Solving Image Registration Problems Using Interior Point Methods

Camillo Jose Taylor and Arvind Bhusnurmath

GRASP Laboratory, University of Pennsylvania

Abstract. This paper describes a novel approach to recovering a parametric deformation that optimally registers one image to another. The method proceeds by constructing a global convex approximation to the match function which can be optimized using interior point methods. The paper also describes how one can exploit the structure of the resulting optimization problem to develop efficient and effective matching algorithms. Results obtained by applying the proposed scheme to a variety of images are presented.

1 Introduction

Image registration is a key problem in computer vision that shows up in a wide variety of applications such as image mosaicing, medical image analysis, face tracking, handwriting recognition, stereo matching and motion analysis. This paper considers the problem of recovering the parameters of a deformation that maps one image onto another. The main contribution is a novel approach to this problem wherein the image matching problem is reformulated as a Linear Program (LP) which can be solved using interior point methods. The paper also describes how one can exploit the special structure of the resulting LP to derive efficient implementations which can effectively solve problems involving hundreds of thousands of pixels and constraints.

One of the principal differences between the proposed approach and other approaches that have been developed [1,2,3] is that the scheme seeks to construct a *global* convex approximation to the matching function associated with the registration problem as opposed to constructing a *local* convex model around the current parameter estimate. The approach is intended for situations where the displacements between frames are large enough that local matches at the pixel level are likely to be ambiguous. For example, in the experiments we consider images that are 320 pixels on side where individual pixels may be displaced by up to 40 pixels along each dimension. The approximation procedure is designed to capture the uncertainties inherent in matching a given pixel to a wide swath of possible correspondents.

One common approach to solving image matching problems proceeds by extracting feature points in the two images, establishing correspondences between the frames, and then using a robust estimation procedure to recover the parameters of the transformation. This approach is exemplified by the work of

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 638–651, 2008.

Mikolajczyk and Schmid [4] who proposed a very effective scheme for detecting and matching interest points under severe affine deformations. This approach works best when the interframe motion is close to affine since more complicated deformation models can distort the feature points beyond recognition. Further, it becomes increasingly difficult to apply robust estimation methods as the complexity of the deformation model increases since an ever increasing number of reliable point matches are required.

Belongie and Malik [5] proposed an elegant approach to matching shapes based on information derived from an analysis of contour features. This approach is similar to [4] in that it revolves around feature extraction and pointwise correspondence. The method described in this work is very different from these in that it avoids the notion of features altogether, instead it proceeds by constructing a matching function based on low level correlation volumes and allows every pixel in the image to constrain the match to the extent that it can.

Shekhovstov Kovtun and Hlavac [6] have developed a novel method for image registration that uses Sequential Tree-Reweighted Message passing to solve a linear program that approximates a discrete Markov Random Field optimization problem. Their work also seeks to construct a globally convex approximation to the underlying image matching problem but the approach taken to formulating and solving the optimization problem differ substantially from the method discussed in this paper.

Linear programming has been previously applied to motion estimation [7,8]. The work by Jiang *et al.*. [7] on matching feature points is similar to ours in that the data term associated with each feature is approximated by a convex combination of points on the lower convex hull of the match cost surface. However, their approach is formulated as an optimization over the interpolating coefficients associated with these convex hull points which is quite different from the approach described in this paper. Also their method uses the simplex method for solving the LP while the approach described in this paper employs an interior point solver which allows us to exploit the structure of the problem more effectively.

2 Image Registration Algorithm

The objective of the algorithm is to recover the deformation that maps a base image onto a target image. This deformation is modeled in the usual manner by introducing two scalar functions $D_x(x, y, \mathbf{p_x})$ and $D_y(x, y, \mathbf{p_y})$ which capture the displacement of a pixel at location (x, y) along the horizontal and vertical directions respectively [9,5,10]. Here $\mathbf{p_x}$ and $\mathbf{p_y}$ represent vectors of parameters that are used to model the deformation. Consider for example an affine deformation where the horizontal displacements are given by $D_x(x, y) = c_1 + c_2x + c_2y$, then $p_x = [c_1, c_2, c_3]$ would capture the parameters of this transformation. In the sequel we will restrict our consideration to models where the displacements can be written as a linear function of the parameters. That is, if we let D_x and

D_y represent vectors obtained by concatenating the displacements at all of the pixels then $D_x = C\mathbf{p_x}$ and $D_y = C\mathbf{p_y}$ for some matrix C. Here the columns of the matrix C constitute the basis vectors of the displacement field [9].

2.1 Formulating Image Matching as an LP

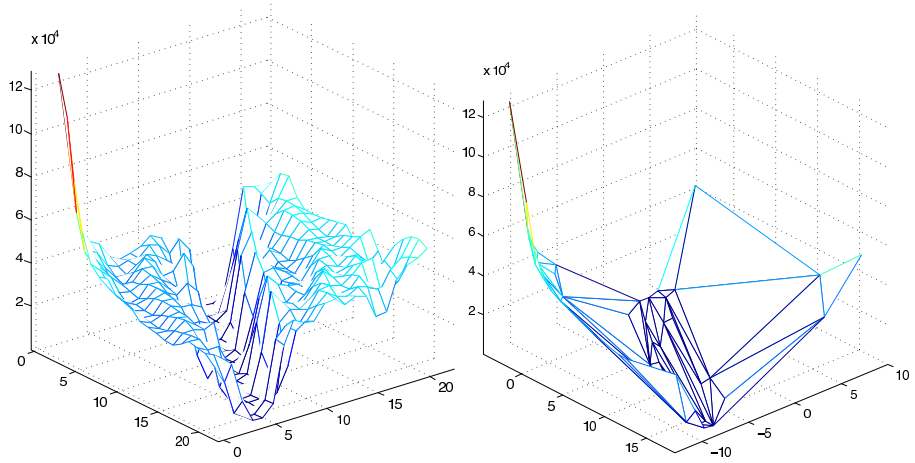
The problem of recovering the deformation that maps a given base image onto a given target image can be phrased as an optimization problem. For every pixel in the target image one can construct an objective function, e_{xy}, which captures how similar the target pixel is to its correspondent in the base image as a function of the displacement applied at that pixel.

Figure 1(a) shows an example of one such function for a particular pixel in one of the test images. This particular profile was constructed by computing the ℓ_2 difference between the RGB value of the target pixel and the RGB values of the pixels in the base image for various displacements up to ± 10 pixels in each direction.

Our goal then is to minimize an objective function $E(\mathbf{p_x}, \mathbf{p_y})$ which models how the discrepancy between the target and base images varies as a function of the deformation parameters, $\mathbf{p_x}$ and $\mathbf{p_y}$.

$$E(\mathbf{p_x}, \mathbf{p_y}) = \sum_{\mathbf{x}} \sum_{\mathbf{y}} \mathbf{e_{xy}}(\mathbf{D_x}(\mathbf{x}, \mathbf{y}, \mathbf{p_x}), \mathbf{D_y}(\mathbf{x}, \mathbf{y}, \mathbf{p_y})) \tag{1}$$

In general, since the component e_{xy} functions can have arbitrary form the landscape of the objective function $E(\mathbf{p_x}, \mathbf{p_y})$ may contain multiple local minima



(a) (b)

Fig. 1. (a) Error surface associated with particular pixel in the target image that encodes how compatible that pixel is with various x, y displacements (b) Piecewise planar convex approximation of the error surface

which can confound most standard optimization methods that proceed by constructing local approximations of the energy function.

The crux of the proposed approach is to introduce a convex approximation for the individual objective functions e'_{xy}. This leads directly to an approximation of the global objective function $E'(\mathbf{p_x}, \mathbf{p_y})$ which is convex in the deformation parameters. Once this has been done, one can recover estimates for the deformation parameters and, hence, the deformation by solving a convex optimization problem which is guaranteed to have a unique minimum.

The core of the approximation step is shown in Figure 1(b), here the original objective function is replaced by a convex lower bound which is constructed by considering the convex hull of the points that define the error surface. This convex lower hull is bounded below by a set of planar facets.

In order to capture this convex approximation in the objective function we introduce one auxiliary variable $z(x, y)$ for every pixel in the target image. There are a set of linear constraints associated with each of these variables which reflect the constraint that this value must lie above all of the planar facets that define the convex lower bound.

$$z(x,y) \geq a_x^i(x,y)D_x(x,y,p_x) + a_y^i(x,y)D_y(x,y,p_y) - b^i(x,y) \;\; \forall i \tag{2}$$

Here the terms a_x^i, a_y^i and b^i denote the coefficients associated with each of the facets in the approximation.

The problem of minimizing the objective function $E'(\mathbf{p_x}, \mathbf{p_y})$ can now be rephrased as a linear program as follows:

$$\min_{p_x,p_y,z} \sum_x \sum_y z(x,y) \tag{3}$$

$$\text{st } z(x,y) \geq a_x^i(x,y)D_x(x,y,p_x) + a_y^i D_y(x,y,p_y) - b^i(x,y) \; \forall x,y,i \tag{4}$$

This can be written more compactly in matrix form as follows:

$$\begin{gathered} \min_{p_x,p_y,z} \mathbf{1^T z} \\ \text{st } A_x D_x + A_y D_y - I_z \mathbf{z} \leq \mathbf{b} \\ D_x = C\mathbf{p_x} \\ D_y = C\mathbf{p_y} \end{gathered} \tag{5}$$

where A_x and A_y are I_z are sparse matrices obtained by concatenating the constraints associated with all of the planar facets and z and b are vectors obtained by collecting the $z(x, y)$ and $b^i(x, y)$ variables respectively.

Note that the A_x, A_y and I_z matrices all have the same fill pattern and are structured as shown in equation 6, the non zero entries in the I_z matrix are all 1. In this equation M denotes the total number of pixels in the image and S_i refers to the number of planar facets associated with pixel i.

$$A = \begin{bmatrix} a_{11} & 0 & \cdots\cdots & 0 \\ a_{21} & 0 & \cdots\cdots & 0 \\ \vdots & 0 & \cdots\cdots & 0 \\ a_{S_1 1} & 0 & \cdots\cdots & 0 \\ 0 & a_{12} & 0 \cdots & 0 \\ 0 & a_{22} & 0 \cdots & 0 \\ 0 & \vdots & 0 \cdots & 0 \\ 0 & a_{S_2 2} & 0 \cdots & 0 \\ 0 & 0 & \ddots \ddots & 0 \\ 0 & \cdots & \cdots 0 & a_{1M} \\ 0 & \cdots & \cdots 0 & \vdots \\ 0 & \cdots & \cdots 0 & a_{S_M M} \end{bmatrix} \tag{6}$$

The linear program shown in Equation 5 can be augmented to include constraints on the displacement entries, D_x, D_y and the z values as shown in Equation 7. Here the vectors b_{lb} and b_{ub} capture the concatenated lower and upper bound constraints respectively. It would also be a simple matter to include bounding constraints on the parameter values at this stage. Alternatively one could easily add a convex regularization term to reflect a desire to minimize the bending energy associated with the deformation.

$$\min_{p_x, p_y, z} \mathbf{1^T z} \tag{7}$$
$$\begin{bmatrix} A_x \; A_y \; -I_z \\ -I \\ I \end{bmatrix} \begin{pmatrix} C & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \mathbf{p_x} \\ \mathbf{p_y} \\ z \end{pmatrix} \leq \begin{pmatrix} b \\ b_{lb} \\ b_{ub} \end{pmatrix}$$

Note that the proposed approximation procedure increases the ambiguity associated with matching any individual pixel since the convex approximation is a lower bound which may significantly under estimate the cost associated with assigning a particular displacement to a pixel. What each pixel ends up contributing is a set of convex terms to the global objective function. The linear program effectively integrates the convex constraints from tens of thousands of pixels, constraints which are individually ambiguous but which collectively identify the optimal parameters. In this scheme each pixel contributes to constraining the deformation parameters to the extent that it is able. Pixels in homogenous regions may contribute very little to the global objective while well defined features may provide more stringent guidance. There is no need to explicitly identify distinguished features since local matching ambiguities are handled through the approximation process.

2.2 Solving the Matching LP

Once the image registration problem has been reformulated as the linear program given in equation 7 the barrier method [11] can be employed to solve the problem. In this method, a convex optimization problem of the following form

$$\begin{aligned} \min \quad & f_0(x) \\ \text{st } & f_i(x) \leq 0,\ i = 1, \ldots, m \end{aligned} \tag{8}$$

is solved by minimizing $\phi(x,t) = tf_0(x) - \sum_{i=1}^{m} \log(-f_i(x))$ for increasing values of t until convergence. At each value of t a local step direction, the Newton step, needs to be computed. This involves the solution of a system of linear equations involving the Hessian and the gradient of $\phi(x,t)$. The Hessian can be computed from the following expression $H = [A^T \text{diag}(s^{-2})A]$ where $s = b - Ax$ and s^{-2} denotes the vector formed by inverting and squaring the elements of s. Similarly the gradient of the $\phi(x,t)$ can be computed from the following expression:

$$g = -tw - A^T s^{-1} \tag{9}$$

Then the Newton step is computed by solving

$$[A^T \text{diag}(s^{-2})A]\delta x = g \tag{10}$$

For our matching problem, it can be shown that this Newton step system can be written in the following form:

$$\begin{bmatrix} H_p & H_z^T \\ H_z & D_6 \end{bmatrix} \begin{pmatrix} \delta p \\ \delta z \end{pmatrix} = \begin{pmatrix} g_p \\ g_z \end{pmatrix} \tag{11}$$

where

$$\begin{aligned} H_p &= \begin{bmatrix} (C^T D_1 C) & (C^T D_2 C) \\ (C^T D_2 C) & (C^T D_3 C) \end{bmatrix} \\ H_z &= \begin{bmatrix} (D_4 C) & (D_5 C) \end{bmatrix} \end{aligned} \tag{12}$$

δp and δz denote proposed changes in the deformation parameters and the z variables respectively and $D_1, D_2, D_3, D_4, D_5, D_6$ are all diagonal matrices.

At this point we observe that since the matrix D_6 is diagonal we can simplify the linear system in Equation 11 via the Schur complement. More specifically we can readily solve for δz in terms of δp as follows: $\delta z = D_6^{-1}(g_z - H_z \delta p)$. Substituting this expression back into the system yields the following expression where all of the auxiliary z variables have been elided.

$$(H_p - H_z^T D_6^{-1} H_z)\delta p = (g_p - H_z^T D_6^{-1} g_z) \tag{13}$$

This can be written more concisely as follows:

$$H_p' \delta p = g_p' \tag{14}$$

In short, computing the Newton Step boils down to solving the linear system in Equation 14. Note that the size of this system depends only on the dimension of the parameter vector, p. For example if one were interested in fitting an affine model which involves 6 parameters, 3 for $\mathbf{p_x}$ and 3 for $\mathbf{p_y}$, one would only end

up solving a linear system with six degrees of freedom. Note that the computational complexity of this key step *does not depend on the number of pixels being considered or on the number of constraints that were used to construct the convex approximation.* This is extremely useful since typical matching problems will involve hundreds of thousands of pixels and a similar number of constraint equations. Even state of the art LP solvers like MOSEK and TOMLAB would have difficulty solving problems of this size.

2.3 Deformation Models

Experiments were carried out with two classes of deformation models. In the first class the displacements at each pixel are computed as a polynomial function of the image coordinates. For example for a second order model:

$$D_x(x,y) = c_1 + c_2x + c_3y + c_4xy + c_5x^2 + c_6y^2 \tag{15}$$

These deformations are parameterized by the coefficients of the polynomials. The complexity of the model can be adjusted by varying the degree of the polynomial. A number of interesting deformation models can be represented in this manner include affine, bilinear, quadratic and bicubic.

Another class of models can be represented as a combination of an affine deformation and a radial basis function. That is

$$D_x(x,y) = c_1 + c_2x + c_3y + \sum_i k_i\phi(\|(x,y) - (x_i,y_i)\|) \tag{16}$$

Once again the deformation model is parameterized by the coefficients c_1, c_2, c_3, k_i and the function ϕ represents the interpolating kernel. Two different variants of this kernel were considered in the experiments, a Gaussian kernel, $\phi(r) = \exp(-(r/\sigma)^2)$ and a thin plate spline kernel $\phi(r) = r^2 \log r$. In the sequel we will refer to the former as the Gaussian deformation model and the latter as the Thin Plate Spline model.

In the experiments the coordinates of the kernel centers,(x_i, y_i) were evenly distributed in a grid over the the image. The complexity of the model can be varied by varying the number of kernel centers employed. All of the experiments that used this model employed 16 kernel centers arranged evenly over the image in a four by four grid.

2.4 Coarse to Fine

It is often advantageous to employ image registration algorithms in a coarse to fine manner [1]. In this mode of operation the base and target images are downsampled to a lower resolution and then matched. The deformation recovered from this stage is used to constrain the search for matches at finer scales. With this scheme, gross deformations are captured at the coarser scales while the finer scales fill in the details. It also serves to limit the computational effort required since one can effectively constrain the range of displacements that must

be considered at the finer scales which limits the size of the correlation volumes that must be constructed.

In the experiments described in section 3.1 the images are first downsampled by a factor of 4 and then matched. The deformations computed at this scale inform the search for correspondences at the next finer scale which is downsampled from the originals by a factor of 2.

Note that as the approach proceeds to finer scales, the convex approximation is effectively being constructed over a smaller range of disparities which means that it increasingly approaches the actual error surface.

3 Experimental Results

Two different experiments were carried out to gauge the performance of the registration scheme quantitatively. In the first experiment each of the images in our data set was warped by a random deformation and the proposed scheme was employed to recover the parameters of this warp. The recovered deformation was compared to the known ground truth deformation to evaluate the accuracy of the method.

In the second set of experiments the registration scheme was applied to portions of the Middlebury stereo data set. The disparity results returned by the method were then compared to the ground truth disparities that are provided for these image pairs.

3.1 Synthetic Deformations

In these experiments the proposed scheme was applied to a number of different images. In each case, a random deformation was constructed using a particular motion model. The base image was warped by the deformation to produce the target image and the registration algorithm was employed to recover this deformation. In these experiments each of the base images was at most 320 pixels on side. The deformations that were applied were allowed to displace the pixels in the base image by up to $\pm 12.5\%$ of the image size. Hence for an image 320 pixels on side each pixel in the image can be displaced by ± 40 pixels along each dimension. The random deformations were specifically constructed to fully exercise the range of displacements so the maximum allowed displacement values are achieved in the applied warps. In order to recover such large deformations, the registration scheme is applied in a coarse to fine manner as described in Section 2.4.

The underlying matching functions associated with each of the pixels in the target image, e_{xy}, are constructed by simply comparing the pixel intensity in the target image to the pixels in a corresponding range in the base image. This is equivalent to conducting sum of squared difference (SSD) matching for each pixel using a 1×1 matching window.

In order to provide a quantitative evaluation of the scheme, the recovered deformation field, $(D_x(x, y), D_y(x, y))$ was compared to the known ground truth deformation field $(D_x^t(x, y), D_y^t(x, y))$ and the mean, median and maximum discrepancy between these two functions over the entire image was computed. The

Table 1. This table details the deformation applied to each of the images in the data set and reports the discrepancy between the deformation field returned by the method and the ground truth displacement field

Image	Deformation Model	no. of parameter	error in pixels		
			mean	median	max
Football	Gaussian	38	0.1524	0.1306	0.5737
Hurricane	Gaussian	38	0.1573	0.1262	0.7404
Spine	Affine	6	0.1468	0.1314	0.4736
Peppers	Gaussian	38	0.1090	0.0882	0.7964
Cells	Thin Plate Spine	38	0.1257	0.1119	0.8500
Brain	Gaussian	38	0.1190	0.0920	0.8210
Kanji	third degree polynomial	20	0.1714	0.0950	2.5799
Aerial	bilinear	8	0.0693	0.0620	0.2000
Face1	Gaussian	38	0.1077	0.0788	0.6004
Face2	Gaussian	38	0.5487	0.3095	4.6354

results are tabulated in Table 1. This table also indicates what type of deformation model was applied to each of the images along with the total number of parameters required by that model.

Note that in every case the deformed result returned by the procedure is almost indistinguishable from the given target. More importantly, the deformation fields returned by the procedure are consistently within a fraction of a pixel of the ground truth values. The unoptimized Matlab implementation of the matching procedure takes approximately 5 minutes to proceed through all three scales and produce the final deformation field for a given image pair.

3.2 Stereo Data Set

The image registration scheme was applied to regions of the image pairs taken from the Middlebury stereo data set. This data set was chosen because it included ground truth data which allows us to quantitatively evaluate the deformation results returned by the registration scheme. Here the vertical displacement between the two images is zero and the horizontal displacement field $D_x(x, y)$ is modeled as an affine function.

The correlation volume was computed using sum of squared difference matching with a five by five correlation window. For the teddy image, the correlation volume was constructed by considering displacements between 12 and 53 pixels while for the venus image the displacement range was 3 to 20 pixels. In this case, the convex lower bound approximations to the individual score functions degenerates to a piecewise linear profile along the horizontal dimension.

In each of the images two rectangular regions were delineated manually and an affine displacement model was fit to the pixels within those regions using the proposed method.

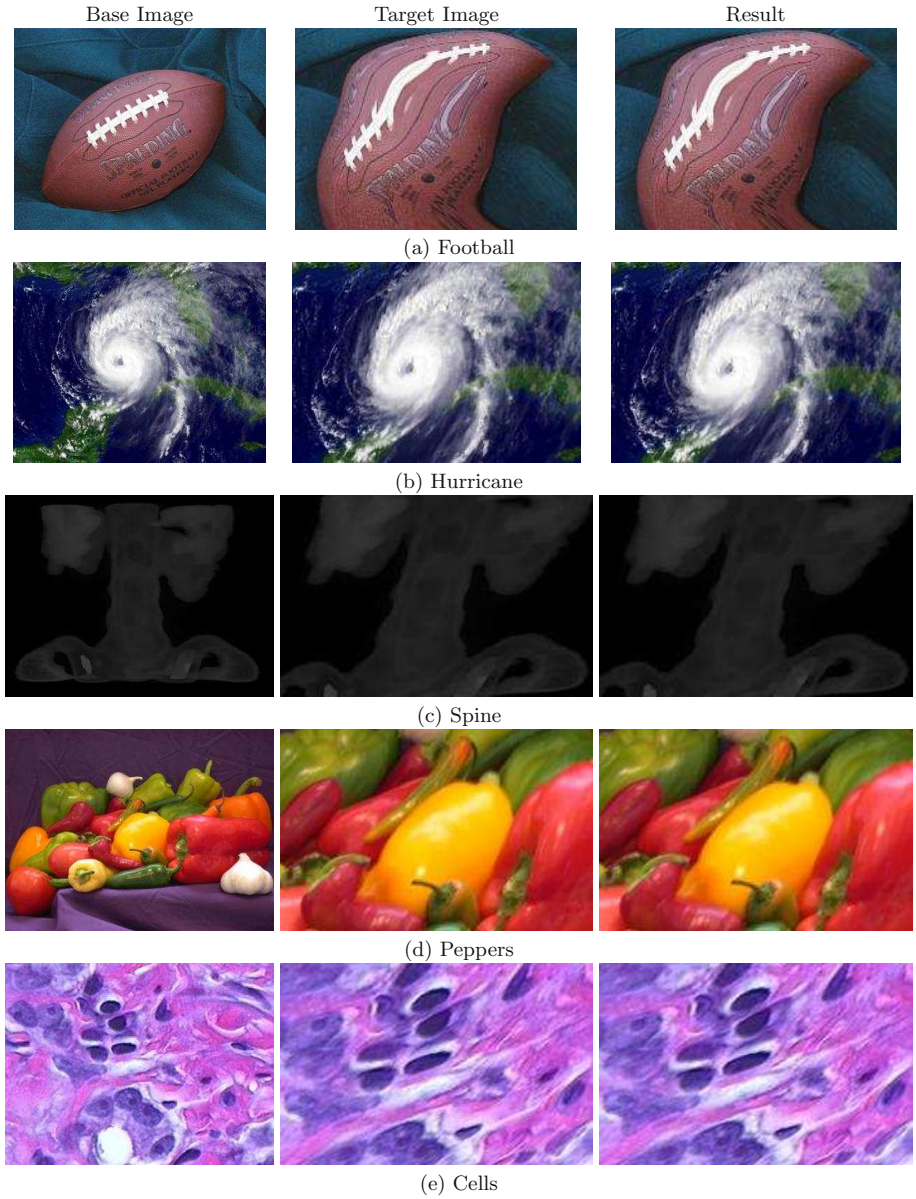
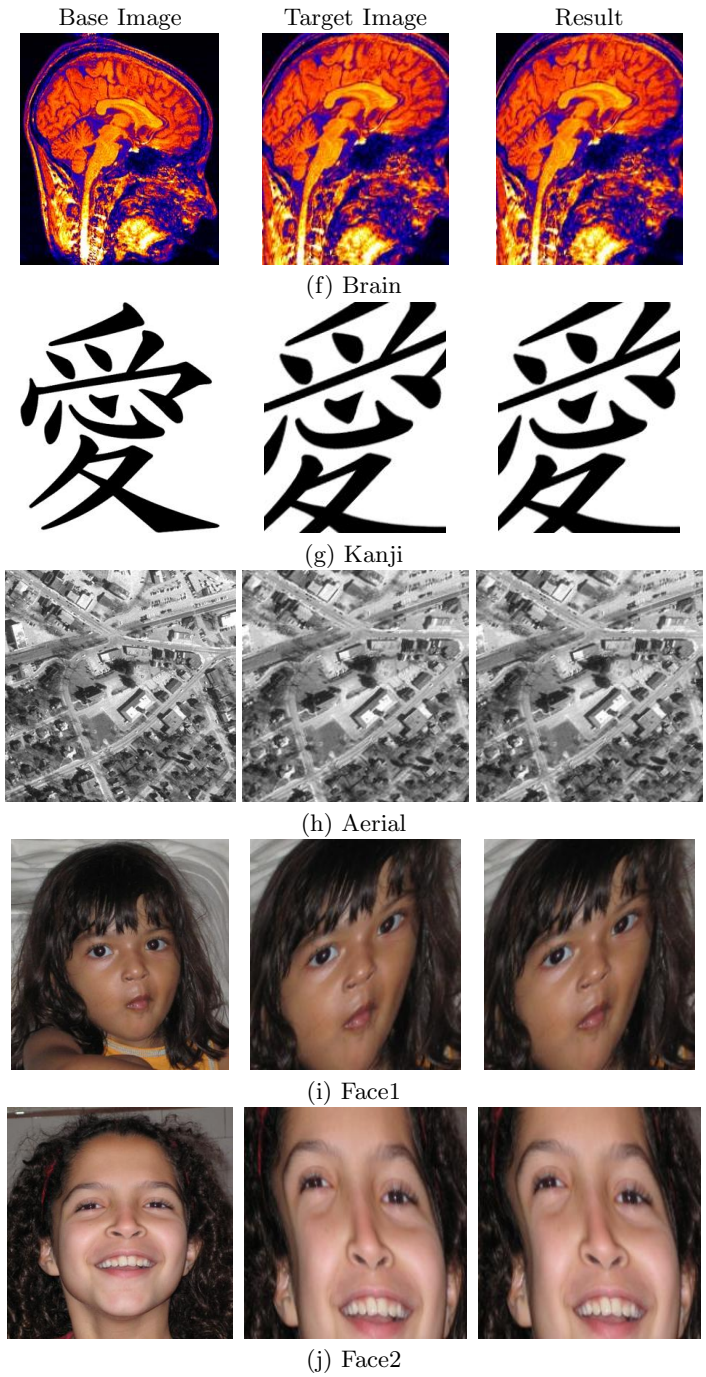


Fig. 2. Results obtained by applying the proposed method to actual image pairs. The first two columns correspond to the input base and target images respectively while the last column corresponds to the result produced by the registration scheme.

The first column of Figure 4 shows the left image in the pair, the second column shows what would be obtained if one used the raw SSD stereo results and the final column shows the ground truth disparities.



(f) Brain

(g) Kanji

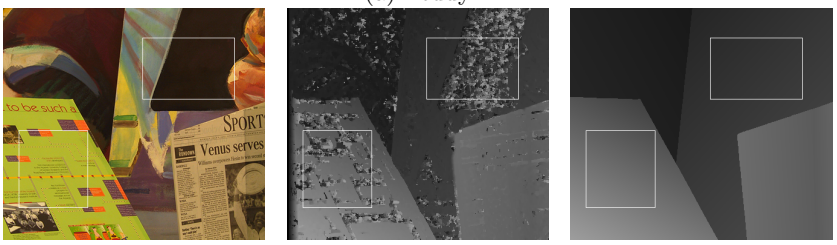
(h) Aerial

(i) Face1

(j) Face2

Fig. 3. More Registration Results

(a) Teddy



(b) Venus

Fig. 4. The proposed image registration scheme was applied to the delineated regions in the Middlebury Stereo Data Set. The first column shows the left image, the second column the raw results of the SSD correlation matching and the last column the ground truth disparity.

Table 2. This table reports the discrepancy between the affine deformation field returned by the method and the ground truth disparities within each region

Image	Region	error in pixels	
		mean	median
teddy	bird house roof	0.2558	0.2245
teddy	foreground	0.9273	0.8059
venus	left region	0.0317	0.0313
venus	right region	0.0344	0.0317

The selected rectangles are overlaid on each of the images. These regions were specifically chosen in areas where there was significant ambiguity in the raw correlation scores to demonstrate that the method was capable of correctly integrating ambiguous data. Table 2 summarizes the results of the fitting procedure. The reconstructed disparity fields within the regions were compared to the ground truth disparities and the mean and median discrepancy between these two fields is computed over all of the pixels within the region.

4 Conclusion

This paper has presented a novel approach to tackling the image registration problem wherein the original image matching objective function is approximated

by a linear program which can be solved using the interior point method. The paper also describes how one can exploit the special structure of the resulting linear program to develop efficient algorithms. In fact the key step in the resulting resulting procedure only involves inverting a symmetric matrix whose dimension reflects the complexity of the model being recovered.

While the convex approximation procedure typically increases the amount of ambiguity associated with any individual pixels, the optimization procedure effectively aggregates information from hundreds of thousands of pixels so the net result is a convex function that constrains the actual global solution. In a certain sense, the proposed approach is dual to traditional non-linear optimization schemes which seek to construct a *local* convex approximation to the objective function. The method described in this work proceeds by constructing a *global* convex approximation over the specified range of displacements.

A significant advantage of the approach is that once the deformation model and displacement bounds have been selected, the method is insensitive to initialization since the convex optimization procedure will converge to the same solution regardless of the start point. This means that the method can be directly applied to situations where there is a significant deformation.

The method does not require any special feature detection or contour extraction procedure. In fact all of the correlation volumes used in the experiments were computed using nothing more than pointwise pixel comparisons. Since the method does not hinge on the details of the scoring function more sophisticated variants could be employed as warranted. The results indicate the method produces accurate results on a wide range of image types and can recover fairly large deformations.

References

1. Bajcsy, R., Kovacic, S.: Multiresolution elastic matching. Computer Vision, Graphics and Image Processing 46(1), 1–21 (1989)
2. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)
3. Baker, S., Matthews, I.: Equivalence and efficiency of image alignment algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1090–1097 (2001)
4. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal of Computer Vision 60(1), 63–86 (2004)
5. Belongie, S., Malik, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(24), 509 (2002)
6. Shekhovstov, A., Kovtun, I., Hlavac, V.: Efficient mrf deformation model for non-rigid image matching. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
7. Jiang, H., Drew, M., Li, Z.N.: Matching by linear programming and successive convexification. PAMI 29(6) (2007)
8. Ben-Ezra, M., Peleg, S., Werman, M.: Real-time motion analysis with linear programming. In: ICCV (1999)

9. Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S.J.: Spatial registration and normalization of images. Human Brain Mapping 2, 165–189 (1995)
10. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)
11. Boyd, S., VandenBerghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)