Automatic robust image registration system: Initialization, estimation, and decision.

Gehua Yang[1] Charles V. Stewart[1] Michal Sofka[1] Chia-Ling Tsai[1,2]
[1] Rensselaer Polytechnic Institute [2]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 highly-reliable, fully-automated image registration technique that takes two images and correctly aligns them or decides that they can not be aligned. The technique should handle image pairs having low overlap, variations in scale, large illumination differences (e.g. day and night), substantial scene changes, and different modalities. Our approach is a combination of algorithms for initialization, estimation and refinement, and decision-making. It starts by extracting and matching keypoints. Rank-ordered matches are tested individually in succession. Each is used to generate a similarity transformation estimate in a small region of each image surrounding the matched keypoints. A generalization of the recently developed Dual-Bootstrap algorithm is then applied to generate an image-wide transformation estimate through a combination of matching and re-estimation, model selection, and region growing, all driven by a new multiscale feature extraction technique. After convergence of the Dual-Bootstrap, the transformation is accepted if it passes a correctness test that combines measures of accuracy, stability and non-randomness; otherwise the process starts over with the next keypoint match. Experimental results on a suite of challenging image pairs shows the effectivenss of the complete system.

1 Introduction

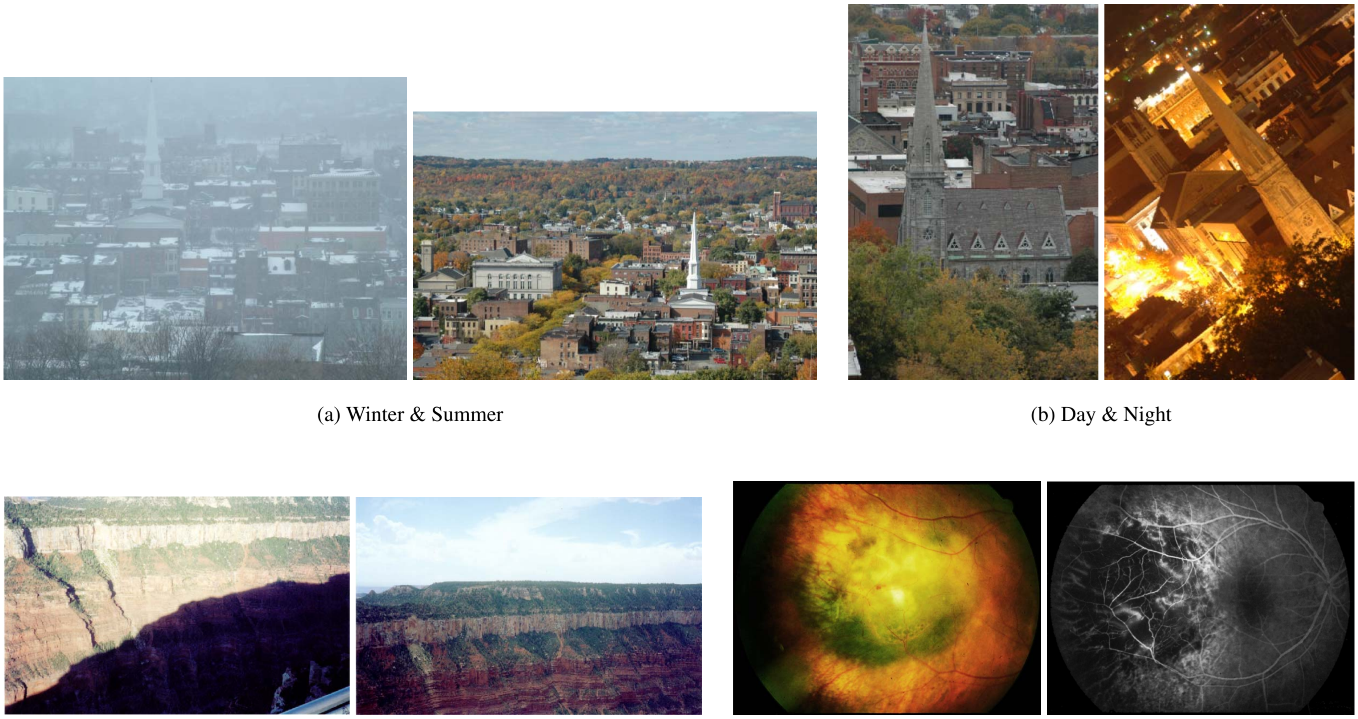
This paper addresses the problem of developing an image registration algorithm that can work on many different types of images, scenes, and illumination conditions. Many of these are illustrated in Figure 1. The algorithm should successfully align pairs of images taken of indoor or outdoor scenes, and 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. It should handle low image overlap and substantial differences in orientation and scale between images. It should be able to align images with high accuracy. Finally, the algorithm should be able to indicate when two images *can not* be aligned either because the images truly do not overlap or because there is insufficient information to determine an accurate, reliable transformation between images. Such a registration algorithm will have numerous applications, but perhaps more importantly will also provide an important milestone toward the development of fully-automatic computer vision systems.

Three primary technical challenges must be addressed to solve this problem: initialization, estimation, and decision.

- While initialization is not a significant problem for aligning images in a video sequence or for multimodal registration of images taken from roughly pre-aligned sensors, it is a major concern for more general-purpose registration. Tolerating a wide range of position, scale and orientation changes implies that simple initialization methods such as just using the identity transformation or using a coarse sampling of parameter space are unrealistic.
- The estimation process must tolerate position, scale and orientation differences, while producing an accurate image alignment. Moreover, estimation must accommodate the possibility that there is no relationship between the intensities for a large fraction of the image pixels. For example, for the winter & summer pair in Figure 1, the leaves are gone and snow has appeared in the winter image, changing both image texture and color. Because of this, an effective estimation technique should automatically and adaptively focus on what is consistent between the images.
- A decision criteria is required that not only chooses between different estimates obtained from different starting conditions, but also decides when the images may not be aligned at all. All of the conditions outlined in the previous two items must be addressed in the design of the decision criteria as well.

Many registration algorithms have been published in the computer vision and related literature. Most of these are fo-

Proceedings of the Fourth IEEE International Conference on Computer Vision Systems (ICVS 2006)
0-7695-2506-7/06 $20.00 © 2006 IEEE



(a) Winter & Summer

(b) Day & Night

(c) Grand Canyon

(d) Melanoma

Figure 1: Image pairs illustrating some of the challenges of general-purpose registration.

cused on the estimation process given an initial estimate. This includes mutual information techniques [10], which have been used widely in multimodal medical image registration problems, and direct methods [1, 2] that align images based on minimizing differences in intensity or some other pixel-by-pixel measure [7]. Neither includes an initialization technique and neither includes a decision criteria. Moreover, significant challenges must be overcome in adapting these measures to the more general-purpose scenario. For example, mutual information is susceptible to local minima and sometimes produces an objective function minimum at an incorrect alignment [10]. Direct methods can not handle substantial scale and orientation differences between images. Both of these can be handled by a sufficiently reliable and accurate initialization method, but such methods do not exist. The closest algorithm is the keypoint matching and random sampling algorithm of [3] but, as we will soon see, this relies too heavily on having a sufficient number of correctly matched keypoints.

1.1 Approach

Our proposed registration algorithm is a combination of existing algorithms (and implementations) and novel techniques. Initialization is based on keypoint extraction and matching, but keypoint matches are considered *individually* to generate an initial transformation estimate accurate in only a small image region. Estimation uses the Dual-Bootstrap ICP algorithm introduced in our earlier work on retinal image registration [13], but now driven by multiscale features extracted using a novel algorithm which is adaptive to local image content. Keypoint matches are tested by the Dual-Bootstrap one-by-one until one resulting estimate passes the decision criteria or all of the top matches are exhausted. Our novel decision criteria is based on a test combining measures of accuracy and non-randomness. Here is a procedure outline of our method:

1. Extract corners and faces
2. Extract keypoints and match through indexing
3. Get next highest ranked keypoint match
4. DBICP refinement
   - i. Generate matches
   - ii. Parameter estimation for each candidate model
   - iii. Model selection
   - iv. Region growing

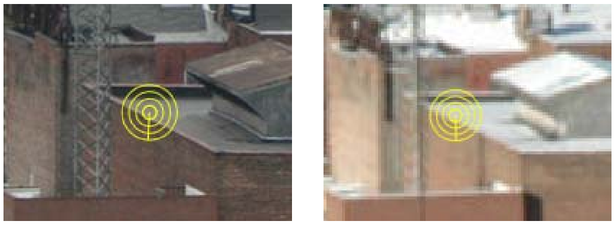


Figure 2: Initial keypoint match and side-by-side alignment for one of our winter-summer pairs. The image region on the right has been scaled by a factor of 2.25, and there are substantial illumination, blurring, and changes (snow) between the regions.

   v. If early termination criteria is met, go to 3

   vi. Go to 4i until region covers the overlap

5. Check decision criteria. If not met, go to 3

2 Initialization

Similar to several recent techniques (e.g. [3], we start by extracting keypoints from each image and matching them. We use Lowe's implementation of his keypoint extraction algorithm and SIFT descriptor [8] and reimplemented his matching algorithm. (Similar results are obtained with [9].) At this point, many techniques for registration and fundamental matrix estimation apply random sampling techniques to search for combinations of matches that generate an estimate optimizing a robust error term. We have found, however, that on challenging image pairs there are too few correct matches, both in terms of percentage and absolute number, for this to succeed (Section 5). This reflects the overall difficulty of initialization. We handle this by expecting less of initialization and designing the estimation algorithm to "bootstrap" the alignment from a weak initialization.

In particular, each keypoint match is used separately, starting with the highest ranking match, to generate an initial similarity transformation. This transformation is established from the positions of the two keypoints, the orientations of their dominant intensity gradients, and their relative scales (Figure 2). A small initial region is established around the matching points in each image. The Dual-Bootstrap procedure starts from this initial estimate and region.

3 DB-ICP

As outlined above, the Dual-Bootstrap ICP (DB-ICP) algorithm iterates steps of (1) refining the current transformation in the current "bootstrap" region R, (2) applying model selection to determine if a more sophisticated model may be used, and (3) expanding region R, growing inversely proportional to the uncertainty of the mapping on the region boundary. Since this algorithm is described in full elsewhere for registering retinal images, we concentrate here on what is needed to generalize this algorithm for a wider class of images. This includes a new technique for multiscale feature extraction, new matching and estimation techniques, and a simplified model selection algorithm. For more details, especially on region growing, see [13].

3.1 Feature Extraction

The most important innovation is our feature extraction method. The original Dual-Bootstrap is driven by points extracted along blood vessel centerlines. In a more general setting, we need features that will be applicable to wider range of images. The advantage of advocating image features are (a) matching features provides direct measurement of the geometric alignment error needed for the region growth and model selection processes and (b) in our setting even though much of the image texture may change between images being aligned, structural outlines usually remain unchanged. These should be captured by properly extracted features. Moreover, the Dual-Bootstrap matching and robust estimation techniques (implicitly) determine the features that are consistent between images, using these to drive registration, while ignoring what is inconsistent.

Two different feature types are located — corners and face points — and these are detected to subpixel accuracy at multiple scales in half-octave steps. No attempt is made to combine features across scales, and all scales are used simultaneously during registration. This is important for several reasons, including aligning images having significant scale differences. The remainder of this section considers feature extraction at a single scale.

At each pixel $\mathbf{x}$, the intensity gradient, $\nabla I(\mathbf{x})$, is computed. A weighted neighborhood outer product matrix is then computed,

$$\mathbf{M}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w(\mathbf{x} - \mathbf{y}) \; (\nabla I(\mathbf{y}))(\nabla I(\mathbf{y}))^{\top}.$$

The eigen-decomposition is computed, $\mathbf{M}(\mathbf{x}) = \sum_{i=1,2} \lambda_i(\mathbf{x}) \boldsymbol{\Gamma}_i(\mathbf{x}) \boldsymbol{\Gamma}_i(\mathbf{x})^{\top}$, with $\lambda_1(\mathbf{x}) \leq \lambda_2(\mathbf{x})$. Potential *corners* are located at pixels where $\lambda_1(\mathbf{x})/\lambda_2(\mathbf{x}) > t_a$. This criterion is similar to the Harris corner detector [5]. Potential *face points* are located at pixels for which $\lambda_1(\mathbf{x})/\lambda_2(\mathbf{x}) \leq t_a$. Decision value t_a has been experimentally set to 0.1, although the choice of values is not crucial. Strength is assigned to each feature as $s(\mathbf{x}) = \mathrm{trace}(\mathbf{M}(\mathbf{x}))$.

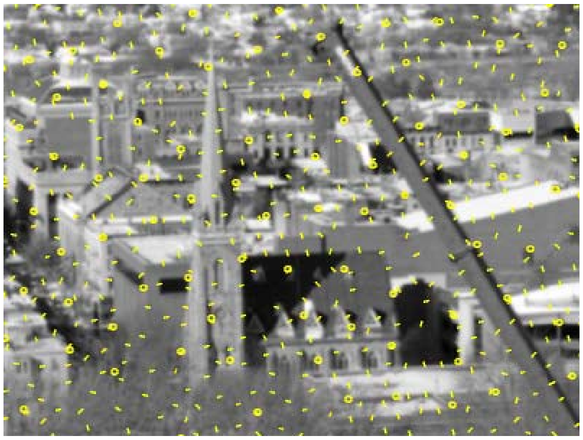


Figure 3: Example intermediate resolution driving features, which are more sparse than matchable features. Circles are corners and line segments are face points, oriented along the direction of greatest eigenvalue.

The next steps are designed to make the selection feature adaptive to image content. (1) A very low threshold, $t_s = 1$, is applied to the strength to eliminate plainly noise points. (2) The median and robust standard deviation of the strength values are computed in overlapping neighborhoods throughout the image. At each pixel, if the strength is below the local median plus a half standard deviation, it is eliminated. (3) Non-maximum suppression is then applied at each pixel for corners and faces separately. (4) Each remaining pixel is tested (faces and corners separately) in order of decreasing strength to ensure that it has locally largest strength and it is not close to other features. Pixels passing this test become 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. The resulting features are called *matchable features*. (5) The 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 [12]). An example is shown in Figure 3. In the matching process, driving features are transformed and matched against matchable features.

3.2 Refinement Within the Bootstrap Region

The first step of DB-ICP is refinement of the transformation estimate within the current bootstrap region R, ignoring everything else in the two images. The 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 Dual-Bootstrap proceeds to model selection and region growing before selecting a new set of matches.

Matching is applied from image I_p to image 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 $\mathbf{p}' = \mathbf{T}(\mathbf{p}; \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the current estimate of the transformation parameters. 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 similarity in scales and (in the case of faces) orientations following application of $\mathbf{T}$. 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.

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

Before constructing the transformation estimation objective function, we define the error distances

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

for corners and face points, respectively. Using this, for a fixed set of matches and weights, the transformation can be re-estimated by minimizing

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

where $w_{s;i}$ is similarity weight and $w_{d;i}$ 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 σ^2 is the variance. In particular, σ_c and σ_f are the robustly computed error variances for corner points and face points respectively. Weight $w_{s;i}$ measures the similarity in scale and (for faces) orientation between matched features (following application of the transformation).

Estimating $\hat{\boldsymbol{\theta}}$ is carried out by iterating step of minimizing (1) and re-estimating the robust weights w_d. Minimizing (1) is straightforward for affine and similarity transformation. When estimating the parameters of a homography or a homography plus radial lens distortion model, we use Levenberg-Marquardt minimization. The Jacobian of the minimization is the basis for approximating the covariance matrix [11, Ch. 15] needed for region growing and for the stability component of the decision criteria.

This minimization process is applied to estimate the mapping from I_q to I_p by reversing the roles of the feature sets but keeping the same correspondences.

0-7695-2506-7/06 $20.00 © 2006 IEEE

3.3 Model Selection Criterion

As the region grows and more constraints are incorporated, higher-order models can be used. The decision to switch to a higher-order model is made using a model selection technique. Model selection transitions from similarity to affine to homography, and in some cases to a homography plus radial-lens distortion. For retinal images, the final model is a quadratic transformation. Since the region grows monotonically, we only consider switching from lower-order to higher-order models. In the generalized Dual-Bootstrap we now use the relatively simple Akaike Information Criteria with a small-sample bias adjustment as recommended in [4]:

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

where k is the degrees of freedom in current model and $n = 2 \mid \mathcal{C}_c \mid + \mid \mathcal{C}_f \mid$ is the effective number of constraints. Expression (2) is evaluated for the current and higher order models using fixed match sets after IRLS is applied for each model (as described above). The model that minimizes (2) is then selected for the next iteration of the Dual-Bootstrap.

4 Decision Criteria

Once the Dual-Bootstrap procedure expands to cover the apparent overlap between images (based on the estimated transformation), and the refinement process has converged, the procedure ends and the final alignment is tested. As discussed above, if this confirms that the transformation is acceptable, the images are considered to be aligned, and the overall registration procedure terminates. Otherwise, the next keypoint match is tested using the Dual-Bootstrap. This ends in failure when a maximum number of matches is unsuccessfully tested.

The decision criteria is composed of three parts — accuracy, stability and consistency — each of which must pass. Accuracy 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. Stability is measured as the transfer error covariance [6, Ch 4] [13] for the mapping of points on the boundary of the final bootstrap region. If there is high variance in this mapping, the mapping is unstable. Both accuracy and stability are measured against user-defined thresholds which have simple, intuitive meaning. Finally, the consistency measure we use is a histogram of the absolute difference in normal directions (measured as an angle) between face point correspondences following alignment. If this histogram is substantially closer to an exponential distribution — i.e. very low orientation differences — than to a uniform distribution, as determined by the Bhattacharyya measure, then the consistency test passes. As a final note, we use this three-part decision criteria as an earlier termination criteria as well, allowing the algorithm to quickly reject incorrect alignments early in the process.

5 Experiments

We have applied the overall registration algorithm just described to the 18 images pairs in our test suite.[1]. The images in the suite range in size from 676×280 to 1504×1000. The algorithm tried up to 100 initial rank-ordered keypoint matches before declaring that the images could not be aligned. It successfully aligned 15 of the 18 pairs, most to subpixel accuracy. To achieve accurate alignment, a transformation model with a homography plus second-order radial distortion is required for 3 pairs, a homography for 10 pairs, and for the final 2 — both retinal image pairs — a quadratic model. Even when there were significant physical changes in the scene, manual inspection showed no visible misalignments for any of the 15 pairs. Example checkerboard mosaics showing the alignment results are shown in Figures 4. In 2 of the 3 cases of failure, no accurate initial keypoint matches were generated. Moreover, when all possible pairs were formed from non-overlapping images from the test suitet, the new technique generated no false alignments. Finally, looking at the behavior of the algorithm system in more detail, experiments showed that about 80% of the time the initial match was roughly correct, the Dual-Bootstrap procedure grew it into an image-wide alignment that the decision criteria accepted as correct.

To reinforce the significance of these results, the publically-available code for the Autostitch keypoint matching algorithm of [3] produced 1 accurately aligned pair and 4 pairs with visible misalignments; on 13 pairs it failed altogether. Autostich was run with the original parameters.

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.

6 Summary and Conclusion

We have presented a complete system for registering pairs of images and analyzed it on a challenging suite of test

[1] The images and the executable are available through `http://www.cs.rpi.edu/research/groups/vision/gdbicp/`

image pairs. This system is built on three main algorithmic components: keypoint matching to generate initial estimates in small image regions, the Dual-Bootstrap procedure for growing and refining initial estimates, and a three-part decision criteria measuring accuracy, stability and randomness. The Dual-Bootstrap procedure, originally designed for retinal image registration, was generalized here by using new, generic, multiscale feature extraction, new matching and estimation techniques, and a simplified model-selection criteria. The technique successfully aligned 15 of the 18 pairs in the challenge suite, and does not falsely align non-overlapping images. Overall, it 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. Thus, although we have reduced the importance of initialization to producing a single correct keypoint match, initialization still remains the most challenging problem for general-purpose registration.

References

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

[2] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. Second ECCV*, pages 237–252, 1992.

[3] M. Brown and D. Lowe. Recognising panoramas. In *Proc. ICCV*, 2003.

[4] K. P. Burnham and D. R. Anderson. *Model Selection and Inference: A practical Information-theorectic Approach*. Springer, 1st edition, 1998.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, Manchester, UK, 1988.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.

[7] M. Irani and P. Anandan. Robust multisensor image alignment. In *Proc. ICCV*, pages 959–966, 1998.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, November 2004.

[9] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

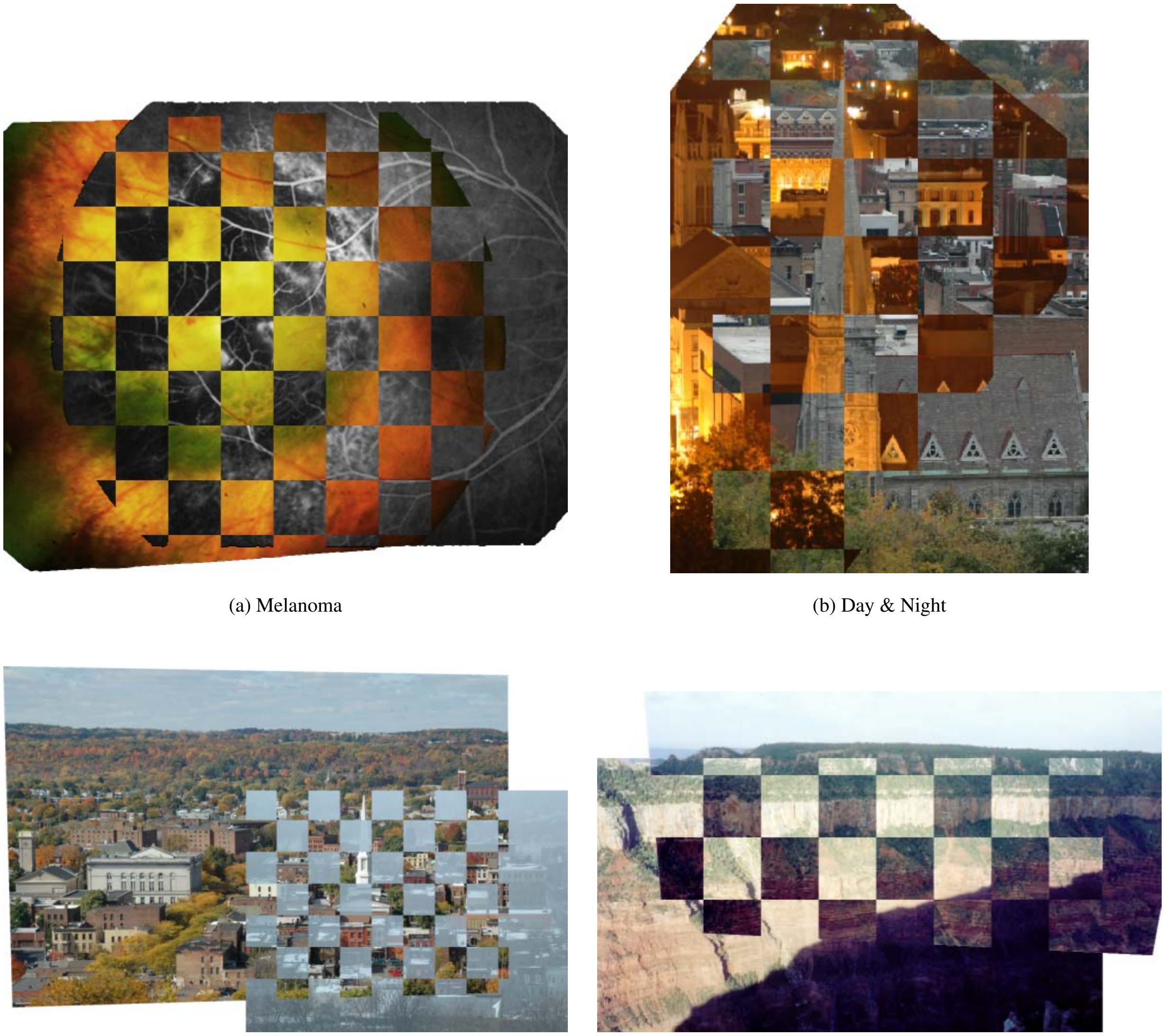
[10] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imaging.*, 22(8):986–1004, 2003.

[11] 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.

[12] D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging.*, 21(11):1421–1439, 2002.

[13] C. Stewart, C.-L. Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans. Med. Imaging.*, 22(11):1379–1394, 2003.

0-7695-2506-7/06 $20.00 © 2006 IEEE



(a) Melanoma

(b) Day & Night

(c) Winter & Summer

(d) Grand Canyon

Figure 4: Checkerboard mosaics showing the alignment accuracy for the pairs from Figure 1.