

# Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence

Gehua Yang<sup>1</sup> Charles V. Stewart<sup>1</sup> Michal Sofka<sup>1</sup> Chia-Ling Tsai<sup>1,2</sup>  
<sup>1</sup> Rensselaer Polytechnic Institute    <sup>2</sup> National Chung Cheng University  
Troy, NY 12180, U.S.A.                      Chiayi, Taiwan 621, R.O.C.  
`{yangg2,stewart,sofka,tsaic}@cs.rpi.edu`

## Abstract

*Our goal is a registration algorithm capable of aligning image pairs having some combination of low overlap, large illumination differences (e.g. day and night), substantial scene changes and different modalities. Our approach starts by extracting and matching keypoints. Ranked-ordered matches are tested individually in succession. Each is used to generate a transformation estimate in a small image region surrounding the keypoints. The growth process works by iterating three steps: 1) refining the estimate by symmetrically matching features on the two images, 2) expanding the region according to the uncertainty in the mapping, 3) selecting an appropriate transformation model. Image features are corner points and face points located by analyzing the intensity structure of image neighborhoods. After convergence, if a correctness test verifies the transformation it is accepted and the algorithm ends; otherwise the process starts over with the next keypoint match. Experimental results on a suite of challenging image pairs shows that the algorithm substantially out-performs recent algorithms based on keypoint matching.*

## 1 Introduction

This paper addresses the problem of developing an image registration algorithm that can work on many different types of images, scenes, and illumination conditions. The algorithm should successfully align pairs of images taken of indoor or outdoor scenes, and in natural or man-made environments. It should be able to align images taken at different times of day, during different seasons of the year, or using different imaging modalities. The algorithm should adjust for rotation and zoom between the images and for low image overlap. Our primary assumption is that the images to be aligned should be spatially-related by a known transformation model — the most common model being a planar projective transformation. Such a registration algorithm will have numerous applications ranging from mosaic

construction to change detection and visualization.

In order to make the difficulty of general-purpose registration concrete, we have gathered a test suite of 18 challenging image pairs, some of which are shown in Figures 1-3.<sup>1</sup> The alignment is clear to the human observer for each pair but difficult for current registration algorithms. Two crucial issues emerge from examining and experimenting with these images. First, initialization is difficult. Recently developed keypoint detection and matching algorithms only produce a small number of correct matches, occasionally none and sometimes fewer than 10 out of the top 100. Second, there is often no relationship between the intensities for a large fraction of the image pixels. For example, in the winter-summer pair from Figure 3, snow on the roofs in winter produces homogeneous intensity regions, where these roofs appear as dark, textured regions in the summer image.

The key idea behind our proposed algorithm is to start at just one corresponding location and grow, discovering consistency between images as part of the alignment process. This intuition is realized in several important steps:

1. The algorithm starts by matching keypoints. Each match is used individually to generate an initial similarity transformation, which is roughly accurate in a small image region. This means the algorithm can succeed even if only one keypoint match is correct.
2. The initial transformation is “grown” into an image-wide alignment by iterating steps of matching, robust refinement, model selection and region growing, controlled by the transformation estimate error and uncertainty. (This process, called the “Dual-Bootstrap,” was originally developed for retinal images — reference omitted.) The growth and refinement process keeps the alignment accurate within the growing region, using robust techniques to select only the constraints that are consistent.

---

<sup>1</sup>All of these will be available through our website.

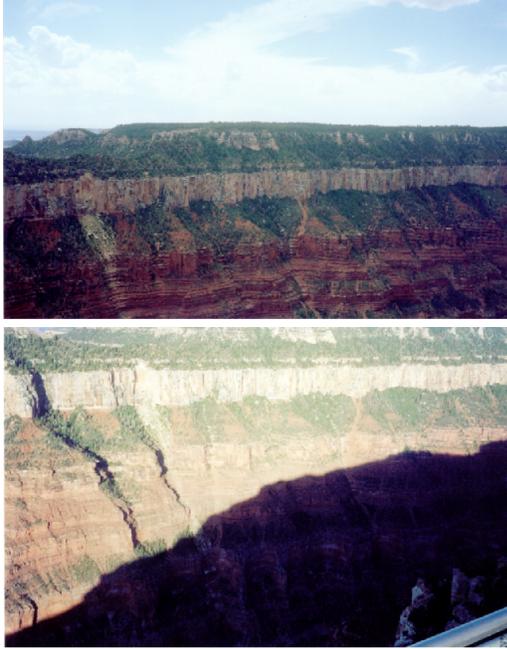


Figure 1: Two images of the North-Rim of the Grand Canyon (test pair “GC-2”) taken morning and midday.

3. When the growth process reaches the image boundaries, the alignment is tested for overall accuracy and consistency. If it passes this test, the algorithm halts with success. If not, another keypoint match is tested using the foregoing steps.

Step 3 matches corners and face points detected across multiple scales using a process that is an extension of Harris corner detection. Each feature point characterizes the intensity structure in small image regions. The overall algorithm will be analyzed on our test suite and compared against the behavior of the keypoint algorithm of [4].

## 2 Background

Many papers have been published on the various aspects of the registration problem. We focus on four approaches most relevant to the problems addressed here.

Much recent work has focused on extracting and matching multiscale keypoints, either Harris corners [11] or scale-scale peaks of the Laplacian of Gaussian [9]. An invariant descriptor is extracted from the intensities near each keypoint, and used by indexing methods to match keypoints between images. Brown and Lowe [4] have described an algorithm that uses random sampling of keypoint matches to align images and form mosaics. This approach handles scale and orientation changes well, but it relies on sufficient number of correct keypoint matches.

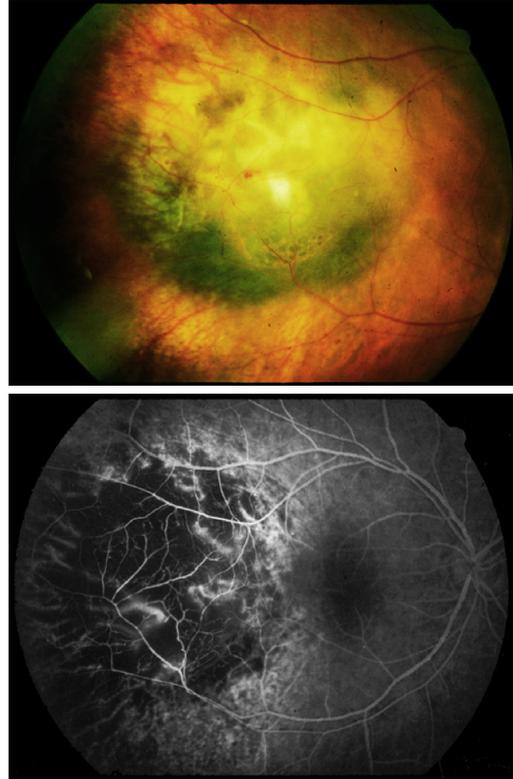


Figure 2: A color image and a fluorescein angiogram (test pair “Melanoma”) of a patient having a choroidal melanoma.

Many algorithms directly minimize intensity differences between images. A number of these fall under the framework of the Lucas-Kanade approach which computes the alignment by incrementally warping one image onto another [1]. This is most effective in a hierarchical setting [2]. The method has been extended to handle multisensor data by computing and normalizing derivatives in four directions as the “intensity” at each pixel [8]. These techniques require good initialization, although coarse searches have proven effective in many cases [14]. Scale and orientation differences between images cause problems for these methods.

Mutual information has been used effectively for aligning multimodal imagery, mostly in medical image analysis [10, 16]. It requires good initialization, and typically uses an expensive, non-differential search. The possibility of incorporating mutual information in the framework proposed here may be considered in future work.

Finally, our technique employs the well-known iterative closest points algorithm [3, 6, 13], which alternates (re)matching and estimation steps. While ICP has been used most often in range image registration, it can be employed using intensity images based on extracted features.

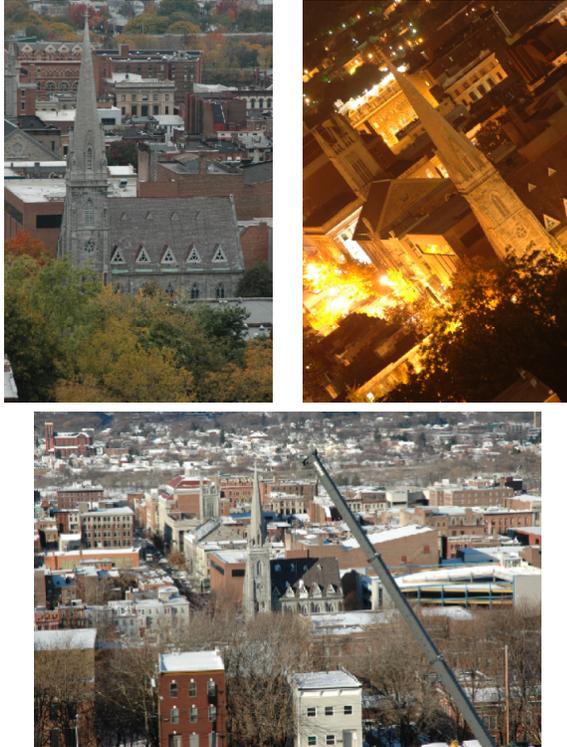


Figure 3: Three images of the same scene near our campus. The top two images were taken day and night during the summer. The bottom was taken during the day in the winter (and with a much smaller scale). They form three test pairs: “Day-Night”, “Winter-Summer” and “DNWS” (Day-Night-Winter-Summer).

### 3 Initialization

Initialization is crucial because nothing is assumed about the relative position, orientation and scales of the images, not to mention affine and projective distortions. Extraction and matching of keypoints is used because keypoint techniques are designed to be invariant to similarity distortion of the image coordinates and linear scaling of intensities. In particular, we use the publically-available implementation of Lowe’s keypoint extraction algorithm [9] and reimplemented his matching algorithm. Similar results are obtained with [11].

Multiscale keypoints are extracted from each image separately. Keypoints are then matched, and rank-ordered based on their distinctiveness. In our algorithm, as discussed earlier, each is used separately, starting with the highest ranking match, to generate an initial similarity transformation. This initial transformation is refined locally and then the Dual-Bootstrap procedure is applied. If the resulting transformation is verified as correct, the procedure stops. Otherwise, the next ranked match is tested. This



Figure 4: Initial keypoint match and side-by-side alignment for one of our winter-summer pairs. The image region on the right has been scaled by a factor of 2.25, and there are substantial illumination, blurring, and changes (snow) between the regions.

greedy procedure continues until one transformation is verified and accepted, or a fixed number have been tried.

The initial similarity transformation is established from the positions of the two keypoints, the orientations of their dominant intensity gradients, and their relative scales (Figure 4). This provides a rough local alignment, which forms the starting point for the Dual-Bootstrap procedure.

### 4 Feature Extraction

Before describing the growth and refinement procedure, we consider the choice of image quantities to drive the alignment process. In contrast to recent trends, we use image features. There are two reasons for this. First, matching image features provides direct measurement of the geometric alignment error. This is needed to drive the growth and model selection processes. The second reason is motivated by the changes seen between images that must be aligned. Much of the image texture and details may change between images — e.g. snow covering rooftops, leaves missing, textured regions falling into deep shadows — but structural outlines usually remain unchanged (Figure 5). These outlines, large or small, can be captured by properly extracted features. Our matching and robust estimation techniques exploit the consistent features between images to drive alignment.

Two different feature types are located — corners and face points. At each pixel, eigenvalues  $\lambda_1 < \lambda_2$  of the gradient outer product matrix  $\mathbf{M} = (\nabla I(\mathbf{x})(\nabla I(\mathbf{x}))^\top$  are computed. Potential *corners* are located at pixels where  $\lambda_1/\lambda_2 > t_a$ . This criterion is similar to the Harris corner detector [7]. Potential *face points* are located at pixels for which  $\lambda_1/\lambda_2 \leq t_a$ . Decision value  $t_a$  has been experimentally determined as 0.1, although the choice of values is not crucial. Strength is assigned to each feature as  $s = \text{trace}(\mathbf{M}) = \sum_i m_{ii}$ . Significantly, the features are extracted at multiple scales. Non-maximum suppression and subpixel localization are computed at each scale separately.



Figure 5: Examples of substantial changes between image regions due to illumination differences, scale differences, and changes (snow).

The next steps are designed to avoid difficulties due to low image contrast and threshold effects. First, a very low threshold,  $t_s$  is applied to the strength — on the order of one grey level — to eliminate plainly noise edges. Next, a minimum number of strongest features is kept without further testing. Then each remaining point is tested in order of decreasing strength to ensure that it has locally largest strength and it is not close to other features. This procedure stops when a maximum number of features is found. A minimum distance between features is set to ensure that these are spread through the image, and this distance grows with scale. The resulting features are called *matchable features*. The second and final step is to extract a reduced subset by increasing the spacing and strength parameters to obtain a set of *driving features* (similar to those in [15]). Driving features are transformed and matched against matchable features.

A course-scale example set of driving features is shown in Figure 6. Features are spread throughout the image. In effect the features represent summaries of local image structure: when a region contains substantial spatial variations in all directions a corner is placed at the location of locally greatest strength; when a region contains variation in one direction, a face point is placed, again at the (subpixel) location of locally greatest strength; when variation is insignificant, no feature is placed.

## 5 DB-ICP

The Dual-Bootstrap ICP (DB-ICP) algorithm begins with an initial transformation estimate and initial matching regions from the two images obtained by keypoint matching. The algorithm iterates steps of (1) refining the current transformation in the current “bootstrap” region by symmetric



Figure 6: Example intermediate resolution driving features, which are more sparse than matchable features. Circles are corners and line segments are face points, oriented along the direction of greatest eigenvalue. As the resolution increases, the feature sets become much denser, and the proportion of face points vs. corner points increases substantially.

matching,  $R$ , (2) applying model selection to determine if a more sophisticated model may be used, and (3) expanding the region, growing inversely proportional to the uncertainty of the mapping on the region boundary (Figure 7). While the framework of this algorithm has been described elsewhere for retinal image registration (ref omitted), many of the details must be changed and extended to make the approach work for a wider class of images. These details are emphasized in the remainder of this section.

### 5.1 Notation

The two images are  $I_p$  and  $I_q$ . The matchable corner and face points are  $\mathcal{P}_c = \{\mathbf{p}_c\}$  and  $\mathcal{P}_f = \{\mathbf{p}_f\}$  from  $I_p$  and  $\mathcal{Q}_c = \{\mathbf{q}_c\}$  and  $\mathcal{Q}_f = \{\mathbf{q}_f\}$  from  $I_q$ . Points from all scales are combined to form these sets. Driving features sets are subsets of  $\mathcal{P}_c, \mathcal{P}_f, \mathcal{Q}_c,$  and  $\mathcal{Q}_f$ . The transformation of point location  $\mathbf{x}$  is  $\mathbf{T}(\mathbf{x}; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is the parameter vector to be estimated. An estimate is  $\hat{\boldsymbol{\theta}}$ , and its covariance matrix is  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ . The initial model computed from the keypoint match is a similarity transformation. Model selection transitions from similarity to affine to homography, and in some cases to a homography plus radial-lens distortion or, in the case of retinal images, a quadratic transformation. Finally, the region over which the transformation is being estimated is called the “bootstrap” region, and is denoted by  $R$ .

### 5.2 Refinement Within the Bootstrap Region

The transformation is refined within current bootstrap region  $R$ , ignoring everything else in the two images. The



Figure 7: Example steps of the Dual-Bootstrap growth and refinement process. The left, center and right show the alignment results and bootstrap region for iterations 0, 6 and 12, respectively. The rectangles outline the bootstrap region. Within the bootstrap region, the alignment is accurate, but perhaps not outside the bootstrap region. As the bootstrap region expands, more and more of the images are accurately aligned. The final alignment is shown in Figure 9.

current transformation is used to generate a new set of correspondences, and these correspondences are used to generate a new transformation. Unlike standard ICP, the algorithm proceeds to model selection and region growing before selecting a new set of matches.

Matching is applied from  $I_p$  to  $I_q$  and symmetrically from  $I_q$  to  $I_p$ . A driving feature  $\mathbf{p}$  from  $I_p$  is mapped into  $I_q$  to produce  $\mathbf{p}' = \mathbf{T}(\mathbf{p}; \hat{\boldsymbol{\theta}})$ , and the  $m = 3$  closest matchable features (of the same type) to  $\mathbf{p}'$  are found. One of these is selected as the best match based on a similarity measure described below. Because the driving features are held to stricter criteria, it is unlikely that a driving feature will be found in one image and its corresponding matchable feature will be missed in the other image, except when there are substantial illumination or physical changes between images. In this case, matching of other structures in the image must constrain the registration process. The corner and face point correspondence sets, computed by matching in both directions, are  $\mathcal{C}_c = \{(\mathbf{p}_{c,i}, \mathbf{q}_{c,i})\}$  and  $\mathcal{C}_f = \{(\mathbf{p}_{f,j}, \mathbf{q}_{f,j})\}$ , respectively. Symmetric matching provides more constraints and more numerically stable estimate.

The similarity measure is a weight, denoted  $w_s$ . For a potential match  $(\mathbf{p}', \mathbf{q})$  between corners, the weight is the ratio of the scales,  $s_q$  and  $s_{p'}$  at which they are detected, with the scale of  $\mathbf{p}'$  multiplied by the scale of the transformation:  $w_s = \min(s_{p'}/s_q, s_q/s_{p'})$ . This biases the selection toward features at similar scales. If the match is between face points,  $w_s$  is multiplied by  $|\mathbf{n}_{p'} \cdot \mathbf{n}|$ , where  $\mathbf{n}_{p'}$  is the transformed normal of  $\mathbf{p}$  and  $\mathbf{n}_q$  is the normal of  $\mathbf{q}$ .

Before defining the transformation estimation objective function, we need to define the error distances. In particular, these are

$$d_c(\mathbf{p}', \mathbf{q}) = \|\mathbf{p}' - \mathbf{q}\| \quad \text{and} \quad d_f(\mathbf{p}', \mathbf{q}) = |(\mathbf{p}' - \mathbf{q})^T \mathbf{n}|$$

for corners and face points, respectively, where  $\mathbf{n}$  is the nor-

mal at  $\mathbf{q}$ .<sup>2</sup> Using this, for a fixed set of matches and weights, the transformation can be re-estimated by minimizing

$$E(\boldsymbol{\theta}; \mathcal{C}_c, \mathcal{C}_f) = \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_c} w_{s,i} w_{d,i} d_c(T(\mathbf{p}_i; \boldsymbol{\theta}), \mathbf{q}_i)^2 + \sum_{(\mathbf{p}_j, \mathbf{q}_j) \in \mathcal{C}_f} w_{s,j} w_{d,j} d_f(T(\mathbf{p}_j; \boldsymbol{\theta}), \mathbf{q}_j)^2 \quad (1)$$

where  $\sigma_c$  and  $\sigma_f$  are the robustly computed error variance for corner points and face points. The weight,  $w_{d,i}$ , is robust alignment error weight. This is  $w_{d,i} = w(d(\mathbf{p}'_i, \mathbf{q}_i)/\sigma)/\sigma^2$ , where  $w$  is the Beaton-Tukey robust weight function,  $d(\cdot)$  is the distance function, and  $\sigma^2$  is the variance. Re-estimating  $\hat{\boldsymbol{\theta}}$  is carried out by iteratively minimizing (1) and then re-estimating the robust weights  $w_d$  — in other words using iteratively-reweighted least-squares (IRLS). The error variances re-estimated in the first iteration from the weighted average of the square distances for corners and face points separately. The objective function (1) for fixed weights may be directly minimized as a linear regression for the affine transformation, with the covariance matrix computed from the scatter matrix.

When estimating a homography, or homography with radial lens distortion, the minimization is no longer analogous to regression, and standard normalization techniques are not effective. We use Levenberg-Marquardt minimization, and the Jacobian of the minimization [12, Ch. 15] as the basis for approximating the covariance matrix. In particular, if  $\mathbf{J}_{\hat{\boldsymbol{\theta}}}$  is the Jacobian, the covariance is approximately  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = (\frac{1}{\sigma^2} \mathbf{J}_{\hat{\boldsymbol{\theta}}}^T \mathbf{J}_{\hat{\boldsymbol{\theta}}})^{-1}$ . The pseudo-inverse is used when there are constraints on the estimate that make it not full-

<sup>2</sup>The distinction between the normal distance for face points and Euclidean distance for corner points is crucial. For example, if Euclidean distances were used for face points, then there would be no apparently uncertainty in the alignment of a line from the two images.

rank. The approach works with the homography combined with radial-lens distortion.

Overall, this minimization process is applied to estimate the mapping from  $I_p$  to  $I_q$  by reversing the roles of the feature sets but keeping the correspondences, from  $I_q$  to  $I_p$ .

### 5.3 Model Selection Criterion

Due to the characteristics of the region growing and the formation of new match sets, we only consider switching from lower order models to high order, more sophisticated models. This must be done carefully: switching to a higher-order model too soon causes the estimate to be distorted by noise; switching too late causes an increase in error variances  $\sigma_c^2$  and  $\sigma_f^2$  and misalignment on the bootstrap region boundary. Model selection techniques have been studied extensively in the literature. In our work we have found that the earlier and quite simple yet effective Akaike Information Criteria, with a small-sample bias adjustment as recommended in [5]:

$$-|\mathcal{C}_c| \log(\sigma_c) - |\mathcal{C}_f| \log(\sigma_f) - E(\hat{\boldsymbol{\theta}}; \mathcal{C}_c, \mathcal{C}_f) + \frac{nk}{n-k-1}, \quad (2)$$

where  $k$  is the degrees of freedom in current model and  $n = 2|\mathcal{C}_c| + |\mathcal{C}_f|$  is the effective number of constraints.

Expression (2) is evaluated for each model using fixed match sets after IRLS is applied for each model as described earlier. The final error distances,  $d_c$  or  $d_f$ , of each correspondence are then used to evaluate expression (2) for each model. The model that minimizes (2) is then selected for the next iteration of the Dual-Bootstrap.

### 5.4 Region Growth

Region growth depends on the uncertainty in the transformation estimate, as represented by covariance matrix  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ . In particular, expansion is inversely proportional to the transfer error — the error in applying the estimated transformation to points on the boundary of the bootstrap region. If  $\mathbf{x}$  is such a point location (not a feature point) and  $\mathbf{J}$  is the Jacobian of transformation  $\mathbf{T}$  with respect to  $\boldsymbol{\theta}$ , evaluated at  $\mathbf{x}$ , then the error covariance of the mapping of  $\mathbf{x}$  is  $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathbf{J}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\mathbf{J}^T$ . Suppose now that  $\mathbf{x}$  is chosen at the center of a side of bootstrap region  $R$  and the outward direction (away from the center of the region) is  $\mathbf{n}$ . Then the mapping error variance in the outward direction is  $\mathbf{n}^T\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{n}$ . Finally, this side of the rectangle expands outward inversely proportional to  $\mathbf{n}^T\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{n}$ , which means that more uncertainty leads to slower growth. This is applied independently to each side of the rectangle.

## 6 Decision Criteria

Once the iterative Dual-Bootstrap procedure just described expands to cover the apparent overlap between images (based on the estimated transformation), and the refinement process has converged, the final alignment must be tested for correctness. As discussed above, if this confirms that the transformation is correct, the images are considered to be aligned, and the algorithm stops. Otherwise, the next keypoint match is tested using the initial refinement and Dual-Bootstrap procedures.

The decision criteria is composed of two parts: accuracy and consistency. The accuracy,  $\tau$ , is measured as the weighted average alignment error for the final match set, using the weights and distance measures defined above, and only face points because these are more accurately positioned. Consistency,  $\rho$ , is harder to define because of differences between image modalities and illuminations, and because of scene changes.

The measure we use is the absolute difference in normal directions (measured as an angle) between face point correspondences  $(\mathbf{p}'_{f,j}, \mathbf{q}_{f,j})$ , where  $\mathbf{p}'_{f,j}$  is the transformed point. We calculate a histogram,  $h$  of orientation differences in the range  $[0, \pi/2]$  using all face correspondences. If the transformation is incorrect, this angle difference will tend toward being uniformly distributed within the range; whereas if the images are well-aligned, the histogram will tend to have a strong peak near 0 degrees. We measure this by computing the Bhattacharya measure between  $h$  and a uniform distribution and between  $h$  and an exponential distribution. Denoting these distances as  $b_u$  and  $b_e$ , our consistency measure is the ratio of distances,  $\rho = b_e/b_u$ . Small values of  $\rho$  correspond to well-aligned images.

We use two thresholds for each of the measures:  $T_L < T_H$  for  $\tau$  and  $P_L < P_H$  for  $\rho$ . When  $\tau < T_L$  and  $\rho < P_L$ , the alignment estimate is simply accepted as correct. DB-ICP alignment results in this case are accurate to less than a pixel ( $T_L = 1.0$ ) and no substantial improvement can be made. When  $\tau > T_H$  or  $\rho > P_H$ , the estimate is apparently incorrect and is thrown away. The algorithm then moves to the next keypoint correspondence, as described above. Otherwise the estimate is saved and the next initial match is considered. At the end, the estimate that has the minimum alignment error  $\tau$  is chosen as the final estimate. If no estimate has both  $\tau < T_H$  and  $\rho < P_L$ , the images are left unregistered.

Finally, we can use the decision criteria above as a termination criteria, allowing the algorithm to quickly reject incorrect alignments early in the process. We let the DB-ICP loop run for 3 iterations before testing. In addition to the measures defined above, we also include a measure that detected extreme scale changes between images. The thresholds used are the same for all experiments.

## 7 Experiments

We have applied the keypoint matching, initial estimation and Dual-Bootstrap algorithm just described to the 18 images pairs in our test suite. The images in the suite range in size from  $676 \times 280$  to  $1504 \times 1000$ . It tries up to 100 initial rank-ordered keypoint matches before declaring that the images can not be aligned. Although we test the Lowe keypoints based on a rank-ordering, we do not apply a threshold on the ratio to restrict the number of matches. This is important for aligning medical images because the keypoint matches are not distinct. Finally, for multi-modal images involving contrast reversals, we invert the intensities of one of the images before extracting keypoints.

The algorithm successfully aligns 15 of the 18 pairs, including pairs from Figures 1-3 except “DNWS” from Figure 3. Example checkerboard mosaics showing the alignment results are shown in Figures 8 and 9. We have also applied the new algorithm to other, easier image pairs with universal success. Finally, when tested on image pairs that have no overlap, the new algorithm has not yet falsely aligned a pair. For some image pairs we achieved subpixel accuracy by using homography with radial lens distortion model, whereas visible misalignments are apparent when using the homography only.

To show the significance of these results, the publically-available code for the Autostitch keypoint matching algorithm of [4] produced 1 accurately aligned pair and 4 pairs with visible misalignments; on 13 pairs it failed altogether. (Autostitch was run with the original parameters.) None of images in Figures 1-3 were successfully aligned.

More about the behavior of our algorithm and the causes for the failure of the random-sampling keypoint match algorithm can be understood by using studying the verified alignment for 9 of these pairs. This alignment is used to determine which matches are correct based on mapped location, orientation, and scale. In two image pairs, there were no correct keypoint matches that passed the threshold ratio test. In two others, there were only three. For the remaining pairs, there were sufficient numbers of correct matches, but only 10%-22% of the overall match set. Finding a good random sample with such a small fraction of inliers leads to an exponential growth in the number of samples required.

By contrast, our algorithm usually succeeded with the first correct match it tested. In most cases, the successful match was among the first ten. The worst case was the 34th match. On the challenging winter-summer pair, we let our algorithm run on each of the correct keypoint matches, and found that on 9 out the 11 pairs it produced an alignment that passed the final consistency tests. In general, the Dual-Bootstrap procedure converges in 10-20 iterations when started from a correct match. Finally, the failures of our algorithm may generally be attributed to initialization:



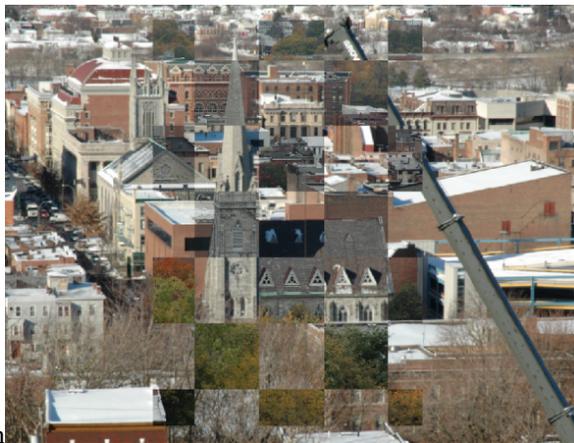
Figure 8: Checkerboard mosaics showing the accuracy of the alignment for the GC-2 and Melanoma pairs from Figures 1 and 2.

on each of the three pairs in our test suite in which the algorithm failed, a simple manual initialization in a small image region, followed by application of the Dual-Bootstrap procedure, led to a consistent alignment.

As a last comment, our algorithm is not as expensive as one would imagine. On the melanoma pair the cost is about 0.25s per initial match, whereas on the larger winter-summer pair the cost is 3.1s per initial match. Aside from image size, the difference in the costs is primarily due to the earlier termination criteria, which is much more effective on melanoma images. All the performance results are measured on a Pentium IV 3.2GHz PC.

## 8 Summary and Conclusion

We have presented an algorithm designed to register a wide variety of images, and analyzed it on a challenging suite of test image pairs. The crucial properties of the algorithm include (1) keypoint matching, (2) generating and testing similarity transformations based on a single keypoint match,



inin

Figure 9: Checkerboard mosaics showing the accuracy of the alignment for the Day-Night and Winter-Summer pairs from Figure 3.

and (3) growing and refining alignments using a combination of symmetric feature matching, robust re-estimation of the transformation, model selection, and region growing. A consistency test successfully determines which transformation estimates are correct and can be used to quickly eliminate obviously incorrect estimates. The algorithm successfully aligned 15 of the 18 pairs in the challenge suite, substantially outperforming a recent algorithm based on keypoint matching alone. The algorithm routinely handles a wide variety of much easier image pairs.

The algorithm works effectively when at least one keypoint match is correct and when there is sufficient consistent structure between the images to drive the Dual-Bootstrap procedure — even when much of the structure is inconsistent due to physical and illumination changes or differences in modality. The algorithm fails primarily when there is

no keypoint match to gain an initial toe-hold on the correct alignment.

Our future work is headed in several directions. First, we are re-addressing the initialization problem — the most important issue for improving the performance of this algorithm in particular and registration techniques in general. Second, we are pursuing a variety of applications. Third, we are generalizing it to a multi-image registration algorithm.

## References

- [1] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004.
- [2] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. Second ECCV*, pages 237–252, 1992.
- [3] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE T. Pattern Anal.*, 14(2):239–256, 1992.
- [4] M. Brown and D. Lowe. Recognising panoramas. In *Proc. ICCV*, 2003.
- [5] K. P. Burnham and D. R. Anderson. *Model Selection and Inference: A practical Information-theoretic Approach*. Springer, 1st edition, 1998.
- [6] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *IVC*, 10(3):145–155, 1992.
- [7] C. Harris and M. Stephens. A combined corner and edge detector. pages 147–151.
- [8] M. Irani and P. Anandan. Robust multisensor image alignment. In *Proc. ICCV*, pages 959–966, 1998.
- [9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, November 2004.
- [10] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging.*, 16(2):187–198, 1997.
- [11] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.
- [12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
- [13] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. Third Int. Conf. on 3-D Digital Imaging and Modeling*, pages 224–231, 2001.
- [14] H. Sawhney, S. Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. In *Proc. 5th ECCV*, volume II, pages 103–119, 1998.
- [15] D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging.*, 21(11):1421–1439, 2002.
- [16] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *IJCV*, 24(2):137–154, 1997.