Registration of challenging image pairs: initialization, estimation, and decision

Tech Report # 05-19

Gehua Yang, Charles V. Stewart, Michal Sofka, and Chia-Ling Tsai

The authors are with the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, 12180, USA. Emails are
{*yangg2,stewart,sofka,tsaic*}*@cs.rpi.edu*

Abstract

Our goal is an automated 2d-image-pair registration algorithm capable of aligning images taken of a wide variety of natural and man-made scenes as well as many medical images. The algorithm should handle low overlap, substantial orientation and scale differences, large illumination variations, and physical changes in the scene. An important component of this is the ability to automatically reject pairs that have no overlap or too many differences to be aligned well.

We propose a complete algorithm, including techniques for initialization, for estimating transformation parameters and for automatically deciding if an estimate is correct. Keypoints extracted and matched between images are used to generate initial similarity transform estimates, each accurate over a small region. These rank-ordered initial estimates are tested individually in succession. Each estimate is refined using the Dual-Bootstrap ICP algorithm, driven by matching of multiscale features. A three-part decision criteria, combining measurements of alignment accuracy, stability in the estimate, and consistency in the constraints, determines whether the refined transformation estimate is accepted as correct. Experimental results on a suite of 22 challenging image pairs show that the algorithm effectively aligns 19 of the 22 pairs and rejects 99.8% misalignments when all possible pairs are tried. The algorithm substantially out-performs algorithms based on keypoint matching alone.

Index terms: Image registration, feature extraction, iterative closest point, radial lens distortion, decision criteria, keypoint

I. INTRODUCTION

This paper addresses the problem of developing an automated image-pair registration algorithm that can work on a wide variety of image types, scenes, and illumination conditions. Much of this variety is captured in a test suite of 22 image pairs we have gathered, some of which are shown in Figure 1.[1] The suite includes image pairs taken of indoor and outdoor scenes, in natural and man-made environments, at different times of day, during different seasons of the year, and using different imaging modalities. It includes image pairs with low overlap (e.g. 2%), substantial differences in orientation (90 degrees), and large changes in scale (up to a factor of 6.4). A general-purpose registration algorithm should be able to align each of these image pairs with high accuracy. Moreover, such an algorithm should be able to indicate that two images *can not* be aligned either when the images truly do not overlap or when there is insufficient

---

[1]Both the test suite and our software are available at `http://www.vision.cs.rpi.edu/gdbicp/`

information to determine an accurate, reliable transformation between images. Such a registration algorithm will have numerous applications ranging from mosaic construction to change detection and visualization.

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

- While automatic 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.

- In combination with initialization, the estimation process must tolerate position, scale, orientation and illumination differences. Moreover, estimation must accommodate the possibility that there is no relationship between the intensities for a large fraction of the pixels in the two images. For example, in the Summer-Winter pair from Figure 1, snow on the roofs in winter produces homogeneous intensity regions, whereas these roofs appear as dark, textured regions in the summer image (Figure 4). Because of this, an effective estimation technique should automatically and adaptively exploit what is consistent between the images.

- Decision criteria are required not only to choose among different estimates obtained from different starting conditions, but also to decide when the images may not be aligned at all. The need for effective decision criteria is particularly acute when handling low overlap and large changes in orientation, illumination and scale due to the extremely large search space of initial estimates.

A. Related Techniques in Registration

The literature on image registration is large; see [6], [25], [55] for reviews. We focus our attention here on four classes of methods that appear most appropriate for the general-purpose registration problem being addressed.

1) Feature-Based Methods and ICP: Our first class of prior techniques is feature-based methods and the Iterative Closest Point (ICP) algorithm [5], [12], [13]. Starting from an initial estimate, the ICP algorithm iteratively (a) maps points (features) from the moving image to the fixed image, (b) finds the closest fixed image point for each mapped point, and (c) re-estimates the transformation based on these temporary correspondences. In registration of range data, the sensor data points (together with estimated normals) are the features. In alignment of intensity

images, feature extraction is applied to obtain the points matched in ICP [11], [22]. There are two common problems with feature-based methods and ICP: (1) ICP has a narrow domain of convergence and therefore must be initialized relatively accurately; and (2) feature extraction can be unreliable and overly sensitive to the choice of parameters and the image content. (We will address both issues.) Finally, several papers [14], [21], [32] have proposed Expectation Maximization (EM) algorithms where multiple correspondences per feature are simultaneously considered. While this approach can improve the robustness and reliability of ICP, it has not been shown to work on data sets as difficult as our test suite. We discuss this further in Section VII.

2) Keypoint Indexing Methods: While ICP correspondences are based on geometric distances, keypoint correspondences are based on appearance similarity measures. Keypoint methods have received growing attention because of their demonstrated ability to tolerate low image overlap and image scale changes. These methods start with keypoint detection and localization followed by computation of the descriptor that summarizes the image around the keypoint. Existing extraction algorithms are based on approaches ranging from the Laplacian-of-Gaussian operator [7], information theory [28], Harris corners [37], and intensity region stability measures [34]. A detailed comparison can be found in [39]. Region descriptors proposed are based on steerable filters [19], moment invariants [20], shape context [3], image gradients [31] and Haar wavelets [8]. They are usually invariant to 2d similarity or affine transformations of the image, as well as linear changes in intensity. A comparison of descriptors can be found in [38].

Keypoint matches have been applied to fundamental matrix estimation [34], stereo [44], [53], registration [7], and recognition [17], [16]. Often these techniques use minimum-subset random-sampling methods [7], [34]. (In essence, this combines our initialization and estimation steps.) Statistical tests of randomness based on the numbers of overall and consistent keypoint matches are then used to decide if the transformation is correct [7] (the decision criterion) . While this approach has all the steps needed for a general-purpose registration system, it has important limitations for the types of image pairs considered in this paper. Most importantly, experiments on our test suite show that keypoint indexing and matching methods only produce a small number of correct matches, occasionally none and sometimes fewer than 10 out of the top 100 matches. Finding an accurate estimate in this case is either impossible, or requires an expensive, combinatorial search of possible match sets. It is possible that more sophisticated keypoint detection, description and matching methods will make this overall approach viable for such

challenging image pairs, but this is not the direction taken in this paper.

3) Mutual Information: The third class of methods is based on mutual information (MI) [33], [54]. MI registration is primarily designed for multimodal alignment, and has been used extensively in medical imaging applications [42], but also in other areas as well [18]. For the current problem it is less useful than it might first appear. First, algorithms based on MI [18], [41] are only estimation techniques and include neither initialization nor decision techniques. The former is often not a problem in medical applications, and the latter is often left to human judgment. Moreover, MI minimization procedures are often quite sensitive to local minima, a crucial concern when aligning low-overlap image pairs.

4) Direct Methods: A fourth class of techniques involves the direct minimization of intensity differences between images [1]. These methods work by using intensity differences and image gradients to compute an update to the estimate of the aligning transformation and then applying this update to warp one image on top of the other. Multiresolution techniques are used to accommodate larger initial misalignments [4]. The approach has been extended to handle multisensor data by computing and normalizing derivatives in four directions as the "intensity" at each pixel [27]. These techniques require good initialization, although coarse searches of the translation-only parameter space have sometimes been proven effective [43]. Moreover, direct methods have not yet been demonstrated on registration problems involving substantial scale and orientation differences between images. Finally, no decision criterion is associated with these methods.

B. Approach

We propose an end-to-end registration algorithm — actually a system of algorithms — for aligning pairs of images using parametric global transformation models, building heavily on existing techniques. We use our test suite both to carefully evaluate the main components of our algorithm and to demonstrate its overall effectiveness. The primary novelties of our work are in the construction of the overall algorithm, the design and interrelationship of the components, and the success of the system and the components. As an illustration of this, the alignment process is driven by image features extracted using auto-correlation matrices [45], which have been widely used in Harris corner detectors and its variants [2], [23], [37]. Our implementation produces edge-like as well as corner-like features distributed throughout the images, even in extremely low-contrast regions. By stressing dense coverage of features in images we ensure that features

will nearly always be available, trusting the robustness of the matching and estimation algorithms, as well as the decision criteria, to avoid the effects of inconsistencies between images. Thus, the novelty is in tailoring feature extraction for general-purpose registration rather than feature extraction per se.

As mentioned earlier, the three primary components of the system are the initialization algorithm, the estimation technique, and the decision criteria:

- The initialization method uses extraction and matching of keypoints [31]. But unlike current techniques it does not attempt to combine matches in any way. Instead, each match is used to generate an initial similarity transformation that is accurate only in a small region surrounding the matched keypoints from each image. There are two reasons for this. First, as suggested above, for challenging image pairs a relatively small fraction of keypoint matches is correct — too small for effective use of minimum-subset random sampling search techniques. Second, for the keypoint matches that are correct, our growth-and-refinement based estimation procedure usually aligns the images accurately. Both reasons will be illustrated experimentally.

- The estimation technique starts with the initial local regions and associated transformation estimates, treating each one individually and in succession. The goal for each region and estimate is to expand the region to cover the entire overlap between images while refining the estimate each time the region changes. This can be thought of as keeping the estimate close to the optimum as the problem grows in complexity. At the same time, as more information is incorporated, it may be possible to switch to a higher-order transformation model that more accurately describes the mapping of larger image regions. These intuitions are realized by generalizing the Dual-Bootstrap ICP algorithm, which we proposed recently for retinal image registration [49]. Several generalizations are needed to make this work. The most important is the use of generic features, as discussed above, in place of detected blood vessels to drive registration. Other enhanced techniques include bidirectional, across-scale matching between images, region growth in both images, use of a more general model selection technique, and an estimation technique that accounts for variations in feature location uncertainty. It is interesting to note that these generalizations allow the new algorithm to handle some retinal image pairs that the original algorithm could not.

- The decision technique determines if an estimate generated by the Dual-Bootstrap algorithm

is indeed a correct alignment of the two images. The technique combines novel decision criteria that measure accuracy, consistency and stability in the alignment.

The overall procedure takes an ordered sequence of initial estimates (generated from keypoint matching) and applies Dual-Bootstrap ICP to each in succession. Following the convergence of each Dual-Bootstrap estimate the decision criteria are applied. If the estimate passes these tests, it is accepted as correct and the two images are considered registered. Otherwise, the process continues to the next initial keypoint estimate. It stops with failure if none succeed. The complete algorithm will be referred to as the Generalized Dual-Bootstrap ICP — GDB-ICP for short.

The remainder of this paper describes the details of GDB-ICP and then evaluates it on our test suite of challenging image pairs (Figure 1). The experimental analysis demonstrates the overall performance of the algorithm and explores the significance of many of the individual components and design decisions. Importantly, this includes testing on not only the image pairs that overlap and should therefore be registered but also all possible image pairs, including those with no overlap. The paper concludes with a discussion of the strengths and limitations of GDB-ICP, and its relationship to other techniques.

II. INITIALIZATION

Our initialization method is based on Lowe's multiscale keypoint detector and SIFT descriptor [31]. This has proven to be the most effective in the experimental evaluation of [36]. We have used both our own implementation and the one used in the above evaluation with approximately equal success. We have also used affine-invariant multiscale Harris corners [37], and again the differences in results were minor. The results reported here used the implementation from [36].

Here is a summary of the initialization method. Each Lowe keypoint is a local maximum of the magnitude of the Laplacian-of-Gaussian operator in both spatial and scale dimensions. A neighborhood is established at each keypoint, with size determined by the keypoint scale and orientation determined by the local gradient direction. The intensity gradient vectors within this neighborhood are collected in histograms to form a 128-component SIFT descriptor vector. All SIFT vectors are stored in a spatial data structure, one for each image. Keypoint matching between images occurs by taking the keypoints from one image and using their descriptors to index into the data structure for the other image. The descriptor distance, measured as the Euclidean distance between vectors, is computed for each candidate match. The two closest
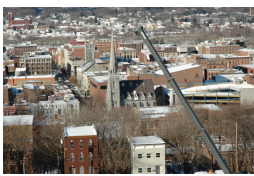
White Tower

Grand Canyon 1

Dashpoint (2% overlap)

Extreme Zoom (factor of 6.4)

Downtown winter day, summer day,
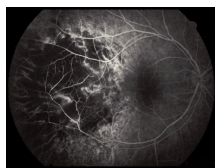
summer night, respectively.

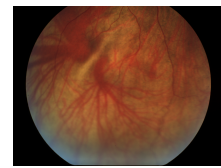Capital Region

Brugge Tower
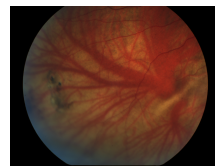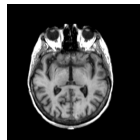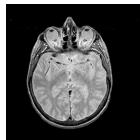

Brugge Square

Boston Library

Brussels

Melanoma

Retina

Brain PD-T1

Grand Canyon 2

Fig. 1

SOME IMAGES FROM OUR 22-IMAGE PAIR DATASET. THE THREE "DOWNTOWN" IMAGES PRODUCE THREE OF OUR TEST

PAIRS.

(a) initialization   (b) 3rd iteration   (c) 6th iteration   (d) 9th iteration
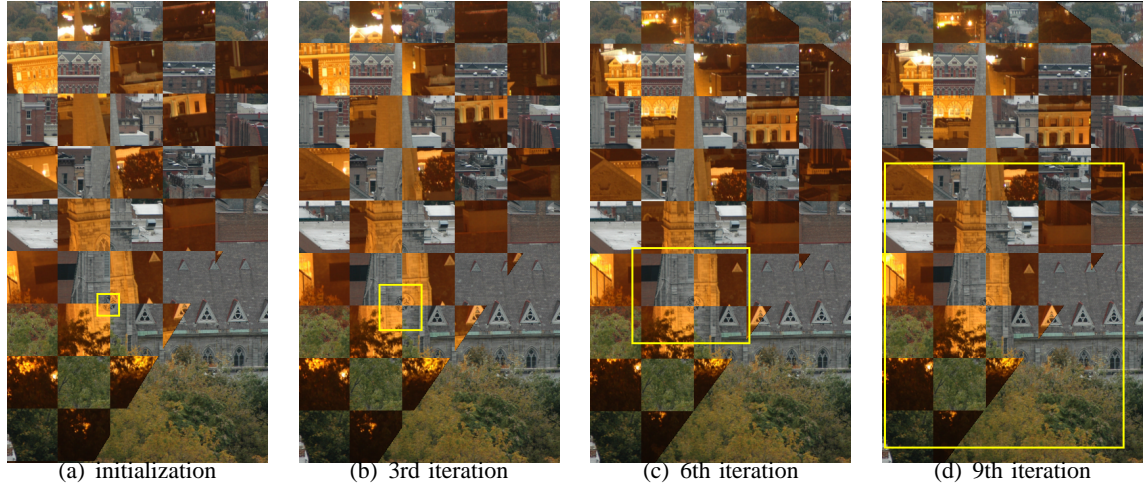
Fig. 2

EXAMPLE STEPS OF THE DUAL-BOOTSTRAP GROWTH AND REFINEMENT PROCESS ON THE DAY-NIGHT SUMMER PAIR. THE CHECKERBOARD IMAGES ABOVE SHOW THE ALIGNMENT RESULTS AND BOOTSTRAP REGION FOR ITERATIONS 0, 3, 6 AND 9. THE YELLOW RECTANGLE OUTLINES THE BOOTSTRAP REGION IN ONE IMAGE. WITHIN THIS REGION, WHICH IS WHERE THE COMPUTATION IS FOCUSED, THE ALIGNMENT IS GENERALLY ACCURATE. OUTSIDE THE REGION, ESPECIALLY FOR THE SMALL REGIONS EARLY IN THE COMPUTATION, THE ALIGNMENT TENDS TO BE INACCURATE. AS THE BOOTSTRAP REGION EXPANDS, MORE AND MORE OF THE IMAGES ARE ACCURATELY ALIGNED. THE FINAL ALIGNMENT IS SHOWN IN FIGURE 10.

matches for each descriptor are found and the ratio of the distances to the closest and second closest is calculated. Lowe establishes a upper bound threshold on this ratio of 0.8, and considers only matches below this threshold. Instead, we sort the matches by the ratio and test the top N in order, stopping when GDB-ICP produces an estimate that the decision criteria accept or when all N matches have been tested unsuccessfully. We have found $N = 50$ to be a good, conservative choice. The initial similarity transformation is established from the positions of the two keypoints, the orientations of their dominant intensity gradients, and their scales (Figure 3). The initial bootstrap region is a square centered at each keypoint location, with half width $30 + 3s_i$, where s_i is the scale of the ith keypoint in the image where it is detected. This width setting ensures that there are enough constraints for Dual-Bootstrap to start successfully for keypoints of various scales. One initial region is generated on each of the two images, centered at the keypoint location. Together with the initial transformation, they are provided as input to

Fig. 3

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.

the Dual-Bootstrap.

One important consideration is why should keypoint descriptors designed only for invariance to linear changes in intensity be useful in multimodal image registration. Indeed, as we will see experimentally, keypoint matching is the least successful component of our algorithm. On the other hand, our algorithm relies on finding only one correct keypoint match, and in generating this match, the descriptors do not have to match exactly. Instead they must only be distinct from other matches. Furthermore, *local* intensity differences between analogous keypoints in different modalities are sometimes well-approximated by linear transformations. These two observations explain why matching of Lowe keypoints successfully generates at least a few correct matches on a variety of multimodal pairs.

III. FEATURE EXTRACTION

The Dual-Bootstrap procedure is feature-driven for two reasons. First, matching image features provides direct measurement of the geometric alignment error. This is needed to compute the statistics that drive the growth, model selection and decision 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 4). These outlines, large or small, can be captured by properly-extracted features.

Fig. 4

EXAMPLES OF SUBSTANTIAL VARIATIONS BETWEEN IMAGE REGIONS DUE TO ILLUMINATION DIFFERENCES, SCALE
DIFFERENCES, AND PHYSICAL CHANGES (SNOW).

As discussed in the introduction, the primary consideration in designing the feature extraction method is not obtaining a complete or a perceptually-significant set of features. It is solely to extract features useful to drive the alignment process. Because of this, our goal is a feature-extraction method that (a) does not depend on thresholds or parameters that must be tuned to individual image content, (b) produces repeatable features and (c) distributes features throughout the image rather than concentrating them in the image regions of highest contrast.

Two different types of features are extracted — corner points and face points. Corner points provide two constraints on registration, whereas face points, because of tangential position ambiguity, provide only one. On the other hand, face points are more densely distributed. We will evaluate the contribution of each feature type to the registration process in the experimental section. Both corners and face points are extracted in scale-space with scale increasing in half-octave steps (e.g. scales $1, \sqrt{2}, 2, 2\sqrt{2}, \ldots$), with no combination across scales. Features at all scales are used simultaneously during registration. The following details of the feature-extraction algorithm are presented for a single scale.

We use a single response measure for both feature types. At each pixel location $\mathbf{x}$ at scale σ, the intensity gradient, $\nabla I(\mathbf{x})$, is computed. A weighted neighborhood outer product (auto-

(a) Initial classification   (b) After adaptive pruning
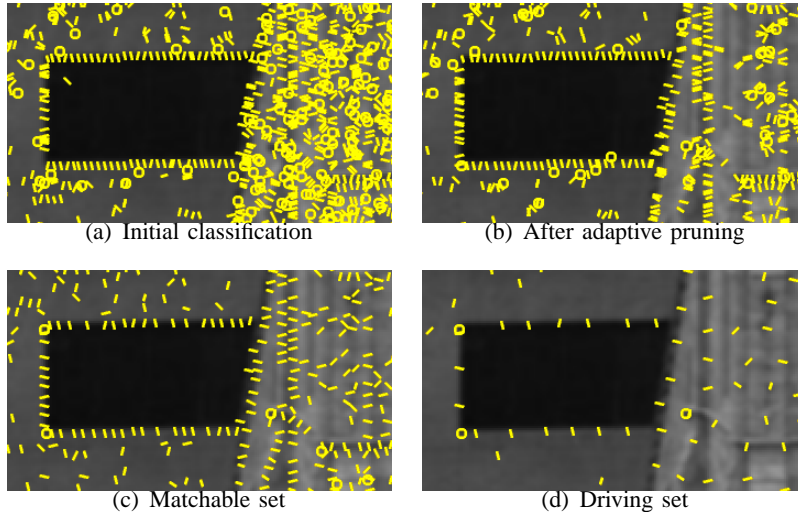
(c) Matchable set   (d) Driving set

Fig. 5

INTERMEDIATE RESULTS OF THE FEATURE EXTRACTION PROCESS: (A) INITIAL CORNERS (SHOWN WITH CIRCLES) AND FACE POINTS (SHOWN WITH LINE SEGMENTS ALONG THE EIGENVECTOR CORRESPONDING TO THE LARGER EIGENVALUE) AFTER APPLYING THE THRESHOLD t_m (ONLY ONE OF EVERY THREE IS SHOWN IN THIS DISPLAY BECAUSE THEY ARE SO DENSE), (B) CORNERS AND FACE POINTS AFTER ADAPTIVE LOCAL PRUNING, (C) "MATCHABLE" CORNERS AND FACE POINTS, AND (D) "DRIVING" CORNERS AND FACE POINTS.

correlation) matrix is then computed,

$$\mathbf{M}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w(\mathbf{x} - \mathbf{y}) \, (\nabla I(\mathbf{y}))(\nabla I(\mathbf{y}))^{\top}, \tag{1}$$

where w is a Gaussian weight function with standard deviation σ and the neighborhood size is approximately 3σ. Next, the eigen-decomposition is computed: $\mathbf{M}(\mathbf{x}) = \sum_{i=1,2} \lambda_i(\mathbf{x})\mathbf{\Gamma}_i(\mathbf{x})\mathbf{\Gamma}_i(\mathbf{x})^{\top}$, with $\lambda_1(\mathbf{x}) \leq \lambda_2(\mathbf{x})$. Potential *corners* are at pixels where $\lambda_1(\mathbf{x})/\lambda_2(\mathbf{x}) > t_a$. This criterion is similar to the Harris corner detector [23]. Potential *face points* are 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. A strength is assigned to each point: $m(\mathbf{x}) = \text{trace}(\mathbf{M}(\mathbf{x}))$.

The next step is designed to make the final selection of feature points adaptive to local image contrast. First, a very low threshold, $t_m = 1$, is applied to the strength to eliminate points that are obviously noise. The result is illustrated in Figure 5(a). The next step, local pruning, starts by computing the median μ_m and median absolute deviation (MAD) [50] σ_m of the values of $m(\mathbf{x})$ in a coarse set of overlapping neighborhoods (30×30 pixels) throughout the image. Then,

pixels with strength $m(\mathbf{x}) < \mu_m + 0.5\sigma_m$ are eliminated from further consideration. As the final preliminary step, non-maximum suppression is applied at each point — in 2D for potential corner points and in 1D, along direction of the eigenvector $\mathbf{\Gamma}_2$, for face points — and surviving point locations are interpolated to subpixel accuracy. Figure 5(b) shows an example result.

The final steps are to extract the actual features, generating corners and face points independently. Points surviving the previous step are sorted by $m(\mathbf{x})$ values. The highest strength point is labelled as a feature. Remaining points within a small neighborhood are eliminated. This continues until a maximum number of features — determined from the size of the images — is extracted or the list is exhausted. This produces a set of features which we call the *matchable features* (Figure 5(c)). The procedure is repeated starting from the noise thresholding step with the elimination neighborhood doubled in size, half the allowed number of features, and a threshold requiring $m(x) > 2t_m$. This produces a set of *driving features* (Figure 5(d)). As will be explained later, driving features are transformed and matched against matchable features, similar to [47]. Since driving features must pass stricter criteria than matchable features, it is less likely that a driving feature in one image will be missed as a matchable feature in the other due to random effects.

An example of extracted driving and matchable feature sets at different scales is shown in Figure 6. Features are spread throughout the image and summarize the local image structure: a corner is placed in regions containing substantial intensity variations in all directions; a face point occurs where the region contains variation in one direction.

IV. GDB-ICP ESTIMATION

The estimation step of the Generalized Dual-Bootstrap ICP (GDB-ICP) algorithm starts with an initial similarity transformation generated from keypoint matching, together with the initial bootstrap region surrounding the keypoint location in each image. The algorithm iterates steps of (1) refining the current transformation inside the current bootstrap region 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 2). The entire algorithm is outlined in Figure 7.
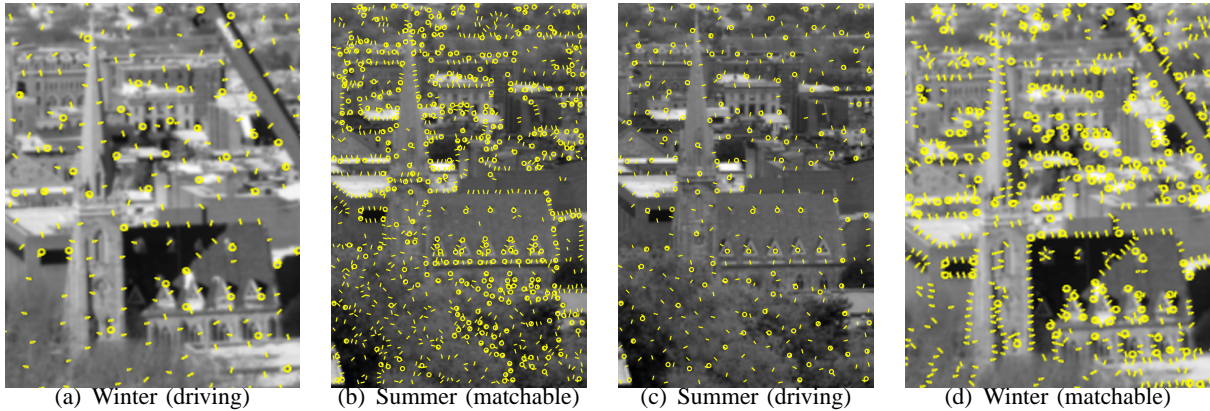
(a) Winter (driving)   (b) Summer (matchable)   (c) Summer (driving)   (d) Winter (matchable)

Fig. 6

EXAMPLE OF "MATCHABLE" AND "DRIVING" FEATURES AND BIDIRECTIONAL MATCHING. IN MATCHING, "DRIVING" FEATURES FROM THE (CROPPED) WINTER IMAGE (A) ARE MAPPED ONTO THE SUMMER IMAGE AND MATCHED AGAINST THE "MATCHABLE" FEATURES (B). SYMMETRICALLY, "DRIVING" FEATURES FROM THE SUMMER IMAGE (C) ARE MAPPED ONTO THE WINTER IMAGE AND MATCHED AGAINST THE "MATCHABLE" FEATURES (D). ALTHOUGH NOT ILLUSTRATED IN THIS FIGURE, MATCHING MAY OCCUR BETWEEN FEATURES AT DIFFERENT SCALE-SPACE SCALES.

A. 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. Driving feature sets are subsets of $\mathcal{P}_c$, $\mathcal{P}_f$, $\mathcal{Q}_c$ and $\mathcal{Q}_f$. Points from all scales are combined to form these sets. Abusing notation, $\mathbf{p}$ and $\mathbf{q}$ represent both the feature and its location. Each feature has associated with it a feature scale, s, at which it was detected, and each face point has a normal direction, $\boldsymbol{\eta}$.

The *forward* transformation of point location $\mathbf{x}$ from I_p onto I_q is $\mathbf{T}(\mathbf{x}; \boldsymbol{\theta}^{pq})$, where $\boldsymbol{\theta}^{pq}$ is the parameter vector to be estimated. An estimate is $\hat{\boldsymbol{\theta}}^{pq}$, and its covariance matrix is $\hat{\Sigma}_{\boldsymbol{\theta}^{pq}}$. The *backward* transformation from I_q onto I_p is $\mathbf{T}(\mathbf{x}; \boldsymbol{\theta}^{qp})$, with an estimate $\hat{\boldsymbol{\theta}}^{qp}$ and covariance estimate $\hat{\Sigma}_{\boldsymbol{\theta}^{qp}}$. Finally, the regions over which the transformation are being estimated are called the "bootstrap" regions, and are denoted by R_p on image I_p and R_q on image I_q. Each is defined as an axis-aligned rectangular box on its own image.

B. Matching Within the Bootstrap Region

The transformation is refined within current bootstrap regions R_p and R_q, ignoring everything else in the two images. Recall that in standard ICP the current transformation is used to

1) Extract keypoints and features from images I_p and I_q (Sec. II and III).

2) Match keypoints from I_p to I_q and rank-order the matches (Sec. II).

3) **Do**

   a) Choose next keypoint match and generate the initial transformations and bootstrap regions (Sec. II).

   b) **Repeat**

      i) Use the current model and parameter estimates $\hat{\boldsymbol{\theta}}^{pq}$ and $\hat{\boldsymbol{\theta}}^{qp}$ to generate the forward and backward match sets $\mathcal{C}_c^{pq}$ or $\mathcal{C}_f^{pq}$, $\mathcal{C}_c^{qp}$ and $\mathcal{C}_f^{qp}$ (Sec. IV-B).

      ii) Use the match sets to re-estimate the forward and backward transformation parameter estimates, $\hat{\boldsymbol{\theta}}^{pq}$ and $\hat{\boldsymbol{\theta}}^{qp}$, and their covariance matrices, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}^{pq}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}^{qp}}$, for the current model and remaining higher order models (Sec. IV-C).

      iii) Use the estimates and covariance matrices to select the model for the next iteration (Sec. IV-D).

      iv) Grow the bootstrap regions in each image separately using the selected model, parameter estimates, and covariance matrices (Sec. IV-E).

      Until region growth, model selection and parameter estimation have converged.

   c) Apply decision criteria using the final model, parameter estimates, covariance matrices, and face correspondence sets. If all pass accept parameter estimates $\hat{\boldsymbol{\theta}}^{pq}$ and $\hat{\boldsymbol{\theta}}^{qp}$ and terminate (Sec. V).

4) **While** fewer than N keypoint matches have been tried.

5) Terminate with failure

Fig. 7

OUTLINE OF THE GENERALIZED DUAL-BOOTSTRAP ALGORITHM

generate a new set of correspondences, in turn these correspondences are used to generate a new transformation, and this process iterates. By contrast GDB-ICP proceeds to model selection and region growing before selecting a new set of matches.

GDB-ICP uses bi-directional matching. This provides more constraints and helps to produce more numerically-stable estimates, especially for small bootstrap regions. A driving feature $\mathbf{p}$ is mapped from I_p to I_q, producing $\mathbf{p}' = \mathbf{T}(\mathbf{p}; \hat{\boldsymbol{\theta}}^{pq})$. The three closest matchable features (of the same type) to $\mathbf{p}'$ are found in I_q, and the best matching feature, $\mathbf{q}$, is chosen from among these three based on a similarity measure described below. The correspondence pair $(\mathbf{p}, \mathbf{q})$ is added to match sets $\mathcal{C}_c^{pq}$ (corners) or $\mathcal{C}_f^{pq}$ (faces). Reversing order, the pair $(\mathbf{q}, \mathbf{p})$ is also added to either $\mathcal{C}_c^{qp}$ or $\mathcal{C}_f^{qp}$. The same procedure is used in the opposite direction to generate correspondences between

driving features from I_q and matchable features from I_p. Once these matches are generated, the forward and backward transformation estimates $\hat{\boldsymbol{\theta}}^{pq}$ and $\hat{\boldsymbol{\theta}}^{qp}$ are calculated. Since these use the same set of constraints, just with the feature roles reversed, the two transformation estimates are close to being inverses of each other (typical root mean square error of 0.1 pixels or less). The remainder of the discussion focuses on the calculation of $\hat{\boldsymbol{\theta}}^{pq}$ using $\mathcal{C}_c^{pq}$ and $\mathcal{C}_f^{pq}$.

A similarity measure is used both in choosing between matches for a feature and in weighting the chosen match during estimation. For corners it depends only on the feature scales, but for face points it depends on orientations as well. For a feature point $\mathbf{p}$, let $s_{\mathbf{p}'}$ be the feature scale after the transformation is applied. For a face point, let $\boldsymbol{\eta}_{\mathbf{p}'}$ be the transformed normal direction. The similarity measures for a prospective match with $\mathbf{q}$, with scale $s_{\mathbf{q}}$ and (for a face point) normal $\boldsymbol{\eta}_{\mathbf{q}}$, are

$$w_c = \min(s_{\mathbf{p}'}/s_{\mathbf{q}}, s_{\mathbf{q}}/s_{\mathbf{p}'}) \qquad \text{and} \qquad w_f = \min(s_{\mathbf{p}'}/s_{\mathbf{q}}, s_{\mathbf{q}}/s_{\mathbf{p}'}) \cdot |\boldsymbol{\eta}_{\mathbf{p}'}^{\top} \boldsymbol{\eta}_{\mathbf{q}}|$$

for corners and face points, respectively. This biases the selection toward features at similar (mapped) scales and orientations, and allows for contrast reversals in face point matches as well.

C. Estimation

Estimation is applied to the current model and, as the basis for model section (Sec. IV-D), to higher-order models under consideration. This section describes estimation for a single model.

Before defining the transformation estimate objective function, we need to define the error distances. For corner points these are Euclidean distances, whereas for face point these are normal distances:

$$d_c(\mathbf{p}, \mathbf{q}; \boldsymbol{\theta}^{pq}) = \|\mathbf{T}(\mathbf{p}; \boldsymbol{\theta}^{pq}) - \mathbf{q}\|/s_{\mathbf{q}} \qquad \text{and} \qquad d_f(\mathbf{p}, \mathbf{q}; \boldsymbol{\theta}^{pq}) = |(\mathbf{T}(\mathbf{p}; \boldsymbol{\theta}^{pq}) - \mathbf{q})^{\top} \boldsymbol{\eta}_{\mathbf{q}}|/s_{\mathbf{q}}.$$

In each case, the distance is normalized by the scale at which the feature is detected, reflecting the fact that feature location uncertainty increases with increasing scale. This makes distances of features at different scales approximately comparable.

Combining the foregoing, the objective function for estimating transformation parameters $\boldsymbol{\theta}^{pq}$ from a fixed set of correspondences is

$$E(\boldsymbol{\theta}^{pq}; \mathcal{C}_c^{pq}, \mathcal{C}_f^{pq}) = \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_c^{pq}} w_{c;i} \rho(d_c(\mathbf{p}_i, \mathbf{q}_i; \boldsymbol{\theta}^{pq})/\sigma_c) + \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_f^{pq}} w_{f;i} \rho(d_f(\mathbf{p}_i, \mathbf{q}_i; \boldsymbol{\theta}^{pq})/\sigma_f) \quad (2)$$

where $\rho(\cdot)$ is the Beaton-Tukey [26], [35], [50] robust loss function:

$$\rho(u) = \begin{cases} \frac{a^2}{6}[1 - (1 - (\frac{u}{a})^2)^3], & |u| \leq a \\ \frac{a^2}{6}, & |u| > a \end{cases} \quad (3)$$

Following standard usage, the constant a is set to 4, which means that normalized alignment error distances beyond 4σ have a fixed cost. The parameters σ_c and σ_f are the robust alignment error scales (standard deviations) for the normalized distances of corners and faces.

Objective function (2) is minimized using the Iteratively-Reweighted Least-Squares (IRLS) technique from the robust statistics literature [35], [50], which alternates computation of a distance-based weight $w_{d,i}$ for each correspondence, i, based on fixed transformation parameters with weighted least-squares re-estimation of the parameters from

$$F(\boldsymbol{\theta}^{pq}; \mathcal{C}_c^{pq}, \mathcal{C}_f^{pq}) = \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_c^{pq}} w_{d,i} w_{c,i} d_c^2(\mathbf{p}_i, \mathbf{q}_i; \boldsymbol{\theta}^{pq}) + \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_f^{pq}} w_{d,i} w_{f,i} d_f^2(\mathbf{p}_i, \mathbf{q}_i; \boldsymbol{\theta}^{pq}). \quad (4)$$

The distance-based robust weight factor for corners is $w_{d,i} = w(d_c(\mathbf{p}_i, \mathbf{q}_i; \boldsymbol{\theta}^{pq})/\sigma_c)/\sigma_c^2$, where $w(\cdot)$, derived from the Beaton-Tukey robust loss function, is

$$w(u) = \begin{cases} 1 - (\frac{u}{a})^2, & |u| \leq a \\ 0, & |u| > a \end{cases}.$$

An analogous computation produces the weights for face points. Normalization factors $1/\sigma_c^2$ (for corners) and $1/\sigma_f^2$ (for face points) make corners and face points comparable.

The robust standard deviation, σ_c, for corner matches is recomputed once (per correspondence set and per Dual-Bootstrap iteration) from the weights and current transformation estimate as

$$(\sigma_c)^2 = \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_f^{pq}} w_{d,i} w_{c,i} d_c^2(\mathbf{p}_i, \mathbf{q}_i; \hat{\boldsymbol{\theta}}^{pq}) \Big/ \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_c^{pq}} w_{d,i} w_{c,i},$$

with a similar computation for face matches. At the start of the Dual-Bootstrap procedure for a given initial transformation, the MUSE algorithm [40] is used to estimate σ_c and σ_f from the first set of matches, since weights are unavailable.

Finally, the computation of the weighted least-squares estimates from (4) and the associated covariance matrix of the parameter estimates, which is needed for the region growth and decision criterion, use standard techniques. In particular, for the transformation models that are unconstrained and linear in their parameters, $\hat{\boldsymbol{\theta}}$ is obtained in closed-form using linear weighted least-squares. The covariance matrix, $\hat{\Sigma}_{\boldsymbol{\theta}^{pq}}$ of the estimate is the inverse of the Hessian of (4). Usually

this must be multiplied by a noise variance term, but this is already built in to the distance-based robust weights $w_{d,i}$. For constrained or non-linear models, such as the planar homography or planar homography plus radial-lens distortion terms, Levenberg-Marquardt is used, with the pseudo-inverse of the Hessian giving the covariance matrix, $\hat{\Sigma}_{\boldsymbol{\theta}^{pq}}$ (see [24, Ch. 4]).

D. Model Hierarchy and Model Selection

The goal of model selection is to select the model from a set (a hierarchy) of transformation models that best describes the current bootstrap region. As the region expands, model selection is applied to choose between the model used for the previous bootstrap region, and the remaining, high-order models. Model selection must be done carefully. Switching to a higher-order model too early, especially when the region is small and there are insufficient constraints, may lead to overfittings and distortions in the estimate. Switching too late causes an increase in mapping errors and a resulting increase in mismatches. In either case, incorrect model selection may drive the estimate into a local minimum representing an incorrect alignment.

Two different model hierarchies are used in GDB-ICP. One, used for retinal images, is a hierarchy moving from similarity to a reduced quadratic to a quadratic model (see [49] for details). The second, used for natural images, is a hierarchy of similarity, affine, homography, and homography plus radial lens distortion (HRD). The HRD model is defined as

$$\mathbf{T}(\mathbf{p}; \boldsymbol{\theta}) = \mathbf{D}(\mathbf{T}_H(\mathbf{D}(\mathbf{p}; k_p); \mathbf{h}); k_q)$$

where $\boldsymbol{\theta}^\top = (\mathbf{h}^\top, k_p, k_q)$, $\mathbf{T}_H(\mathbf{x}; \mathbf{h})$ is the usual planar homography ($\mathbf{h}$ is a 9-component vector formed from the 3×3 homography matrix), and $\mathbf{D}(\mathbf{x}; k) = (1 + k\|\mathbf{x} - \mathbf{x}_0\|^2)\mathbf{x}$ is the radial distortion function, given image center $\mathbf{x}_0$, assumed to be the center of the pixel array. This model is important for accurate alignment of digital photographs taken with off-the-shelf cameras.

Model selection techniques have been studied extensively in the literature[10], [9], [29], [51]. In our earlier work [49] we used a Bayesian technique derived in [9] that depends on computing the determinant of the parameter-estimate covariance matrix, $\hat{\Sigma}_{\boldsymbol{\theta}^{pq}}$. For homographies this is problematic because $\hat{\Sigma}_{\boldsymbol{\theta}^{pq}}$ is not full rank. Rather than developing an appropriate projection onto a full-rank covariance, we have replaced the Bayesian criteria with a modified version of Akaike Information Criteria (AIC), derived from the Kullback-Leibler measure, and found it to be quite effective. Using our robust objective function (2) and taking the advantage of having

transformation estimates and match sets available in both directions, a second-order Akaike Information Criteria may be written as

$$
\begin{aligned}
\mathcal{I} = & - 2 \left[\mid \mathcal{C}_c^{pq} \mid \log(\sigma_c^{pq}) + \mid \mathcal{C}_f^{pq} \mid \log(\sigma_f^{pq}) + E(\hat{\boldsymbol{\theta}}^{pq}; \mathcal{C}_c^{pq}, \mathcal{C}_f^{pq}) \right] \\
& - 2 \left[\mid \mathcal{C}_c^{qp} \mid \log(\sigma_c^{qp}) + \mid \mathcal{C}_f^{qp} \mid \log(\sigma_f^{qp}) + E(\hat{\boldsymbol{\theta}}^{qp}; \mathcal{C}_c^{qp}, \mathcal{C}_f^{qp}) \right] + 2nl/(n-l-1),
\end{aligned}
\tag{5}
$$

where l is the degrees of freedom in current model, $n = 2 \mid \mathcal{C}_c^{pq} \mid + 2 \mid \mathcal{C}_c^{qp} \mid + \mid \mathcal{C}_f^{pq} \mid + \mid \mathcal{C}_f^{qp} \mid$ is the effective number of constraints (each corner match provides with two constraints, while each face point match provides one), and the term $2nl/(n-l-1)$ adjusts for small-sample bias [10, p. 51].

Expression (5) is evaluated for each candidate model using a fixed set of matches found using the transformation estimate of the best model from the previous Dual-Bootstrap iteration. The final objective function value of (2) (after IRLS converges) is used for each model to evaluate expression (5). The model that minimizes (5) is chosen as the current model and its estimated parameters become the current parameters. Model selection is turned off once the selection procedure reaches the highest-order model.

E. Region Growth

Region growth, illustrated in Figure 2, is based on the uncertainty in the transformation estimate, represented by the covariance matrix $\hat{\Sigma}_{\boldsymbol{\theta}^{pq}}$. Expansion of the axis-aligned rectangle representing the current bootstrap region is inversely proportional to the transfer error — the error in applying the estimated transformation to points on the boundary of the bootstrap region. The following is a summary of the details of this procedure taken from [49]. Subsequently, a simple modification is given to make the algorithm more effective in registering image pairs with large scale variations.

Let the center of the bootstrap region be $\mathbf{y}_0$, let a point location centered on one of the four sides of the region be $\mathbf{y}$, and let $\eta_y = (\mathbf{y} - \mathbf{y}_0)/\|\mathbf{y} - \mathbf{y}_0\|$ be the outward-pointing normal to the rectangle. The mapping error covariance at the mapped point $\mathbf{y}' = \mathbf{T}(\mathbf{y}; \hat{\boldsymbol{\theta}}^{pq})$ is computed from the Jacobian of the mapping, $\mathbf{J}_\mathbf{y} = \partial \mathbf{T}(\mathbf{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ evaluated at $\mathbf{x} = \mathbf{y}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{pq}$, together with the covariance of the transformation parameters:

$$
\Sigma_{\mathbf{y}'} = \mathbf{J}_\mathbf{y} \hat{\Sigma}_{\boldsymbol{\theta}^{pq}} \mathbf{J}_\mathbf{y}^\top.
\tag{6}
$$

The outward growth rate is inversely proportional to the error variance in the mapped outward normal direction, η'_y:

$$\delta_y = \beta \frac{(\mathbf{y} - \mathbf{y}_0)^\top \eta_y}{\max(1, \eta'^\top_y \boldsymbol{\Sigma}_{\mathbf{y}'} \eta'_y)},$$

where $\beta = 2.0$ is a fixed constant. The new center of the side is given by $\hat{\mathbf{y}} = \mathbf{y}_0 + (1 + \delta_y)(\mathbf{y} - \mathbf{y}_0)$. The new region is obtained after all the side centers are updated with the above method.

The extension made here is to form and grow bootstrap regions independently in each image, I_p and I_q. The initial regions are determined from the separate keypoint locations and scales, as described in Section II. The above procedure is applied separately for the two regions at each Dual Bootstrap iteration separately. These regions are implicitly kept relatively consistent through the use of bi-directional matching, which keeps the transformations close to inverses of each other and the covariance matrices commensurate with each other. Keeping separate regions in the two images is important for handling large scale differences between images.

V. DECISION CRITERIA

Once the GDB-ICP refinement 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 is tested for correctness. 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 GDB-ICP.

Three tests — accuracy, stability and consistency — form the decision criteria. The tests are applied in each direction using the final match sets. A transformation that passes all three tests in both directions is accepted as correct.

Accuracy is measured as the weighted average error $\zeta_e(\hat{\boldsymbol{\theta}}, \mathcal{C}_f)$, computed on the final face matches, $\mathcal{C}_f$. Face points are used because their positions (along the normal direction) are more accurate than corner points. Using the measures introduced above, accuracy is

$$\zeta_e(\hat{\boldsymbol{\theta}}^{pq}, \mathcal{C}_f) = \left(\sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_f} w_{f,i} w_{d,i} |d_f(\mathbf{p}_i, \mathbf{q}_i; \hat{\boldsymbol{\theta}}^{pq}))| \right) \Big/ \left(\sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}_f} w_{f,i} w_{d,i} \right). \tag{7}$$

Stability is measured by the error covariance — the mapping transfer error introduced in the context of region growth in Section IV-E. To check this, points are uniformly sampled in the overlap area between aligned images. For each sample point $\mathbf{y}_i$, the mapping error covariance $\boldsymbol{\Sigma}_{\mathbf{y}'_i}$
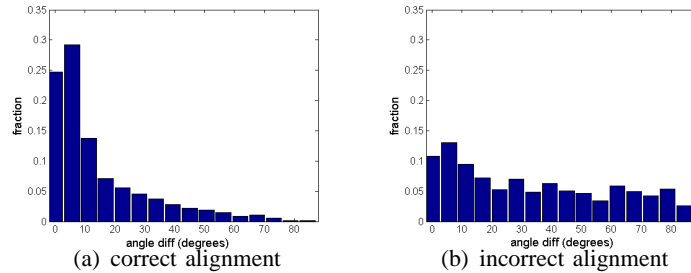
(a) correct alignment (b) incorrect alignment

Fig. 8

EXAMPLE HISTOGRAMS OF THE ORIENTATION DIFFERENCE OF A SET OF FACE POINT MATCHES FROM CORRECT (A) AND

INCORRECT (B) ALIGNMENTS USING THE SUMMER-WINTER PAIR.

is computed from (6). The overall measure is $\zeta_t(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}) = \max_i \operatorname{trace}(\boldsymbol{\Sigma}_{\mathbf{y}'_i})$. This is particular effective at avoiding incorrect low-overlap transformations.

The consistency measure is derived from the orientation differences of the face point match set $\mathcal{C}_f$ after the application of the transformation estimate $\hat{\boldsymbol{\theta}}$. These differences, measured in absolute angle difference, are put into a histogram $h(\hat{\boldsymbol{\theta}}, \mathcal{C}_f)$ of the range $[0, \pi/2]$. (Absolute angle differences greater than $\pi/2$ are subtracted from π, effectively accommodating intensity reversals.) If the transformation is incorrect, this angle difference will tend toward being uniformly distributed, whereas if the images are well-aligned, the histogram will tend to have a strong peak near 0 degrees (Figure 8). The consistency measure is based on the Bhattacharyya measure against an exponential distribution ($\lambda = 4.7$ dictates 70% of the face point matches have orientation differences no greater than 10 degrees). This exponential distribution, denoted as e, is represented as a second histogram. Then the consistency measure $\zeta_c(\hat{\boldsymbol{\theta}}, \mathcal{C}_f)$ is

$$\zeta_c(\hat{\boldsymbol{\theta}}, \mathcal{C}_f) = 1 - \sum_i \sqrt{h_i(\hat{\boldsymbol{\theta}}^{pq}, \mathcal{C}_f) e_i} \tag{8}$$

To make a decision with these measures — ζ_e, ζ_t and ζ_c — lower and upper thresholds are introduced for each: $Z_e^L \leq Z_e^H$ for ζ_e, $Z_t^L \leq Z_t^H$ for ζ_t, and $Z_c^L \leq Z_c^H$ for ζ_c. When $\zeta_e \leq Z_e^L$, $\zeta_t \leq Z_t^L$ **and** $\zeta_c \leq Z_c^L$, the transformation estimate is accepted as correct. When $\zeta_e > Z_e^H$, $\zeta_t > Z_t^H$ **or** $\zeta_c > Z_c^H$, the transformation is rejected. Otherwise, the transformation is saved. If all initial transformations have been tested and none have been accepted, the saved transformation with the minimum value of alignment error ζ_e is accepted. If there are no saved transformations, the algorithm rejects the image pair, indicating that it can not be aligned. Transformations tend

to fall into the "saved" category for image pairs that involve significant changes or that can not be precisely aligned using the final transformation model.

These thresholds are fixed at $Z_e^L = 1$, $Z_e^H = 2$, $Z_t^L = 0.3$, $Z_t^H = 1$, $Z_c^L = 0.09$, and $Z_c^H = 0.2$ for all experiments here. For efficiency, the algorithm also applies a set of higher thresholds, starting after the third Dual Bootstrap iteration when the estimate has begun to stabilize, to identify and terminate estimates that are clearly wrong. We refer to this step as the *early termination criteria*.

VI. EXPERIMENTS

This section presents experiments designed to illustrate the overall performance of the GDB-ICP algorithm (Section VI-B), compare it to minimum-subset random sampling methods (Section VI-C), and analyze in detail the most important aspects of the algorithm. The focus of the latter is on the success of growth and refinement (Section VI-D), the choices of features and matching criteria (Section VI-E), and the effectiveness of the final decision criteria (Section VI-F) — the newest aspects of the algorithm. See [49] for analysis that shows the significance of the Dual-Bootstrap refinement, growth and model selection procedures in the context of retinal image registration.

A. Data Set

All experiments are on the test suite of 22 image pairs discussed in the introduction. This suite was constructed from our own digital photographs, from pairs found on the web, and from challenging pairs suggested by colleagues. Many easier pairs have been left out in order to keep the tests manageable. As an example of this, we included one pair, with 2% overlap (the "Dashpoint" pair here), from the test suite of [8] (GDB-ICP registers all overlapping pairs from this suite). On the other hand, some types of pairs, such as PET-CT images, which have no common geometric structure, have been purposely left out. We discuss this more in Section VII. The results are clearly conditioned on the test suite, but the range of challenging pairs shown should be suggestive of the broad effectiveness of our algorithm. In order to allow the community to test GDB-ICP beyond the experiments presented here, an executable version of the software has been posted on the web.

The images range in size from 676×280 to 2500×2500. Image pairs overlap as little as 2%, differ in scale by a factor as high as 6.4, and differ in orientation by as much as 90 degrees. Five pairs are multi-modal (retina angiogram vs. red-free photograph, two infrared vs. video airport scenes, pan-chromatic and infrared satellite images, and proton density vs. T1 weighted brain MRI slices). Four pairs involve substantial illumination changes and two other pairs are of different seasons. The selection of scenes includes aerial, urban, landscape, indoor, and medical. On the Melanoma and EO-IR 1 pairs, the intensity of one image is negated before keypoint generation since SIFT is not invariant to intensity reversal. Finally, the retinal images involve quadratic transformations, whereas the others involve the use of the homography or the homography plus radial lens distortion (HRD) models. The choice of final model is specified by a command-line argument. All other parameter settings are fixed for these experiments.

B. Overall Results

GDB-ICP successfully aligned 19 of all 22 image pairs in our test suite with error less than a pixel. Success is defined here as no visible misalignments between homologous structures following application of the transformation, as judged independently by a graduate student who is not one of the authors. The successful transformations, one for each pair, are labeled as "verified" transformations to be used in subsequent experiments. Example alignments are shown in Figure 10; complete results are posted at our web site, including animations. Interestingly, for pairs "Brugge", "Brugge Square" and "Brussels", the 10 degree-of-freedom "Homography plus Radial lens Distortion" (HRD) model eliminated small, but visible misalignments produced by using only a homography (see Figure 9 for details).

Table II shows, for each pair, the index number of the first keypoint match in the rank ordering for which the algorithm succeeded, the index of the same successful keypoint match among only those consistent with the verified transformation for which the algorithm succeeded, the final alignment error, and the chosen transformation model. A consistent keypoint match is somewhat arbitrarily defined to have a location error of less than 6 pixels, a scale ratio within the interval 0.8 to 1.25 (one step in scale space), and an orientation difference of 15 degrees, all computed following application of the verified transformation. Intuitively, these are matches that appear to be geometrically close to correct. The remaining keypoint matches are labeled "inconsistent". As can be seen from the table, in most cases a consistent match appears among the first five in

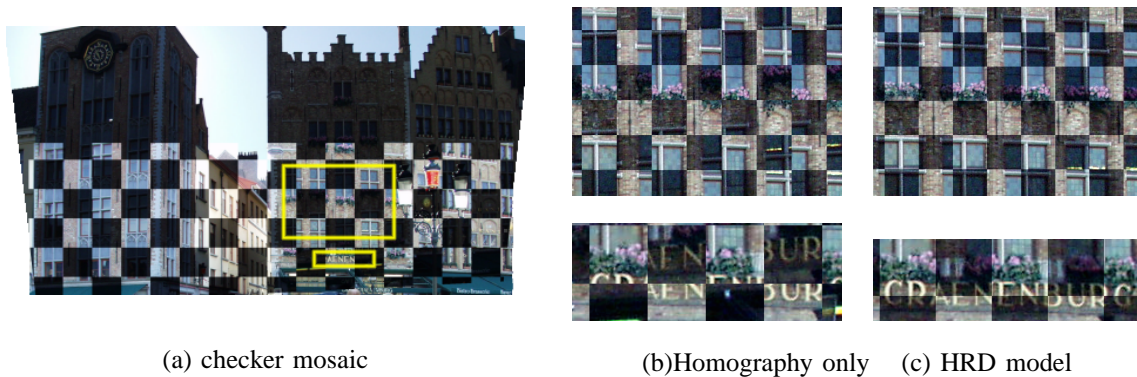(a) checker mosaic   (b)Homography only   (c) HRD model

Fig. 9

(A) SHOWS A CHECKER MOSAIC OF THE BRUGGE SQUARE PAIR USING THE HOMOGRAPHY MODEL, IN WHICH THERE ARE

PLACES, NOTABLY INSIDE THE BOXES, THAT HAVE MISALIGNMENTS. COLUMN (B) SHOWS ZOOM-IN OF THE BOXES.

COLUMN (C) SHOWS THE SAME AREA AFTER USING THE HOMOGRAPHY WITH RADIAL LENS DISTORTION (HRD) MODEL.

THE WINDOW FRAMES AND LETTERS ARE NOW WELL ALIGNED. THE RESULTING ALIGNMENT IS IS SUBPIXEL ACCURATE.

the rank ordering and in 14 cases GDB-ICP successfully refined this initial transformation to a verified final transformation.

GDB-ICP failed for three pairs. In each case, manual specification of three initial correspondences in a small initial region and computation of an initial estimate of an affine transformation, following by application of the Dual-Bootstrap growth and refinement procedure led to a verified transformation. This indicates that the failures are caused by keypoint detection and matching, by keypoint-based initialization, or by the early stages in the Dual-Bootstrap growth and refinement procedures. In one case in particular — Capital Region — the projective distortions are too severe be handled starting from a local similarity transformation.

Image sizes and timing results are summarized in Table I. Clearly algorithm speed is mostly affected by image size and matching difficulty. The failures and the image pairs requiring testing of all 50 keypoint matches (because no match provided results below the lower decision thresholds) are the only ones other than the huge "Satellite" pair requiring more than a minute.

C. Comparison to Keypoint Matching Algorithms

As one indication of the significance of these results, the publicly-available code for the Autostitch keypoint matching algorithm of [7] (with default parameters) produced five alignments

| Image Pair Name | Image Dimension | On the successful keypoint match | | Grand Total | |
		# of Iterations	Time	# of Keypoint matches tried	Time
Boston	1712×1368	14	12	1	12
Boston Library	1504×1000	13	5.7	1	5.7
Brain T1 to PD	512×512	12	2.0	3	2.9
Brugge Square	1712×1368	17	52	2	58
Brugge Tower	1712×1368	20	52	1	52
Brussels	1712×1368	17	37	2	41
Capital Region	1712×1368	NA	NA	50	1225
Day-Night	1000×1504	24	38	2	39
Dashpoint	2048×1536	12	10	1	10
Eiffel	1712×1368	16	15	2	19
EO- IR 1	300×236	15	6.8	50	127
EO- IR 2	676×280	NA	NA	50	51
Extreme Zoom	1504×1000	25	34	50	761
Grand Canyon 1	1184×780	15	9.6	2	12
Grand Canyon 2	900×568	16	26	50	592
Melanoma	1156×880	14	13	50	58
Retina	1600×1200	17	9.9	5	24
Satellite	2878×2878	18	40	4	122
Whiteboard	1504×1000	19	9.5	4	12
White Tower	1504×1000	18	21.5	2	23
Winter-Summer	1504×1000	19	30	1	30
Winter Day- Summer Night	1504×1000	NA	NA	50	300

TABLE I

Timing results in seconds. The first two columns are the image-pair name and the dimension of the larger image. The next two columns are the number of iterations and the time that the Dual Bootstrap growth-and-refinement procedure spent on the keypoint match that led to the successful alignment. The last two columns are the total number of keypoint matches tried and the total time used before GDB-ICP terminates. The performance is measured on a Pentium 4 3.2GHz PC with 2GB memory.

White Tower

Winter-Summer Day (cropped)
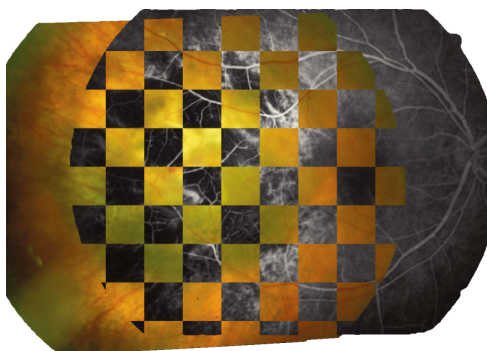

Extreme Zoom (ratio 6.4, cropped)
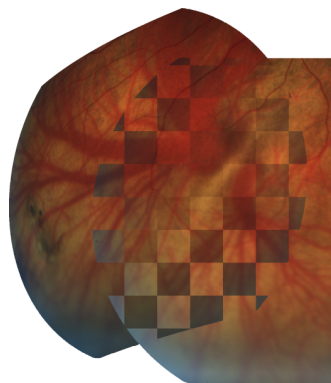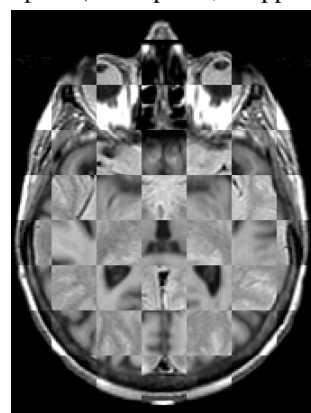

Brugge Square (cropped)

Day-Night Summer

Dashpoint(overlap 2%, cropped)

Melanoma (cropped)

Retina (cropped)

Brain PD-T1

Fig. 10

FINAL ALIGNMENT CHECKERBOARD IMAGES

("Boston", "Boston Library", "Eiffel", "Brugge Square", and "Brussels"). The latter three have visible misalignments, partly due to the fact that the homography is insufficient for these pairs. On the other 17 pairs Autostitch failed altogether. We obtained similar results with our own implementation using RANSAC and other random-sampling-based algorithms ([40], [52]). The failures are due to both the small number and the small fraction of consistent keypoint matches, as shown in the last three columns of Table II.

D. Success of the Growth and Refinement Procedure

The following experiment shows the effectiveness of starting from individual keypoint matches rather than combining them, as in a random-sampling approach. We use the top 50 keypoint matches of the 19 pairs that GDB-ICP aligned. The GDB-ICP estimation process is applied to each of these keypoint matches, without any decision criteria. The resulting "test" transformation estimate is then compared to the verified transformation. Those that agree to within an average distance of less than 2 pixels are considered correct.

Among the 19 pairs of images, there are 950 keypoint matches in total, 363 are "consistent" and 587 "inconsistent". Among the 363 consistent ones, 282 led to correct final transformations, while among the 587 labeled "inconsistent", 28 of these led to correct final transformations, resulting in a total of 310 correct alignments. Examination of the 28 shows that the estimation procedure recovered from initial location errors as high as 12 pixels and orientation differences as much as 18 degrees.

To interpret the significance of these results, based on a probability of $P_a = 282/363 \approx 0.78$ of succeeding from a consistent keypoint match, the overall probability of GDB-ICP producing the correct alignment given n good matches is $1 - (1 - P_a)^n$, which is 0.99 when there are just $n = 3$ consistent matches. By contrast, minimal subset random sampling techniques require 4 matches just to instantiate a transformation. Clearly GDB-ICP can succeed despite an extremely small number of keypoint matches.

E. Choice of Features and Matching

The next set of experiments evaluates several variations on the choice of features, the scale of the features, and the directionality of matching. This is important to show the influence of these

Image pair name	With GDB-ICP				Should RANSAC be used		
	1st success in all keypoint matches	1st success in only the consistent matches	Final alingment error	Final model	Total # keypoint matches	position consistent within 6 pixels	ratio
Boston	1	1	0.36	HRD	1228	506	41%
Boston Library	1	1	0.32	HRD	1030	252	40%
Brain T1 to PD	3	2	0.98	H	40	2	5%
Brugge Square	2	1	0.43	HRD	758	287	37%
Brugge Tower	1	1	0.49	HRD	390	178	45%
Brussels	2	1	0.35	HRD	2728	1213	44%
Capital Region	NA	NA	NA	HRD	176	22	12%
Day-Night (Summer)	2	1	0.62	HRD	124	16	12%
Dashpoint	1	1	0.30	HRD	160	27	16%
Eiffel	2	1	0.30	HRD	442	117	26%
EO - IR 1	2	1	1.59	HRD	62	3	4.8%
EO - IR 2	NA	NA	NA	H	46	0	0%
Extreme zoom	1	1	0.33	HRD	448	89	19%
Grand Canyon 1	2	1	0.88	HRD	90	40	44%
Grand Canyon 2	5	1	1.20	HRD	54	14	26%
Melanoma	27	5	1.88	Q	14	2	14%
Retina	5	0[1]	0.58	Q	102	14	13%
Satellite	4	2	0.26	HRD	2152	989	46%
White board	4	2	0.20	HRD	342	96	28%
White Tower	2	1	1.14	HRD	58	7	12%
Winter-Summer (Day)	1	1	0.60	HRD	204	15	7%
Winter Day-Summer Night	NA	NA	NA	HRD	NA	NA	NA

TABLE II

SUMMARY STATISTICS ON ALL PAIRS FROM OUR DATA SET. THE 2ND TO 5TH COLUMNS SHOW THE PERFORMANCE OF GDB-ICP: THE INDEX (STARTING FROM 1) OF THE FIRST SUCCESSFUL KEYPOINT MATCH IN THE RANK-ORDERED LIST OF ALL KEYPOINT MATCHES, THE INDEX OF THE SAME SUCCESSFUL KEYPOINT MATCH IN ONLY THE **CONSISTENT** KEYPOINT MATCHES , THE FINAL ALIGNMENT ERROR, AND THE CHOICE OF FINAL TRANSFORMATION MODEL — HOMOGRAPHY (H), HOMOGRAPHY WITH RADIAL LENS DISTORTION (HRD), AND QUADRATIC (Q). THE LAST THREE COLUMNS SHOW ABOUT KEYPOINT STATISTICS SHOULD MINIMUM-SUBSET RANDOM SAMPLING TECHNIQUES BE USED: THE TOTAL NUMBER OF KEYPOINT MATCHES (LOWE'S SIMILARITY RATIO < 0.8), THE NUMBER WHOSE POSITIONS ARE WITHIN 6.0 PIXELS FOLLOWING APPLICATION OF THE VERIFIED TRANSFORMATION, AND THE RATIO BETWEEN THE TWO. [1]ON THE "RETINA" PAIR, GDB-ICP SUCCEEDED ON AN "INCONSISTENT" KEYPOINT MATCH — ONE WITH 8.0 PIXELS OF POSITION ERROR.

design decisions on the performance of the overall algorithm. These experiments show that face points and bidirectional matching are important and indispensable.

Just as in the previous test, we evaluate all 50 keypoint matches from the 19 pairs GDB-ICP succeeds upon. We study the change in the aforementioned 310 successes with changes in the feature extraction and matching. We also determine whether one of these changes causes the entire GDB-ICP to fail on a pair on which it originally succeeded.

The results are summarized in Table III, which shows several important results. First, using corner points alone without faces results in a 44% drop in the number of successful initial keypoint matches, and a loss of 9 successful pairs (4 pairs if the "correct" tolerance is raised from 2 to 5 pixels). Apparently, corners are not widely-enough and densely-enough distributed for GDB-ICP to succeed consistently on our challenging test suite. Using faces alone, the drop is only 3%, and no pairs are lost.

Interestingly, using forward matching alone instead of bidirectional matching causes the loss of 28% of the successful keypoint initializations and the three of the most difficult pairs (EO-IR 1, Extreme zoom, and Melanoma). In a related result, not shown in Table III, the percentage of driving features that are mapped to within 2 standard deviations of their corresponding matchable features, thereby creating "inlier" correspondences, ranges from 58% to 83%. This indirectly justifies (a) the ability of GDB-ICP to adapt to substantial differences between images and (b) the decision to push feature extraction toward covering as much of an image as possible, trusting the rest of the algorithm to automatically determine which features are consistent between images.

The final test, shown in the last four rows of Table III, explores multiscale feature extraction. Using scale 1.0 (standard deviation of Gaussian smoothing) results in a loss of nine initializations, but one image-pair — the Melanoma pair. When the single scale at which the features are extracted is increased, the success rate drops quickly. Finally, when using features combined across scales, similar to the scale-space detection technique of many keypoint matching algorithms [37], [30], there is a 13% drop in the number of successful initializations, but no loss of any pairs.

F. Decision Criteria

To analyze the strength of the three-component decision criteria, we compared them with simplified versions. The results show that all three components of the decision criteria are necessary, that bi-directional decision increases robustness, and that the complete decision criteria

variations	succ. init.	succ. pairs	variations	succ. init.	succ. pairs
corners and faces	310	19	forward matching	224	16
corners only	174	10	scale 1.0	291	18
faces only	300	19	scale 1.4	175	8
driving only	314	18	scale 2.0	96	5
matchable only	300	19	combined across scales	272	19

TABLE III

GDB-ICP SUCCESS NUMBERS BASED ON VARYING THE FEATURE SET AND THE MATCHING. THE 2ND COLUMN IS THE NUMBER OF INITIALIZATIONS (KEYPOINT MATCHES) THAT LED TO SUCCESSFUL ALIGNMENTS, WHILE THE 3RD COLUMN IS THE PAIRS FOR WHICH AT LEAST ONE INITIALIZATION SUCCEEDED.

are effective in distinguishing correct alignments from incorrect ones, even in the presence of low overlap, scale differences and physical changes.

One of the simplified versions is the use of alignment error alone, a natural measure for registration based on geometric constraints and the one used in [11], [49] for retinal image registration. We then considered the importance of the three criteria by leaving each out in turn. Finally, we considered the effect of several other aspects of the decision criteria. In these experiments, all $42 \cdot 41 = 1722$ possible ordered pairs of images are used, with both orderings used for each pair because each can produce different initial keypoint matches (see Section II for details), and therefore different initializations. The final model is always the homography (for speed considerations). Alignments passed by the modified decision criteria were examined by a graduate student (not one of the authors) to determine correctness. This turned out to be crucial because this test discovered some small overlaps in our image set that we did not realize existed. Based on this human judgment and based on our verified results, the decisions made by GDB-ICP under various decision criteria could be classified as True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). No changes were made in the parameter values of the decision criteria throughout the experiment.

The results are summarized in Table IV. The first observation concerns the effectiveness of the full decision criteria. The six false negatives come from the three pairs which GDB-ICP known to fail on. The three false positives are image pairs that appeared locally consistent, with

one of the regions having very low contrast. On the other hand, most such pairs are rejected. In fact, 99.8% of the incorrect pairs are rejected. Stated even more strongly, among the 1671 rejected pairs there are $1671 \times 50 = 83,550$ incorrect initializations, including many with low overlap, all of which are rejected.

	TP	TN	FP	FN
Full	42	1671	3	6
Accuracy alone	41	1515	158	8
No Stability	42	1614	60	6
No Orientation	42	1646	28	6
No Accuracy	42	1638	36	6
Only forward	42	1641	33	6

TABLE IV

EFFECTS OF VARYING THE DECISION CRITERIA WHEN APPLYING GDB-ICP TO ALL POSSIBLE PAIRS. TP STANDS FOR TRUE POSITIVES, TN FOR TRUE NEGATIVES, FP FOR FALSE POSITIVES, AND FN FOR FALSE NEGATIVES. SEE TEXT IN SECTION VI-F FOR DETAILS.

As seen in the next four rows of the table, when using only part of the criteria, the number of false positives increases significantly — jumping to 158 for use of the accuracy measure only, but with fewer when two of the three measures are used. These false positives are due to locally-consistent structures, especially near image boundaries (producing apparent low-overlap between images). These experiments show convincingly that all three decision criteria are important. Finally, when the decision criteria are applied only in the forward direction — from the moving image to the fixed image — the false positive rate increases substantially.

VII. DISCUSSION

The experiments on our challenging test suite have demonstrated that GDB-ICP is a powerful registration algorithm, capable of aligning a wide variety of image pairs. Overall, our experience with this data set and with other pairs shows that GDB-ICP succeeds when keypoint matching produces a small number of consistent matches, when there is sufficient common structure between the images to drive the dual-bootstrap process and the decision criteria, and when the similarity transformation is a reasonable initial model. In this section, we examine this success, using the experiments to show how the design of the algorithm allows it to handle the

image registration challenges outlined in the introduction. We also discuss some limitations of the algorithm. We conclude this section by re-examining several individual components of the algorithm.

Our experiments have shown that GDB-ICP can succeed with as few as one consistent keypoint match and with as little as 58% correct matches between the driving features of one image and the matchable features of the other. Remembering that matchable features must pass less-stringent tests than driving features and recalling that contrast reversals are ignored by the matching process, this result explains why the algorithm does so well with substantial changes in image illumination and structure, and even changes in modality. The tolerance for differences in feature extraction and matching is allowed in the algorithm because the decision critera can be trusted to reject nearly all incorrect alignments. The effectiveness of the decision criteria is also crucial to the algorithm's success in handling low-overlap pairs. Using the criteria GDB-ICP accepts a small number of correct, low-overlap alignments while rejecting the extremely large number of low-overlap alignments generated by incorrect keypoint matches. Finally, the ability to generate matches across scales is crucial to handling substantial differences in scale.[2] While another recent algorithm has shown the ability to handle large scale variations [15], it has not be demonstrated in as challenging a context as our test suite.

Despite the demonstrated success, GDB-ICP does have limitations:

- It can not handle extreme appearance differences between image pairs. In our test suite this is mostly due to initial keypoint matching, but we anticipate that the algorithm will fail on other multimodal pairs such as PET-CT pairs, where it is unlikely that the features will capture enough structural similarity. Still, GDB-ICP did succeed on all but one of the multimodal pairs in our test suite, because multimodal images often do have sufficient common structure. Intuitively, the structural and textural differences between the color image and the fluorescein angiogram in the "Melanoma" pair put it near the limit of what GDB-ICP can handle.

- Currently, the decision criteria do not eliminate incorrect alignments of an image-wide repetitive structure such as a checkerboard. On the other hand if only a moderate fraction of a scene involves repetitive structure, the decision criteria make the right decision: although

---

[2]After these tests were completed we tested a pair with a zoom factor of 9.5 and GDB-ICP succeeded.

incorrect alignments in the repetitive region will appear accurate, these produce inconsistent matches image-wide. An example of this occurs in the "Brussels" pair of our test suite.

- As a 2d registration algorithm, GDB-ICP currently only tolerates a small amount of parallax. In a similar vein, it is currently limited to global transformation models.

- There is no convergence proof in the overall algorithm, just as there is no convergence proof in the ICP using anything but Euclidean match distances. In practice, however, on thousands of tests, GDB-ICP has always converged. One reason for this that both region growth and model selection are monotonic.

- Finally, GDB-ICP, while consistently running in less than a minute for two mega-pixel image pairs, is still somewhat slow.

All of these issues are topics of our ongoing work.

Finally, we make a few observations about the individual components of GDB-ICP:

- The experiments show the importance of using multiscale face point (edge-like) features during the alignment process instead of more sparse features such as corners, even when corners are detected at multiple scales.

- While it is straight-forward to replace Lowe's LoG keypoint detector and SIFT descriptor vector with other current keypoint techniques, it would be surprising if this would improve keypoint matching substantially on our data set. Still, a thorough exploration of this question is worthy of future study.

- Our earlier work in the context of retinal image registration showed the advantages of the Dual-Bootstrap ICP growth and refinement procedure over robust ICP when starting from a single initial correspondence [49]. Intuitively, the combination of re-estimation in the bootstrap region and model selection keeps the estimate close to the local minimum while gradually increasing the problem complexity through region growth. In contrast, EM-ICP and RPM [14], [21] attempt to disambiguate matches and estimate the transformation through simultaneous, image-wide consideration of multiple matches per feature. These have not yet been proven to handle low image overlap and a large percentage of features that have no match whatsoever. Our informal experience has been that achieving robustness to unmatchable features pushes the design of the cost function toward heavy reliance on the distance to the nearest match for each feature, significantly reducing the importance of

considering multiple matches per feature. A definitive answer to this question is beyond the scope of this paper.

- The axis-aligned, rectangular shape of the region is a simple, efficient representation. The new use of two independent regions introduced here is important for handling large differences in scale. Region models more sophisticated than our rectangular model could be developed — e.g. ones encompassing only the points where the trace of the transfer error covariance matrix is below a threshold — but the rectangular shape has proven sufficient for all our experiments and has not been the cause of algorithm failure. Note that other region growth techniques have recently been proposed in the literature [17], [48]. Ours differs in that its growth is monotonic and is controlled by the uncertainty in the transformation estimate.

- The three-part decision criteria have proven to be essential for handling the challenging image pairs studied here. Other techniques include Brown and Lowe's combinatorial analysis of keypoint matches [7] and Belongie's use of distance, brightness variation and bending energy for recognition [3]. Clearly, keypoint-based measures alone are insufficient. Measures based on intensity variation or gradient magnitude differences [46] are not appropriate for the range of appearance variation seen here. Finally, although our three-part criteria have proven highly successful, they are not perfect and further improvements are possible.

VIII. Summary and Conclusion

This paper has presented the fully automatic Generalized Dual-Bootstrap ICP (GDB-ICP) image registration algorithm designed to handle a wide variety of image pairs, including those showing scale changes, orientation differences, low overlap, illumination differences, physical changes and different modalities. Building extensively on existing work, the algorithm is in fact a series of algorithms designed to work together to solve the problem. Extensive experiments on a 22 image-pair test suite representative of these challenges have shown the effectiveness of the design and demonstrated that a broadly-applicable, fully-automatic image registration is possible. The experiments have also highlighted areas of potential improvement. The most important of these is initialization, especially when there are large appearance variations between images, caused by physical or illumination changes or differences in image modalities. Despite this, the experiments reported here and experience by both our group and others who have tested the

GDB-ICP executable demonstrate that it is effective enough for widespread use.

REFERENCES

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

[2] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 675–682, 1998.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, June 2002.

[4] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the Second European Conference on Computer Vision*, pages 237–252, 1992.

[5] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[6] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[7] M. Brown and D. Lowe. Recognising panoramas. In *Proceedings of the IEEE International Conference on Computer Vision*, 2003.

[8] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 510–517, 2005.

[9] K. Bubna and C. V. Stewart. Model selection techniques and merging rules for range data segmentation algorithms. *Computer Vision and Image Understanding*, 80:215–245, 2000.

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

[11] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):347–364, 2002.

[12] G. Champleboux, S. Lavallee, R. Szeliski, and L. Brunie. From accurate range imaging sensor calibration to accurate model-based 3-d object localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 83–89, 1992.

[13] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[14] H. Chui, A. Rangarajan, J. Zhang, and C. M. Leonard. Unsupervised learning of an atlas from unlabeled point-sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):160–172, 2004.

[15] Y. Dufournaud, C. Schmid, and R. Horaud. Image matching with scale adjustment. *Computer Vision and Image Understanding*, 93:175–194, 2004.

[16] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[17] V. Ferrari, T. Tuytelaars, and L. V. Gool. Simultaneous object recognition and segmentation by image exploration. In *Proceedings of the Eigth European Conference on Computer Vision*. ECCV, ECCV, 2004.

[18] R. Fransens, C. Strecha, and L. V. Gool. Multimodal and multiband image registration using mutual information. In *ESA-EUSC*, 2004.

[19] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.

[20] L. V. Gool, T. Moons, and D. Ungureanu. Affine/ photometric invariants for planar intensity patterns. In *Proceedings of the Fourth European Conference on Computer Vision*, 1996.

[21] S. Granger and X. Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *Proceedings of the Seventh European Conference on Computer Vision*, pages 418–432, 2002.

[22] W. Grimson, T. Lozano-Perez, W. Wells, G. Ettinger, and S. White. An automatic registration method for frameless stereotaxy, image, guided surgery and enhanced reality visualization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 430–436, 1994.

[23] 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.

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

[25] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3), 2001.

[26] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.

[27] M. Irani and P. Anandan. Robust multisensor image alignment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 959–966, 1998.

[28] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *Proceedings of the Eigth European Conference on Computer Vision*, 2004.

[29] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, 1996.

[30] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[31] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[32] B. Luo and E. R. Hancock. Iterative procrustes alignment with the EM algorithm. *Image and Vision Computing*, 20(5-6):377–396, Apr. 2002.

[33] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[34] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, Sept. 2004.

[35] P. Meer. Robust techniques for computer vision. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*. Prentice Hall, 2004.

[36] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages II:257–263, 2003.

[37] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[38] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[39] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1–2):43–72, 2005.

[40] J. V. Miller. *Regression-Base Surface Reconstruction: Coping with Noise, Outliers, and Discontinuities*. PhD thesis, Rensselaer Polytechnic Institute, Aug. 1997.

[41] J. P. W. Pluim, J. B. A. Maintz, and M. A. Vierveger. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8):809–814, Aug 2000.

[42] J. P. W. Pluim, J. B. A. Maintz, and M. A. Vierveger. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[43] H. Sawhney, S. Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. In *Proceedings of the Fifth European Conference on Computer Vision*, volume II, pages 103–119, 1998.

[44] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or How do i organize my holiday snaps? In *Proceedings of the Seventh European Conference on Computer Vision*, volume 1, pages 414–431, 2002.

[45] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[46] Y. Shan, H. S. Sawhney, and R. Kumar. Unsupervised learning of discriminative edge measures for vehicle matching between non-overlapping cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

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

[48] K. L. Steele and P. K. Egbert. Correspondence expansion for wide baseline stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

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

[50] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Reviews*, 41(3):513–537, 1999.

[51] P. Torr. An assessment of information criteria for motion model selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 47–52, 1997.

[52] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, April 2000.

[53] T. Tuytelaars and L. V. Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 1(59):61–85, 2004.

[54] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[55] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.