fig.6 two kinds of Laplacian and Gabor filters are tested: (1) size = 9×9 pixels with $\sigma = 1.5$ and (2) size = 5×5 pixels with $\sigma = 0.7$. As we can see filtered representations yield better performance than using the image-based representation. Also, the comparison between Fig.6(a,b) and Fig.6(c,d) shows that Laplacian filters tend to be more robust to the choice of filter parameters than Gabor filters.

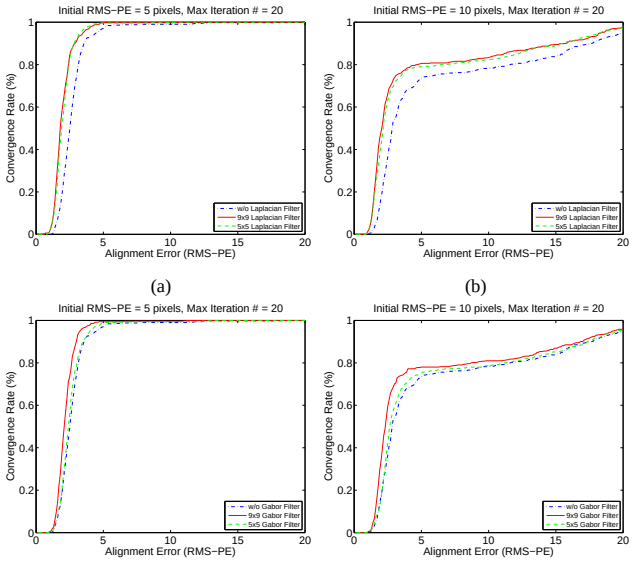


Figure 6. Comparison between the performance of the patch-based correlation method using different filters: (1) Laplacian filters with (a) size of 9×9 pixels with $\sigma = 1.5$ and (b) size = 5×5 pixels with $\sigma = 0.7$; (2) Gabor filters with (c) size of 9×9 pixels with $\sigma = 1.5$ and (d) size = 5×5 pixels with $\sigma = 0.7$. Different initial root mean squared point errors (RMS-PEs) are tested: 5 pixels in (a,c) and 10 pixels in (b,d). The number of iterations for each image is 20. The size of the face template is 110×110 pixels and the patch size is 9×9 pixels. The same horizontal and vertical axes are used as in Fig.5. We can see that filtered representations yield better performance than using the image-based representation. Also, the comparison between (a,b) and (c,d) shows that Laplacian filters tend to be more robust to the choice of filter parameters than Gabor filters.

For comparison purposes Fig.7 shows some example fitting results from each method, i.e., the holistic gradient descent method, the patch-based correlation method, and the patch-based correlation method using local filters. As we can see the proposed patch-based correlation method outperforms the holistic gradient descent method, and use of image filters further improves the performance. For all three methods 20 iterations are used in the fitting process.

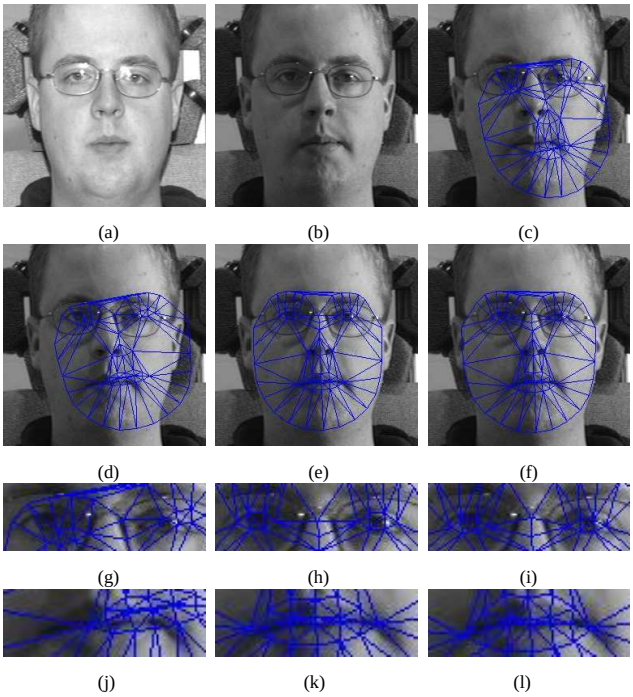


Figure 7. Example fitting results. (a) The front-lit template image. (b) The side-lit test image. (c) The initial shape position for the fitting process with 10 pixel RMS-PE. (d) The fitting result from the holistic gradient descent method. (e) The result from the patch-based normalized cross-correlation method. (f) The fitting result of the patch-based method using a 9×9 Laplacian filter. More details can be found in (g-i) and (j-l), which are the close-up views of the eye and mouth areas, respectively. As we can see the proposed patch-based correlation method outperforms the holistic gradient descent method, and use of image filters further improves the performance.

Furthermore, because the appearance variation of an object can also be caused by non-rigid motion, such as facial expressions, another set of experiments are conducted to evaluate the performance of each method in presence of non-rigid object deformations. More specifically, for each subject in the above dataset an image of the neutral expression is selected as the template (Fig.8(a)). We test the method on another image of the same subject with different facial expressions, such as smiling, under the same illumination condition (Fig.8(b)). The same initialization step is used as shown in Fig.8(c). As we can see the patch-based correlation methods (Fig.8(e,f)) achieve much better performance than the holistic gradient descent method (Fig.8(d)). The full comparison is reported in Fig.9, where the convergence rate of the patch-based methods is much higher than the holistic gradient descent method even with different initialization errors. Again, for all three methods 20 iterations are used in the fitting process.

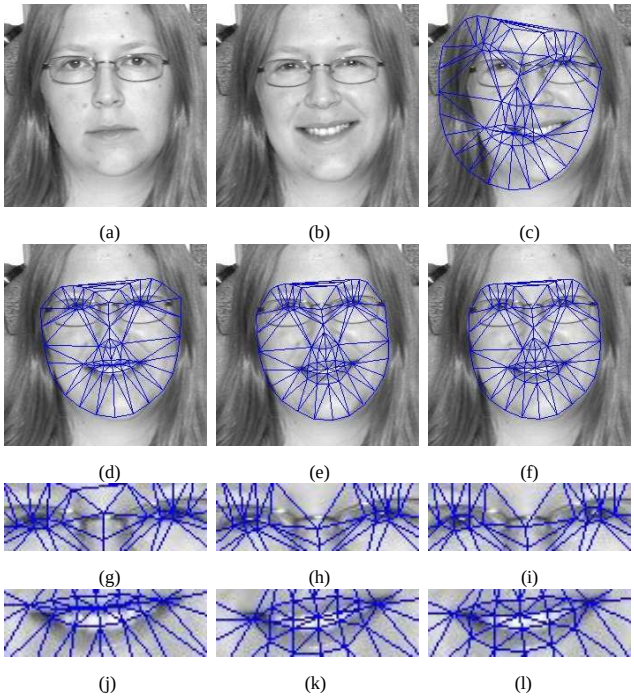


Figure 8. Example fitting results in presence of non-rigid deformations. (a) The template image with the neutral expression. (b) The test image with the smiling expression. (c) The initial shape position for the fitting process with 10 pixel RMS-PE. (d) The fitting result from the holistic gradient descent method. (e) The result from the patch-based normalized cross-correlation method. (f) The fitting result of the patch-based method using a 9×9 Laplacian filter. More details can be found in (g-i) and (j-l), which are the close-up views of the eye and mouth areas, respectively. As we can see both patch-based correlation methods (d,e) outperform the holistic gradient descent method (f).

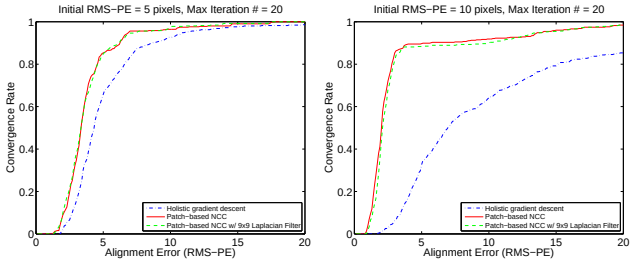


Figure 9. Comparison between holistic and patch-based methods on objects with non-rigid deformations. Different initial root mean squared point errors (RMS-PEs) are tested: 5 pixels in (a) and 10 pixels in (b). 20 iterations are used for each image. The size of the face template is 110×110 pixels and the patch size is 9×9 pixels. For the patch-based correlation method using filters, a 9×9 Laplacian filter is applied on both the source and template images. The horizontal axis shows the resulting alignment error (RMS-PE) in pixels and the vertical axis shows the convergence rate in relation to the alignment error. As we can see, the patch-based correlation methods outperforms the holistic gradient descent method for both 5 and 10 pixel initialization errors.

6. Conclusion and Future Work

In this paper we presented a new method for non-rigid object alignment, using a single appearance template of the target object and a generic (i.e. subject independent) shape model. Even when the source image and the template image do not match due to unknown appearance variation, such as the changes caused by different illumination conditions and non-rigid motion, our approach still performs well. Instead of relying on a holistic representation for the alignment process, which requires multiple training images to capture the appearance variance, our method fits the model to an unseen image based on an exhaustive local search and constrains the local warp updates within a global warping space. Furthermore, our method is not limited to intensity values or gradients and therefore offers a natural framework to integrate multiple local features, such as filter responses, to improve the robustness to large initialization errors, changing illumination conditions and non-rigid deformations. Finally, the performance of our framework is demonstrated through various experimental results, including the improvement to the accuracy and robustness of feature alignment and image registration. For future work, we will optimize the local search and global warp update jointly to further improve the fitting performance and extend our framework to handle large deformations and occlusion. Also, we will study different local feature representations to further improve the performance of our method.

7. Acknowledgements

This work was supported by the U.S. Government VACE program and by the NIH Grant R01 MH051435.

References

[1] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, February 2004.

[2] M.J. Black and A.D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *IJCV*, 26(1):63–84, January 1998.

[3] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *PAMI*, 23(6):681–685, June 2001.

[4] T.F. Cootes and C.J. Taylor. On representing edge structure for model matching. In *CVPR*, pages I:1114–1119, 2001.

[5] D. Cristinacce and T.F. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, page III:29, 2006.

[6] F. de la Torre, A. Collet Romea, J. Cohn, and T. Kanade. Filtered component analysis to increase robustness to local minima in appearance models. In *CVPR*, 2007.

[7] F. de la Torre, J. Vitria, P.I. Radeva, and J. Melenchon. Eigenfiltering for flexible eigentracking (efe). In *ICPR*, pages Vol III: 1106–1109, 2000.

[8] N.D.H. Dowson and R. Bowden. N-tier simultaneous modelling and tracking for arbitrary warps. In *BMVC*, page II:569, 2006.

[9] P.F. Felzenszwalb and D.P. Huttenlocher. Pictorial structures for object recognition. *IJCV*, 61(1):55–79, January 2005.

[10] R.S. Feris, V. Krueger, and R.M. Cesar, Jr. A wavelet subspace method for real-time face tracking. *Real-Time Imaging*, 10(6):339–350, December 2004.

[11] L. Gu and T. Kanade. 3d alignment of face in a single image. In *CVPR*, pages I: 1305–1312, 2006.

[12] G.D. Hager and P.N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20(10):1025–1039, October 1998.

[13] C. Hu, R.S. Feris, and M. Turk. Active wavelet networks for face alignment. In *BMVC*, 2003.

[14] T. Kanade and B.D. Lucas. An iterative image registration technique with an application to stereo vision. In *IJCAI*, pages 674–679, 1981.

[15] L. Liang, F. Wen, Y.Q. Xu, X. Tang, and H.Y. Shum. Accurate face alignment using shape constrained markov network. In *CVPR*, pages I: 1313–1319, 2006.

[16] I. Matthews and S. Baker. Active appearance models revisited. *IJCV*, 60(2):135–164, November 2004.

[17] M.B. Stegmann and R. Larsen. Multi-band modelling of appearance. *IVC*, 21(1):61–67, January 2003.

[18] P. Viola and M.J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, pages I:511–518, 2001.

[19] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *CVPR*, pages II: 535–542, 2004.

[20] Y. Zhou, L. Gu, and H.J. Zhang. Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In *CVPR*, pages I: 109–116, 2003.