Structure-driven SAR image registration

Jeremy S. De Bonet[a] & Alan Chao[b]

[a] Learning & Vision Group
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139

[b] ALPHATECH, Inc.
50 Mall Road
Burlington, MA 01803

ABSTRACT

We present a fully automatic method for the alignment SAR images, which is capable of precise and robust alignment. A multiresolution SAR image matching metric is first used to automatically determine tie-points, which are then used to perform coarse-to-fine resolution image alignment. A formalism is developed for the automatic determination of tie-point regions that contain sufficiently distinctive structure to provide strong constraints on alignment. The coarse-to-fine procedure for the refinement of the alignment estimate both improves computational efficiency and yields robust and consistent image alignment.

Keywords: Registration, Alignment, Coarse-to-fine, Multiresolution, Tie-Point, Texture

1. INTRODUCTION

The ability to bring multiple synthetic aperture radar (SAR) images into alignment is crucial in many applications. For example, a need for accurate alignment commonly occurs when multiple partially overlapping image are "stitched" together to form a single larger image. Even with geocoding information, precise pixel-level registration is need to accurately align the images along seams. Precise alignment is also needed in applications performing change detection; for example, detecting vehicle or missile movement. Here alignment must be accurate enough to guarantee that each stationary target is properly registered, ensuring that correspondence is correctly determined.

Here we present a fully automatic method for alignment SAR images, which is capable of precise and robust alignment.

2. OVERVIEW OF THE ALGORITHM

The automatic registration algorithm presented here can be broken into several stages, as illustrated by Figure 1. In an initial preprocessing stage, the input images are amplitude equalized to enhance the relative importance of scene elements such as roads or trees, which can often be critical in accurately determining alignment. These images are then aligned by a coarse-to-fine registration procedure, in which processing is first done on low resolution versions of the input image, producing an initial alignment transformation. This alignment is then refined by reestimating the transformation with higher resolution images, beginning from the estimate obtained at the lower resolution.

At the coarsest resolution, a set of tie-point regions are automatically determined from the base image. Alignment transformations are evaluated by comparing the tie-point regions in the base image to the corresponding points in the transformation of the second image. Regions are compared using the flexible histogram texture comparison method.[1–3]

A stochastic hill climbing algorithm is used to find successively better alignment transformations. When the hill climbing procedure converges, refinement of the alignment transformation begins at the next higher resolution. For

EMAIL: jsd@ai.mit.edu & alan_chao@alphatech.com
HOMEPAGE: http://www.ai.mit.edu/~jsd & http://www.alphatech.com

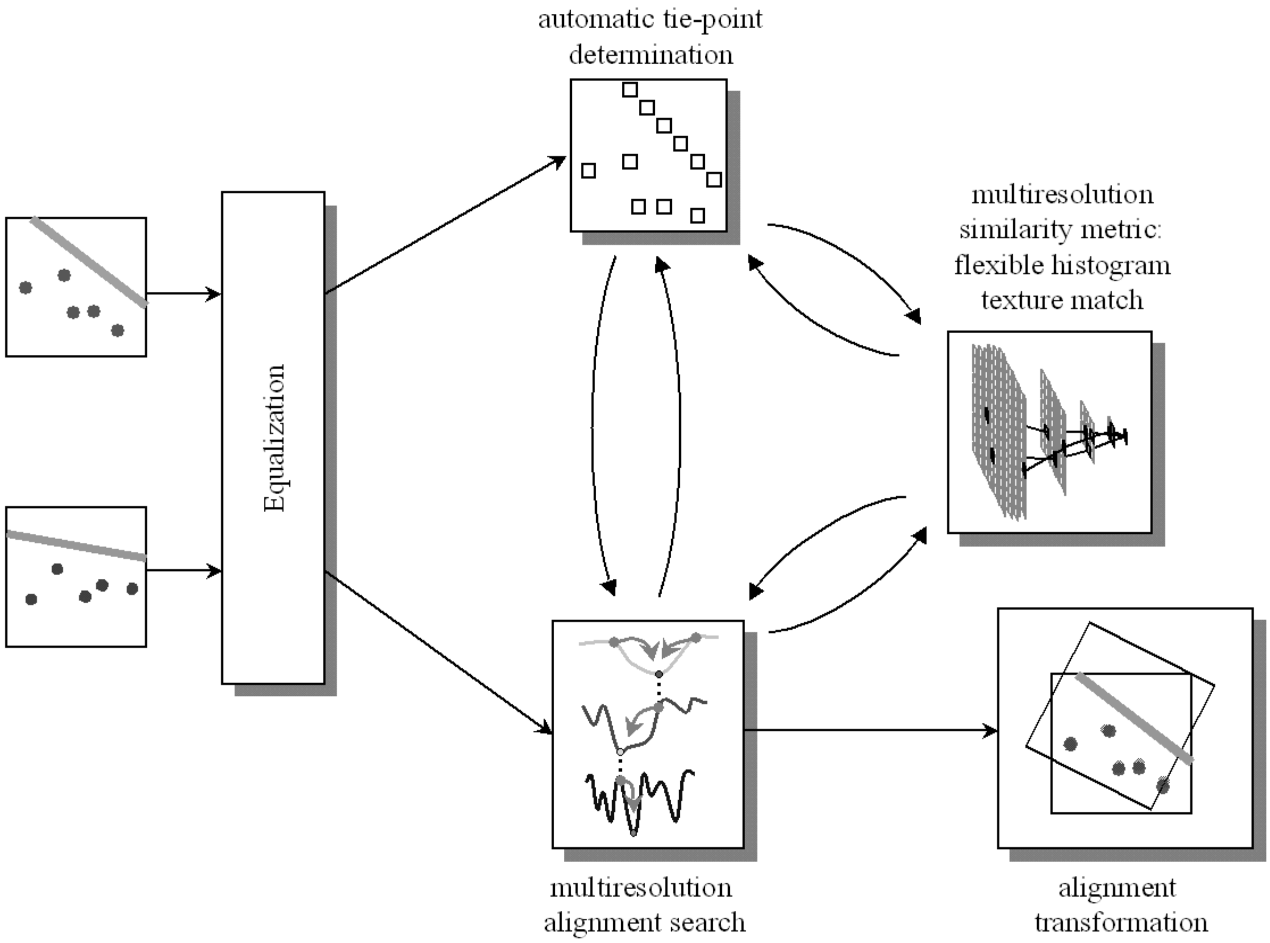

Figure 1. An overview of the image alignment pipeline

each successive resolution, the tie-points used in the previous resolution are mapped into the higher resolution and then refined. The hill climbing procedure is then re-initiated from the convergence point achieved at the previous resolution.

After convergence at the finest resolution, in the experiments described here, an alignment transformation was obtained to within a pixel of perfect alignment as measured by manual registration.

3. HISTOGRAM EQUALIZATION

Unlike the problem of vehicle detection, where natural objects are distractors, when performing SAR image registration low-reflection scene elements – such as roads, water, or forest – contain useful information. To enhance the relative importance of these regions, the input SAR images are preprocessed by dynamically redistributing the pixel brightnesses.

This is done by standard histogram equalization, in which the brightness of each pixel is reassigned by an equalization function $F(\cdot)$ chosen to satisfy the following:

$$\left|\left\{(x,y)\middle|F(I(x,y)) = V_j\right\}\right| = K \qquad\qquad \forall i, j \tag{1}$$

While simultaneously guaranteeing that if $I(x,y) < I(x',y')$ for any two original pixels, then $F(I(x,y)) < F(I(x',y'))$, i.e. that $F(\cdot)$ is stictly monotonic. Thus a constant number of pixels are histogram equalized to each value V_j, and the brightnesses of pixels are not reordered by the histogram equalization. Though it clearly cannot *increase* the

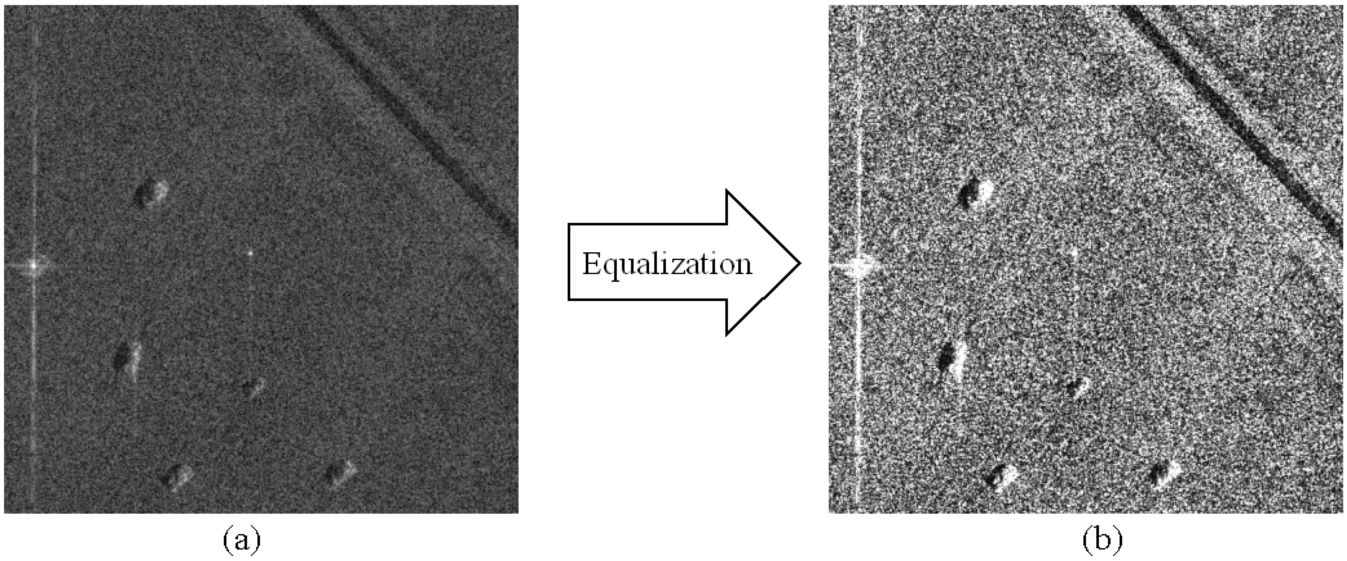

Figure 2. Equalization of the input images enhances the relative importance of low-reflection scene elements

information available in the image, equalization does enhance the effective contrast of elements such as roads, trees, and brick or wood structures, while retaining the highlights and detail on the vehicles.

In Figure 2(a) a SAR scene is shown as a standard log-magnitude image. In this image only the vehicles form high contrast, distinctive regions. After equalization the effective contrast in the road has improved significantly while maintaining high contrast in the vehicle regions. By preprocessing the images in this way the weight of evidence contributed by less reflective objects in the scene can be increased.

4. AUTOMATIC TIE-POINT DETERMINATION

One method for bringing images into alignment is by choosing several distinctive image regions and carefully registering them. These regions are commonly known as “tie-points” or “landmarks”. The quality of a proposed alignment transformation can be determined by evaluating the similarity between the tie-points and the corresponding regions proposed by the transformation. This, counterintuitively, reduces computational cost while simultaneously improving alignment quality.

Using tie-points has several advantages over the alternative technique of feature matching. Because the nature of SAR imagery, finding features is computationally intensive, and often requires an initial denoising stage.[4] Due to the image speckle and geometry of the SAR imaging process, the features found tend to be unstable and highly variable.[5,4]

The notion of using tie-points to improve measurement in this way is not new, and has been used before.[6,7] Each of these, however, requires that the tie-points be selected by human operators. We present a method for automatically selecting good tie-points. This technique is highly generalizable and can be used given any particular matching metric for measuring the similarity between two image regions. In the results presented here, we use one such matching metric which has been shown to work well with SAR imagery, the flexible histogram texture matching metric described in a companion paper in these proceedings,[8] and elsewhere.[1–3]

In Figure 3(a) three distinct types of image elements have been identified. Localized scene elements – such as the vehicle highlighted in the square (1) – provide the strongest constraints on the possible alignment between two images. Because each localized element is only similar to a few regions in the second image, specifically the region around the true corresponding region and the regions around other similar localized elements, the similarity measured between two images at these localized scene elements is a good indication of the quality of an alignment transformation. Even if one were to use a relatively weak matching metric, such as a CFAR statistic or correlation, the distinctness of such localized scene elements can be used to greatly reduce the number of possible alignment

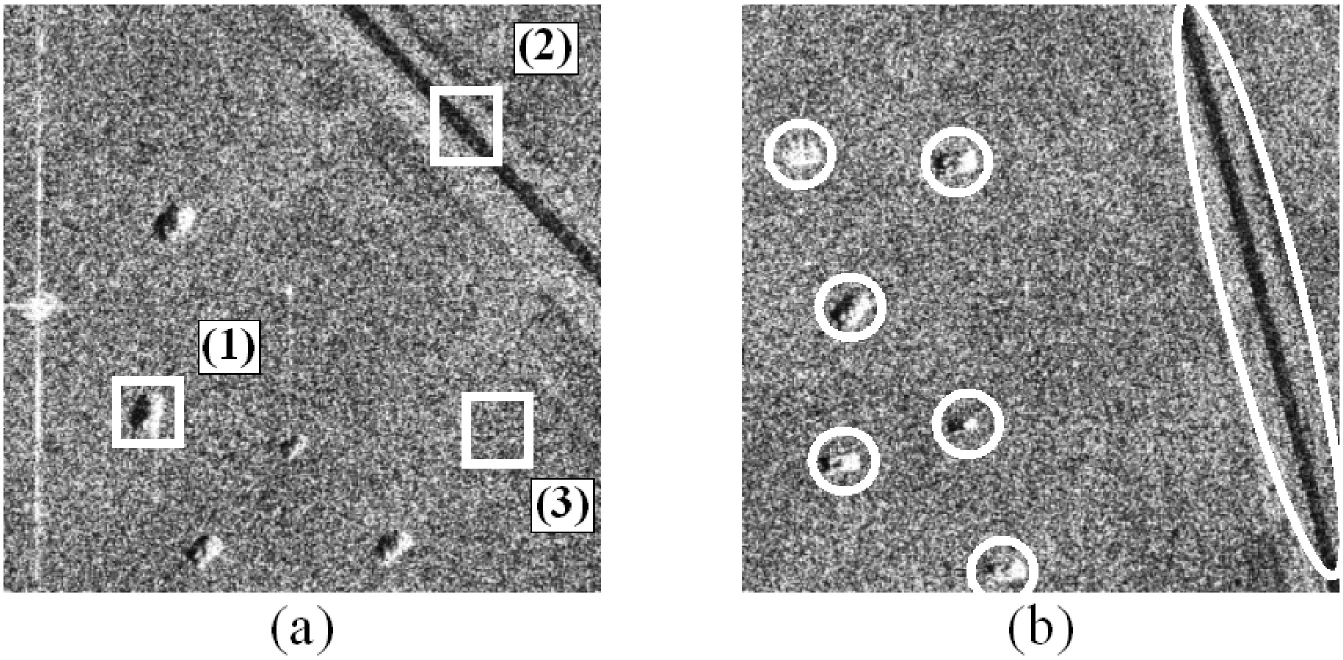

Figure 3. Two SAR images of the same region measured at different aspect angles. If chosen as a tie-point, region (1) constrains alignment to match the circled regions in (b); choosing (2) provides a weaker constraint along region within the oval in (b); while (3) provides no constraint, as it matches with most regions in (b).

transformations under consideration. Each such local element effectively eliminates all but a small set of possible transformations from consideration. For Figure 3 it rules out any transformation which does not map (1) onto one of the circled regions in image (b). In ideal situations enough such localized elements exist, and are shared by the two images, so that using these regions as tie-points will be sufficient constraints to align the two images closely. Other examples of localized scene elements which are commonly seen are: buildings, isolated trees, or bridges. However, such localized elements tend to be rare, and a general and robust SAR registration algorithm cannot reasonably assume that for all input images, sufficient localized scene elements will exist. Additionally, localized scene elements tend to correspond to objects which could potentially move between the times that two SAR images were acquired. This is particularly evident in the case of vehicles, but is also true to a lesser degree to other local scene elements, such as mobile equipment, supplies or even small buildings.

Regions selected from extended scene elements – such as the section of road highlighted in the square marked (2) in Figure 3a – provide weaker, though still valuable, constraints on the possible alignment transformation. Other examples of such extended scene elements include transitions between different types of terrain, such as tree or water lines, fences or power lines, or large building complexes. Because such extended elements are self-similar, a region in one of these elements will tend to be similar to other regions along the same scene element. Thus, a region chosen from along the road in Figure 3 will match well with any of the regions in the oval in (b). This is true even when using a strong matching metric – such as the flexible histogram texture match,[1–3,8] or evaluation by a human observer. Such tie-points selected from within extended elements constrain the final solution to a curve or small local region.

Common scene elements – such as the grass in square (3) in Figure 3a – provide virtually no constraint on the possible alignment transformation. Other examples of scene elements which tend to occur too frequently to be useful include the interior regions of farmland, forest, or water. Because the interior of these types of regions are self-similar, any region within them will be similar to any other such region. Thus, even if a such a region matches well under a proposed alignment transformation, little information is gained – because there are a large number of other transformations under which it will match equally well.

By selecting as many localized scene elements and a sufficient number of extended scene elements to use as tie-points, accurate of alignment can be obtained from comparing just these regions. Clearly there is a significant computational advantage of comparing only those pixels within tie-points regions versus comparing all pixels which overlap under the proposed transformation. Intuitively this corresponds with examining only those regions which are likely to be different if the proposed alignment is inexact.

More importantly, however, limiting the alignment quality estimate to tie-point regions eliminates the additional noise which would otherwise be introduced by the small variations in similarity between the uninformative common scene elements. Even though it is expected that the variation between any two common element regions (i.e. two patches of grass) would be relatively small, the combined effect of all of such regions could introduce sufficient noise to "drown out" the real information from the more localized scene elements.

We present here a formalism for determining tie-point regions in a completely automated fashion. The utility of using a given region as a tie-point is inversely proportional to the frequency with which similar types of regions occur in the input images. Because we can assume that if the images contain views of roughly the same scene it is sufficient to locate distinct points in a single image by comparing them to other regions in the same image.

4.1. Tie-Points are High Entropy Regions

Given a statistical model of a SAR imagery we can in principle measure the entropy of a candidate tie-point, t. The entropy of such a patch, $H_{Model}(t)$, is directly proportional to the frequency at which similar patches are expected to appear. The search for distinctive patches is then a search for the image regions with high entropy:

$$\max_t \left[H_{Model}(t)\right] \tag{2}$$

This notion is quite general; there are potentially many possible estimators of tie-point entropy. For example, edges are useful features in imagery precisely because they have high entropy.[4] We have chosen to use the multi-scale statistical models of SAR imagery defined by De Bonet and Viola.[1,3] These models can be trained directly from SAR imagery. In this case the above entropy can be written as follows:

$$\max_t \left\{E_x\left[H(t|x)\right]\right\} \tag{3}$$

where x is a patch of SAR imagery drawn at random directly from the image to be registered, and $E_x(\cdot)$ is an expectation taken over all possible image patches.

By manipulating equation (3) we can obtain an expression in terms of mutual information:

$$\begin{aligned} \arg\max_t \left\{E\left[H(t|x)\right]\right\} &= \arg\min_t \left\{-E\left[H(t|x)\right]\right\} && (4)\\ &\overset{(a)}{=} \arg\min_t \left\{E\left[H(x)\right]-E\left[H(t|x)\right]\right\} && (5)\\ &\overset{(b)}{=} \arg\min_t \left\{E\left[H(x)-H(t|x)\right]\right\} && (6)\\ &\overset{(c)}{=} \arg\min_t \left\{E\left[I(x;t)\right]\right\} && (7)\end{aligned}$$

where $I(x;t)$ is the mutual information between x and t. Equality (a) holds because $E\left[H(x)\right]$ is independent of t and therefore does not change the result of $argmin_t$; (b) follows from the fact that expectation is linear; and (c) follows from the definition of mutual information. Thus the best tie-point selections are those which have the lowest expected mutual information with other regions in the image.

The output of any image similarity metric, $S(\cdot)$ can be viewed as an approximation of the mutual information between two regions. Thus optimal tie-points can be found from:

$$\arg\min_t \left\{E\left[S(x,t)\right]\right\} \approx \arg\min_t \left\{E\left[I(x;t)\right]\right\} \tag{8}$$

The quality of this approximation is directly related to the quality of the similarity metric. However, if we use the same similarity metric $S(\cdot)$ to determine tie-points as we use to evaluate alignment transformations, then it can be shown equation (8) is optimal with respect to the later measurement.

To reduce computational cost the expectation of $S(\cdot)$ is stochastically approximated. The number of comparisons used in this approximation can be easily tuned to meet specific computational criteria.

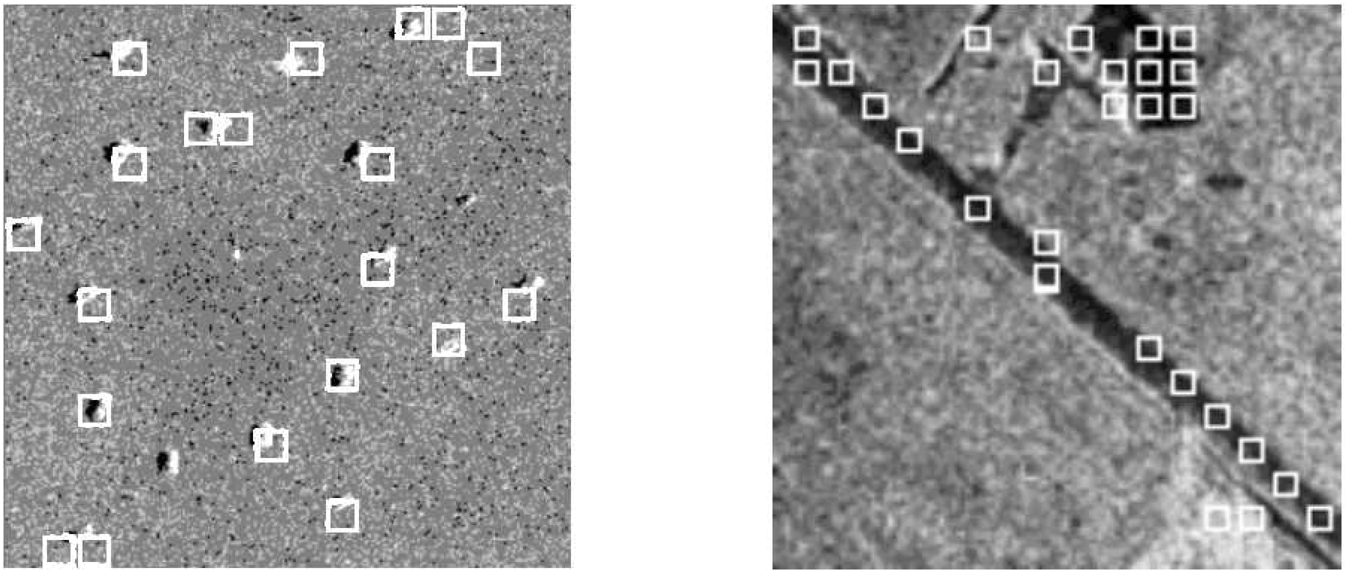

Figure 4. Two typical examples of the tie-points automatically found by this algorithm.

In a similar fashion, a set of good set of tie-points $T = t_1, t_2, \ldots, t_n$ can be found by finding the n most distinct regions in the image as measured by equation (8). In practice, however, it is imperative that the tie-points used not only be distinctive, but that they also cover the image, so that at least some tie-points will lie in the overlapping region under any alignment transformation. To guarantee this, any of a number of schemes can be employed. Here we use a scheme which subdives the image into multiple partitions, considers candidate regions from within each partition, and requires that at least one candidate from each partition be chosen as a tie-point.

In later sections we discuss the coarse-to-fine refinement of the alignment estimate, in which sets of tie-points are needed at each resolution. Each set of tie-points need not be recomputed from scratch if we make the following assumption: distinctive regions at some resolution are likely to be at or around the distinctive regions at lower regions. Clearly this assumption is not true in the general case, as it prohibits small objects, or those which are distinct because of high frequency detail, from ever being chosen as tie-points. Empirically, however, they are valid for mm-wave SAR imagery because high frequency detail tends to be unstable due to speckle. Given this assumption, it is sufficient to locate higher resolution tie-points only in those areas which are within a small region surrounding the tie-points used at the previous resolution.

Two typical examples of the tie-points found by this algorithm are shown in Figure 4. In (a), which contains several vehicles, tie-points are selected in regions around each vehicle; thus, evaluating an alignment transformation, to a first approximation, corresponds to measuring how many vehicles line up with other vehicles. The quality of each match only becomes important when fine tuning the alignment transformation. In (b), fewer localized elements exist and tie-point selections include some of the extended elements in the scene.

5. STOCHASTIC ALIGNMENT OPTIMIZATION

Given a set of tie-points in the base image, we can evaluate the quality of a proposed alignment transformation. One could imagine simply computing the quality of every possible alignment and returning the maximum quality alignment. If we were to consider only the class of translations, such an approach might be feasible though still computationally taxing, requiring $O(X \times Y \times N)$ region similarity comparisons for an X by Y image with N tie-points. However, when we even consider adding another dimension (e.g. rotation) this direct approach clearly becomes infeasible.

Instead we employ a directed search through the space of transformations, in an attempt to determine the best transformation. We do this by using simulated annealing, a standard non-linear optimization search technique, at each resolution. We initiate the search with the convergence point reached at the previous (lower) resolution.

In employing a multiresolution gradient based technique, the following two assumptions are made:

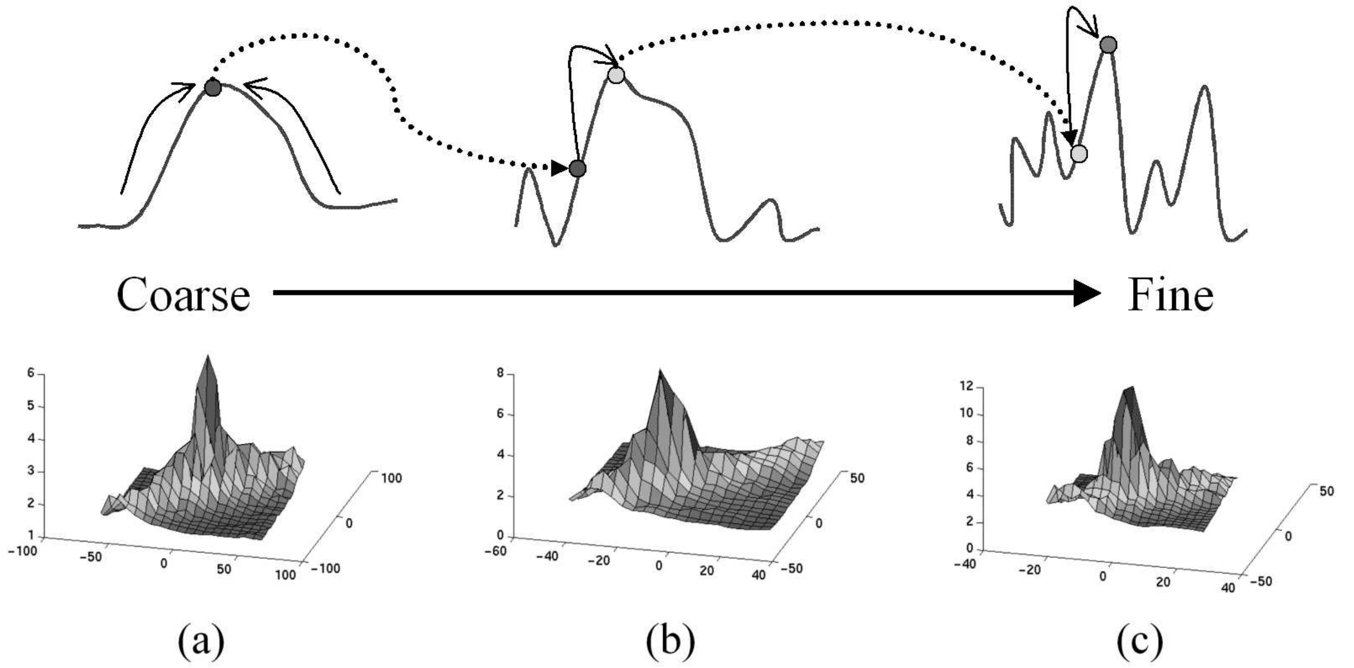

Figure 5. Coarse-to-fine registration is used to improve computational efficiency and to avoid local extrema. A typical example of surface of the alignment quality as a function of translation, at coarse (a), medium (b) and fine (c) scales.

- though there may be noise in the objective function – the alignment transformation quality measurement – its maximum will be surrounded by points whose values are also good.
- the objective function will be smoother at lower resolutions, though the maximum at lower resolutions may deviate from the true fine-resolution maximum.

The rationale behind this methodology is shown schematically in the the curves shown on the top of Figure 5

Using the low resolution estimate to initialize the search at the next higher resolution obtains a significant computational benefit in three ways:

- alignment quality evaluations are less computationally expensive at lower resolutions
- as the low resolution objective function surface is smoother, we can use a lower "temperature" in the search
- since low resolution estimates are close to the higher resolution optimum, fewer local optimal are likely to exist between the global maximum and the low resolution estimate than if we were to begin at some random initialization point; this also allows for the use of a lower search temperature.

For SAR imagery these assumptions are in most cases valid. This can be visualized by examining the objective functions shown in Figure 5. In these graphs, alignment quality is shown as a function of translation. At the lowest resolution (a) the surface is smoothest, however, at this resolution each pixel represents an uncertainty of 4 pixels at the finest resolution, so the estimate of $(0,0)$ is effectively 8 to 12 pixels away from the true alignment. At the next finer resolution (b) the estimate of $(2,0)$, though farther from the true alignment (which is at $(0,0)$), it is nevertheless slightly more accurate due to the increased resolution yielding an offset of 6 pixels. At the finest resolution, the global optimum is exact; however, the surface begins to show local optima which could trap a hill climbing procedure. By initiating the search at the estimate obtained from (b) such local optima can be avoided altogether.

6. OVERVIEW OF SIMULATED ANNEALING

Though we do not have space to discuss it in depth here, we present a brief description of the simulated annealing phase of the alignment procedure. A more complete discussion of simulated annealing can be found in.[9–12]

Simulated annealing is a stochastic hill climbing algorithm in which the objective function is explored around a single estimate of the function's optimum. During this exploration, the estimate of the optimum is updated and search begins from this new estimate. In this work we explore the objective function by choosing potential alignment transformations from a probability distribution centered at the current estimate. In this work, new transformations are chosen from a Gaussian distribution about each parameter in the transformation. I.e.:

$$\begin{aligned} x'_1 &= x_1 + \eta_1 \qquad && \eta_1 \sim \mathcal{N}(0, \sigma_1) \\ x'_2 &= x_2 + \eta_2 && \eta_2 \sim \mathcal{N}(0, \sigma_2) \\ &\vdots \\ x'_n &= x_n + \eta_n && \eta_n \sim \mathcal{N}(0, \sigma_n) \end{aligned} \tag{9}$$

where $X = (x_1, x_2, \ldots, x_n)$ are the parameters for the current estimate of the optimal alignment transformation, and $X' = (x'_1, x'_2, \ldots, x'_n)$ the parameters for the new point under consideration.

If the objective score at X' is better than that at the current estimate X, then estimate of the optimum is updated to X'. Additionally, if the objective score at X' is worse than at X then the estimate is updated with probability:

$$\Pr(X \rightarrow X' | S(X) > S(X')) = e^{-\frac{S(X)-S(X')}{\tau}} \tag{10}$$

where $S(X)$ is the objective score at X, and τ is the temperature of the search.

In the case of image alignment $S(X)$ is the combined similarity measurement between all of the tie-points in the base image and their corresponding regions in the second image under the transformation defined by X.

For SAR image alignment, we consider full six dimensional affine transformations. However, in practice most of this transformation consists of a rigid transformation (translation, and rotation) between the two images. We therefore decompose the affine transformation in the following way:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \cos r & \sin r \\ -\sin r & \cos r \end{pmatrix} \begin{pmatrix} m & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} \cos s & \sin s \\ -\sin s & \cos s \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} dx \\ dy \end{pmatrix} \tag{11}$$

in which there is a translational component (dx,dy) a rotation (r) and stretching along a major and minor axis (m, n, along direction defined by s). In decomposing the affine transformation in this way, the simulated annealing search can be directed to explore each of these dimensions with different size steps, as defined by the corresponding $\{\sigma_i\}$ in equation (10)

The temperature parameter τ in equation (10) is annealed according to a fixed schedule for each resolution. The values for the parameters σ's and τ were chosen manually to optimize search time on several pairs of SAR images.

The analysis up to this point has only used the similarity metric $S(\cdot)$ abstractly. Many algorithms have been shown to be good similarity metrics for SAR imagery[6,7,13–16] and would be reasonable candidates for $S(\cdot)$ In this work we use the flexible histogram texture matching metric, which is described in a companion paper in these proceedings,[8] and elsewhere.[1–3]

7. RESULTS

Several examples of aligned pairs of SAR images are shown in Figures 6 and 7. Automatically selected tie-point regions, highlighted by the white squares, cluster around the distinctive regions in the images.

Though alignment performance is difficult to fully quantify, we have informally been able to make some preliminary measurements of the systems performance. The parameters of the alignment procedure were manually tuned using several image pairs. Alignment was then measured for a collection of 36 different image pairs, with randomly selected initial alignment. Alignment was generally within a few pixels of the optimal affine transformation, as defined by manual alignment. Furthermore, alignment quality was consistent regardless of initial alignment of image pairs.

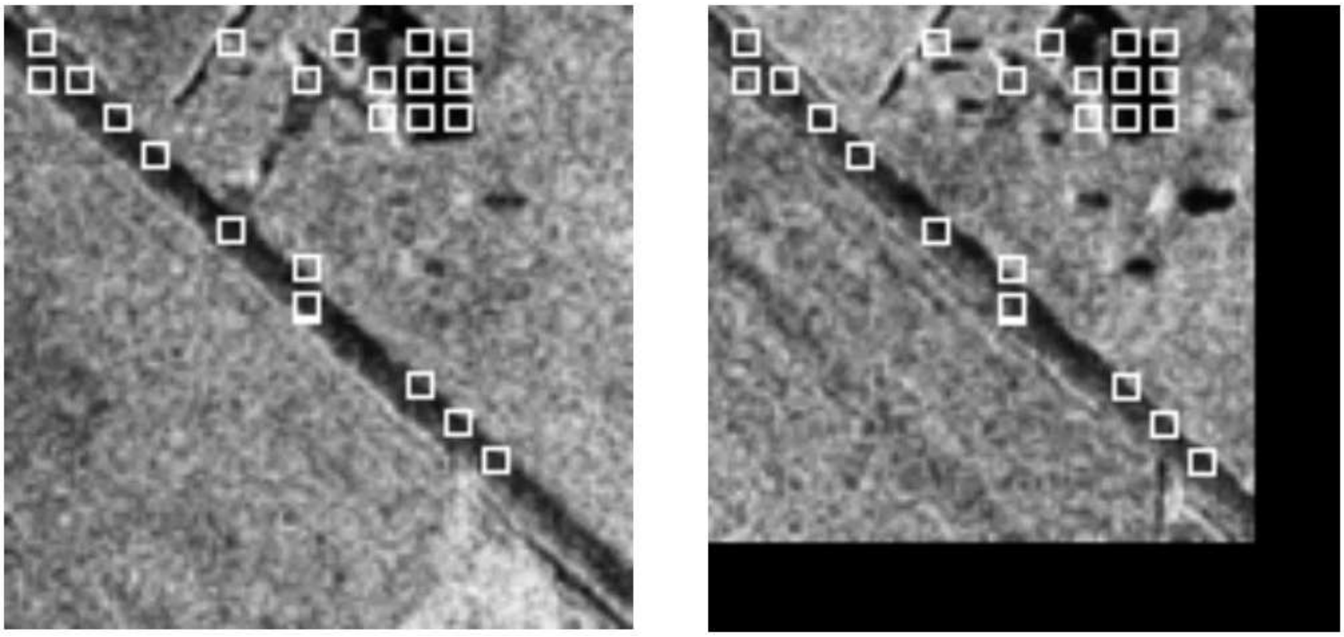

Figure 6. A typical registration

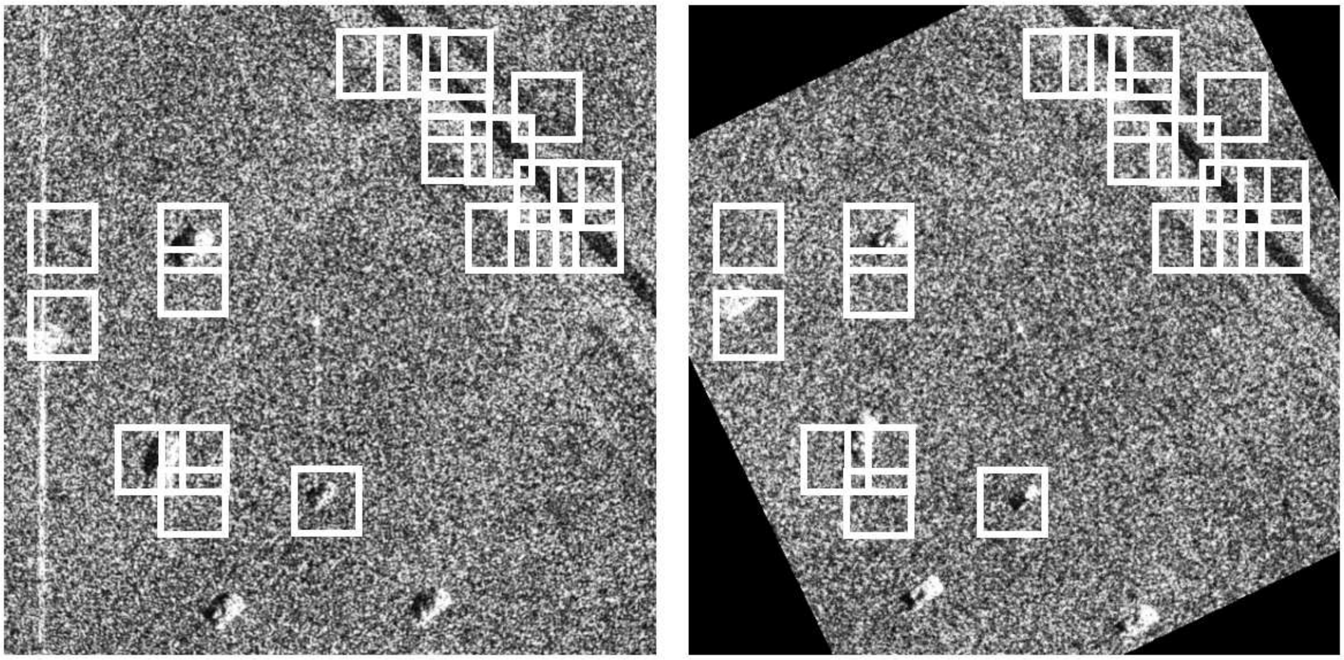

Figure 7. A typical registration

8. DISCUSSION

The two key developments we have presented, which make robust and consistent alignment possible, are the process for the automatic determination of tie-points and the use of coarse-to-fine alignment. Useful tie-points correspond to regions which have a low expected mutual information with other regions in the SAR images. Regions selected in this way correspond to the distinct elements in SAR imagery which are inherently useful in achieving accurate alignment. This formulation for the automatic determination of tie-point regions is extremely general and can can be used by many tie-point based techniques. Using a coarse-to-fine progression to refine the estimate of the image alignment improves computational efficiency and results in more robust alignment. At each resolution, simulated annealing is used to determine the optimal alignment, and is greatly aided by the initialization obtained from the alignment at the previous (coarser) resolution.

The alignment of all of the vehicles in Figure 7 is not perfect even though the optimal *affine* alignment has been found indicating that the class of affine transformations are not sufficient to completely align two images. In future research we intend to examine larger classes of transformations to attempt to achieve even tighter registration between images.

Although the flexible histogram texture matching metric performs well on SAR imagery, as seen by the smoothness of the objective functions in Figure 5, in this direct framework it can only make SAR to SAR comparisons. One of the long term objects of image registration is the alignment of data from several sensors into a single reference frame. The mutual information technique used by Viola and Chao[6] presents a method for measuring the mutual information between pixel brightnesses. We hope to extend this work by examining the mutual information between the flexible histogram texture measures obtained from multiple sensors.

REFERENCES

1. J. S. De Bonet, “Novel statistical multiresolution techniques for image synthesis, discrimination, and recognition,” Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA, May 1997.
2. J. S. De Bonet, “Multiresolution sampling procedure for analysis and synthesis of texture images,” in *Computer Graphics*, ACM SIGGRAPH, 1997.
3. J. S. De Bonet and P. Viola, “Texture recognition using a non-parametric multi-scale statistical model,” in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.
4. P. Dare and I. J. Dowman, “An automated procedure for regisering SAR and optical imagery,” *Proc. SPIE* **2958**, pp. 140–151, 1996.
5. A. Chao and J. S. De Bonet, “SAR image registration using information maximization.” ALPHATECH, Inc., August 1997.
6. P. Viola and A. Chao, “Multiple sensor image alignment by maximization of mutual information.” Massachusetts Institute of Technology and ALPHATECH, Inc., 1995.
7. P. A. Viola, *Alignment by Maximization of Mutual Information.* PhD thesis, Massachusetts Institute of Technology, 1995. MIT AI Laboratory TR 1548.
8. J. S. De Bonet, P. Viola, and J. Fisher, “Flexible histograms: A multiresolution target discrimination model,” *Proc. SPIE* , 1998.
9. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, “Equations of state calculations by fast computing machines,” *Journal of Chemical Physics* **21**, pp. 1087–1091, 1953.
10. McCalla, *Introduction to Numerical Methods and FORTRAN Programming*, John Wiley and Sons, New York, 1967.
11. S. Kirkpatrick, C. G. Jr, and M. Vecchi, “Optimization by simulated annealing,” *Science* **220**(4598), pp. 671–680, 1983.
12. L. Ingber, “Simulated annealing: Practice versus theory,” *Mathl. Comput. Modelling* **18**(11), pp. 29–57, 1993.
13. A. W. W.W. Irving and L. Novak, “A multiresolution approach to discriminating targets from clutter in SAR imagery,” *Proc. SPIE* **2487**, 1995.
14. M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, “Modeling and estimation of multiresolution stochastic processes,” *IEEE Transactions on Information Theory* **38**(2), pp. 766–784, 1992.
15. L. M. Novak, G. J. Owirka, and C. M. Netishen, “Performance of a high-resolution polarimetric SAR automatic target recognition system,” *The Lincoln Laboratory J.* **6**(1), pp. 11–24, 1993.

16. V. Larson, L. M. Novak, and C. Stewart, “Joint spatial-polarimetric whitening filter to improve SAR target detection performance for spatially distributed targets,” *SPIE Conf. on Alg. for SAR Imagery* , Apr. 1994.