The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Multi-Modal Image Registration for Localization in Titan's Atmosphere

Adnan Ansar and Larry Matthies

Abstract— **We study the problem of localizing a balloon in the atmosphere of Saturn's moon Titan by registering onboard imagery with orbital imagery. This is critical for both autonomous navigation purposes and acquisition and sampling of scientifically interesting sites. Because of Titan's atmospheric opacity, we require the ability to match combinations of visible, infrared (IR) and synthetic aperture RADAR (SAR) images. For both localization and direct use as a multi-modal data product for science analysis, match results must be sub-pixel accurate. We demonstrate the feasibility of matching orbital SAR data to visible and IR imagery and outline a framework for using this data as a navigation product. We demonstrate a technique to compensate for local distortions to enable accurate data registration in spite of differences in sensor return and imaging geometry. Finally, we show match results using both terrestrial imagery and the limited amount of available Titan data acquired by the Cassini orbiter and Huygens probe.**

I. INTRODUCTION

While many approaches to image registration exist, the majority assume similarity in the underlying image intensities at corresponding pixels, possibly modulo some easily modeled transformation. See [2], [16] for broad surveys of such techniques. In cases where registration is required between sensors capturing different physical phenomena, the majority of these techniques fail. We adopt a class of algorithms relying on statistical similarity measures such as Mutual Information (MI). These methods have been used successfully by the medical imaging community [9] and, to a lesser extent, by the remote sensing community [6] to match images across different sensor modalities. The requirement is that while the images to be matched may have different appearances, they must have similar information content. MI and similar methods have also been used in computer vision for stereo correlation and image correspondence [7], [16]. We apply them in the context of space exploration.

Our motivating case is a future long-term balloon mission to Saturn's moon Titan, a high priority exploration target for a variety of reasons, including its potential to host prebiotic or protobiotic chemistry that could reveal steps in the origin of life [10]. Such a balloon would have a science payload and would require a significant degree of autonomy due to communications delays arising from the distances involved.

This work was supported by a grant from the NASA Applied Information Systems Research Program (NASA Research Announcement NNH07ZDA001N-AISR)

A. Ansar is a senior member of staff in the Computer Vision Group, NASA Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA. Adnan.Ansar@jpl.nasa.gov

L. Matthies supervises the Computer Vision Group, NASA Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA. Larry.Matthies@jpl.nasa.gov

For both science return and autonomous navigation purposes, it is necessary to precisely localize the balloon relative to Titan's surface. A highly probable scenario, as confirmed through communications with members of the Mission Study Team at the Jet Propulsion Laboratory (JPL) responsible for planning a future Titan mission, is to match imagery acquired from the balloon to imagery acquired by a spacecraft in orbit. Because of Titan's atmospheric opacity, this requires the ability to match combinations of visible, IR and synthetic aperture RADAR (SAR) images. For both localization and direct use as a multi-modal data product for science analysis, match results must be sub-pixel accurate.

In §II we explain how an image matching approach fits into a balloon navigation framework. Details of the proposed match algorithm are then presented in §III. We show match results on both the limited Titan data available (§IV-A) and on a more extensive set of terrestrial data (§IV-B). We present a brief analysis of match sensitivity to errors in scale and orientation in §IV-C. In §IV-D we show an example of accounting for non-linear distortions arising from inherent differences in sensor response and imaging geometry between SAR and Short Wave Infrared (SWIR) instruments.

II. NAVIGATION FRAMEWORK

Our scenario assumes that a survey of Titan, or at least of the areas to be explored by a balloon mission, will have been performed by an orbiter and that the orbital data provides a global context for low altitude navigation. Presently, there is limited hyperspectral and SAR coverage of Titan's surface by instruments aboard the Cassini spacecraft [11], [3]. This coverage is increasing and is expected to be extensively augmented by future orbiter missions.

Because of the large differences in image scale between an orbiter at hundreds of kilometers above the surface and a balloon at a few kilometers altitude, it is necessary to match mosaics from the balloon to single frame orbital images. Observe that in addition to localizing the balloon, this implies the ability to generate high-resolution global maps of the surface that are correctly geo-referenced.

A potential complication is that orbital images are effectively orthographic, while Titan has enough surface relief that image mosaics acquired from a few kilometers must account for perspective effects. Ultimately, this requires that the underlying 3D scene structure be addressed in some way [15] prior to mosaicking. We assume for the purposes of this paper that the problem has been solved or that we are focusing on areas of low surface relief. In either case, composing a sequence of images into a mosaic with coverage adequate for orbital matching is straightforward. Fig. 1 shows

an example using imagery from an experimental JPL airship over a desert terrain.

Relative image scale between orbital and balloon imagery is expected to be known from a combination of accurate navigation solutions for the orbiter and laser or RADAR altimetry for the balloon. A good inertial measurement unit (IMU) also provides highly accurate relative attitude for the balloon, while absolute attitude for the orbiter is known to high precision [8][1]. Once a global attitude is initialized[1], the primary uncertainty in pose knowledge becomes translational. The inherently poor translational solution provided by inertial navigation is further complicated by the presence of wind in Titan's atmosphere. Therefore, our focus is primarily on establishing position from image matching assuming good knowledge of relative scale and orientation. In §IV-C we show that the requirements for this scale and orientation knowledge are mild and easily accommodated.

Depending on mission requirements, image matching may be carried out aboard the balloon. This requires that context imagery (i.e. a geo-referenced map) be available. Because of storage limitations, we envision this map being updated periodically, as the balloon navigates out of the current map region into another, from either a spacecraft in orbit or directly from Earth. The position estimate from the match algorithm will serve as one component of a filtered navigation solution.

III. ALGORITHM OVERVIEW

The image data expected to be available from orbit, primarily SAR and some IR based on the example of the Cassini spacecraft, will differ substantially in appearance from mosaicked imagery from a balloon, for which the nominal imaging payload is a visible camera. The implication for image matching is that traditional intensity-based techniques are unlikely to succeed. As mentioned in §I, we intend to use statistical similarity measures to match orbital imagery to balloon mosaics.

Our scenario differs from typical terrestrial applications of these techniques in that our images are often low contrast and consist of unstructured scenes. We show that the proposed methods remain reliable in these cases and are suitable for the intended application. We begin with a brief overview of the technique and mention a few details specific to our implementation. A thorough summary of the basic approach can be found in [6].

A. Basics

If I_1 and I_2 are a pair of image patches, then a measure of similarity between them is the degree to which their joint probability distribution $P_{1,2}$ can be written as a product of the marginal distributions, $P_1 \otimes P_2$. Here, we approximate the probability distributions by the intensity histograms of the image patches. If the image patches are completely uncorrelated, then we expect

$$P_{1,2} = P_1 \otimes P_2 \tag{1}$$

[1]e.g. by an explicit search over attitude using the methods presented here

It follows that any distance measure f in the space of probability distributions provides a similarity measure S between image patches via

$$S(I_1, I_2) = f(P_{1,2}, P_1 \otimes P_2) \tag{2}$$

where similarity increases with S. The general class of such probability distances, called $f-$divergences, includes several common candidates, some of which we list in Table I. In all cases, M and N are probability distributions. In our application, they represent $P_{1,2}$ and $P_1 \otimes P_2$ as in Eqs. 1 and 2. In the last row of Table I, KL / MI refers to Kullback-Leibler divergence or Mutual Information. We found that in

Name	Expression
Kolmogorov	$f(M,N) = \frac{1}{2}\int \lvert N - M \rvert$
χ^2	$f(M,N) = \frac{1}{2}\int \frac{(N-M)^2}{M}$
Toussaint	$f(M,N) = \int M - \frac{2NM}{N+M}$
KL/MI	$f(M,N) = \int M \log(\frac{M}{N})$

TABLE I
LIST OF CANDIDATES SIMILARITY MEASURES FOR COMPARING PROBABILITY DISTRIBUTIONS M AND N.

practice that there was little to distinguish the four measures listed in Table I in terms of match success. We use MI for the results presented in the following, since the latter has a clear information theoretic interpretation as the difference between the Shannon entropy of M and the cross entropy of M and N via

$$\begin{aligned} \int M \log(\frac{M}{N}) &= -\int M \log(N) + \int M \log(M) \quad (3) \\ &= H(M,N) - H(M) \quad (4) \end{aligned}$$

where H is Shannon entropy[12].

B. Implementation Details

We now list some implementation specific details. The overall match mechanism is similar to stereo correlation. A small template T (representing the balloon mosaic) is localized in a large map M (representing an orbital image) by moving the template over the map and computing a similarity measure as in Eqn. 2. In effect, the location (x, y) of the template in the map is given by

$$(x, y) = \arg\max_{(x,y)} S(T, M_{(x,y)}) \tag{5}$$

where $M_{(x,y)}$ represents a portion of M centered at (x, y) of dimensions equal to T. Unlike stereo correlation, there is limited scope for information re-use along the lines of running sum optimizations. This is due to the need to compute a joint histogram at each trial position, a process for which information cannot be re-used in a simple fashion following a one pixel shift of the template. We compensate in part for the implied computational burden by adopting a coarse to fine strategy. A search over the full map is accomplished only at the coarsest level of refinement. Each subsequent refinement assumes at most a one pixel error

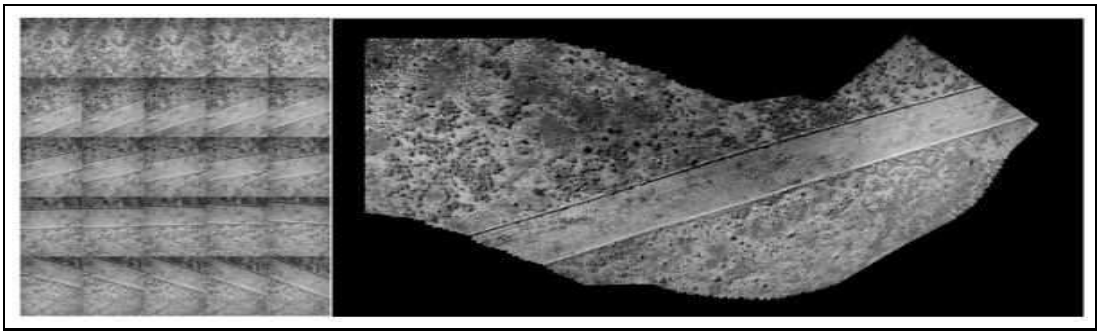

Fig. 1. Several frames of a sequence taken from an experimental airship are shown on the left. A mosaic generated from these images is shown on the right. Observe that the road appears perfectly straight in the mosaic.

in match location. After the integer match is found at full image resolution, a subpixel match is computed following the example of stereo correlation. We use the values of S in Eqn. 2 in the 3×3 neighborhood around the maximum to estimate a quadratic surface. The location of the maximum of that surface represents the subpixel best match of T in M.

Prior to matching, imagery is prefiltered to remove noise. This is critical for SAR imagery in particular. We experimented with Gaussian convolution but found slightly more robust match response by using bilateral filtering[14]. The latter accomplishes a degree of noise suppression without smearing fine structures.

Probability distributions are approximated by intensity histograms smoothed by a Gaussian filter. Results can become erratic without this smoothing as the theory underlying this approach assumes some degree of regularity in the probability distributions. The number of bins in the histogram (i.e. the effective bit depth to which the image is reduced) is based on template size and a global analysis of intensity distributions at full bit depth. In practice, we found that 5 or 6 bits (32 or 64 intensity bins) always sufficed for our examples with fewer bins being required as template sizes were reduced. This becomes especially relevant since we adopt a coarse to fine approach for matching. At the coarsest level, where we assume no prior information on match location, the smallest bit depth feasible is used to reduce computational burden. In flight operations, it may be necessary to learn these minimum bit depths for a particular terrain type by iteratively reducing from a nominal value.

C. Match Refinement

In some cases, the nature of the sensor data returned will vary between instruments, and there may be inherent differences in the image formation process. For example, SAR imagery directly measures topography, while intensity imagery measures albedo and illumination. This leads to local image-relative distortions even when a gross match between map and template is accurate. In an effort to further refine the match result for navigation while simultaneously providing improved data co-registration for better mapping and science analysis, we find it necessary to correct these local distortions. We do not rely on direct reconstruction of underlying scene structure (e.g. via Structure from Motion), since surface relief alone will not account for all cross-modal image variation. Instead, we focus directly on the imagery as follows:

- Compute an initial match as described, using the full match template T.
- Find salient features p_i in either T or matched region of M depending on image type (say T for notational simplicity).
- Using small search regions T_i centered at p_i and the initial match, refine the location of each T_i in M to q_i using the chosen similarity measure (e.g. MI)
- Perform outlier rejection:
  –Prune correspondences with low similarity score
  –Perform robust detection of remaining outliers via thresholding of mapping errors within a RANSAC[4] framework
- Compute Delaunay triangulations based on p_i and q_i.
- Warp convex hull of triangulated region in T to match triangulated region in M using cubic warp in interiors of triangles [5].

Note that the refinement step depends on the initial match derived from the large support region provided by T. Since the process is driven by statistics in the support region rather than pixel to pixel intensity matches, a large degree of local distortion can be accommodated depending on the template size. Salient features can be as simple as Harris corners provided the imagery is suitable, but alternatives include maxima of local KL divergence (more useful for SAR) or even a uniform gridding of the image. The local refinement of each T_i is accomplished in the same manner as the gross match, using Eqn. 5. Note that in addition to filtering correspondences on simple match score, we use RANSAC as a robust statistical outlier rejection technique [4].

IV. RESULTS

Our results focus primarily on terrestrial examples since those are the most readily available. However, we have acquired some of the limited relevant data from Titan.

Note that the lack of ground truth for these datasets makes validation somewhat subjective and that our primary

evaluation tool consists of comparison to hand-registered data. Similar techniques used in terrestrial remote sensing [6] have relied on man-made structures and high contrast regions with distinctive features for match validation. Our problem requires more challenging test cases with the attendant difficulty in validating match accuracy. Nevertheless, for a proof-of-concept application of information theoretic matching techniques to Titan and Titan-like scenarios, we believe that our results are convincing.

A. Titan

The sole existing example that matches our application scenario perfectly is a comparison between a SAR image (2.8 cm Ku-band) taken from Cassini and a mosaic created from imagery (660-1000 nm) acquired by the Descent Imager-Spectral Radiometer (DISR) system aboard the Huygens probe as it landed on Titan. In Fig. 2, we show a hand match done by the U.S. Geological survey [13] and our result. While no ground truth exists, the automatic match is identical to the hand registration in localizing the mosaic in the SAR image. Since this example matches mosaicked imagery in

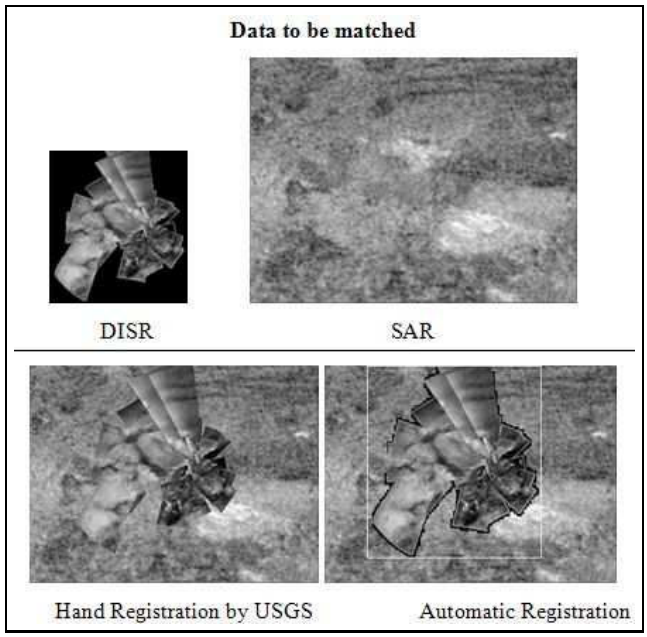

Fig. 2. DISR mosaic matched to and overlayed on Cassini SAR image. Automatic registration produces results identical (to the pixel level) to hand expert hand registration by USGS.

atmosphere to orbital data, the result is compelling for our intended application. Furthermore, it establishes that imagery acquired through Titan's narrow spectral visibility window can be matched to orbital imagery, a fact not obvious without this example.

We have also successfully matched near IR imagery from the Cassini Imaging Science Subsystem (ISS) and IR imagery from the Visible and Infrared Mapping Spectrometer (VIMS) to SAR. The latter is a hyperspectral instrument from which an image consisting of 3 IR bands was matched to SAR. These results also match registrations done by hand

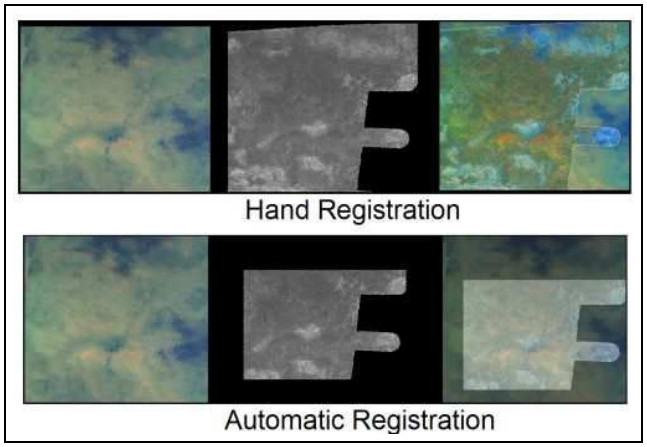

Fig. 3. VIMS IR image (left), Cassini SAR (center), match overlay results (right). Automatic registration produces results identical to expert hand registration.

[13]. The VIMS example is shown in Fig. 3. Note that a smaller template is used for the automatic registration method and that while the automatic match did not color balance the overlay, the result is identical to the pixel level.

B. Terrestrial

Terrestrial multi-modal data is available in much greater abundance. We show in Fig. 4 an example from Death Valley, CA of 4 sensors imaging the same region. The appearance change between instruments is substantial and cannot be compensated for using direct manipulation of the intensity images. Thus, this type of terrestrial data is a good candidate to test image matching in the absence of appropriate Titan data. Our experiments used SPOT (500-730 nm) imagery of Death Valley, CA and ASTER (SWIR, 1600-1700 nm) imagery of Rancho Cima, TX at a pixel resolution of 1500 x 1500 as map images. The coarse to fine match started with up to an 8X reduction in image dimensions with a 50 x 50 minimum size restriction on the template. Various instruments were used to derive templates. A few match results are shown in Fig. 5. Ground truth was not available, so the estimates of match accuracy are subjective. A full list of successful trials, using multiple locations in each map image is listed in Table II.

Map	Template
SPOT	Landsat TM, band 3 (630-690 nm)*
SPOT	SEASAT, L-band SAR (23.5 cm)
SPOT	SIR-C, C-band SAR (5.8 cm)
ASTER	Landsat TM, band 5 (1550-1750 nm)*
ASTER	TIMS, band 3 (9-9.4μm)
ASTER	SIR-C, C-band SAR (5.8 cm)
ASTER	AirSAR, C-band SAR (5.8 cm)

TABLE II
TEST CASES USED FOR AUTOMATIC MATCHING. TEN DIFFERENT TEMPLATES WERE USED FOR EACH CASE LISTED.

There were no failure cases, provided a template size of 500 x 500 (or 11% of the map size by area) was used.

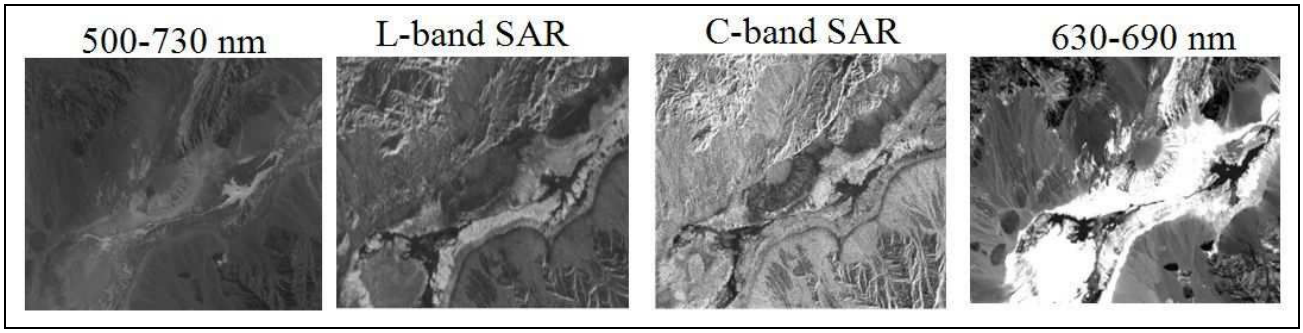

Fig. 4. Identical location in Death Valley acquired by four different instruments with wide variation in appearance between images.

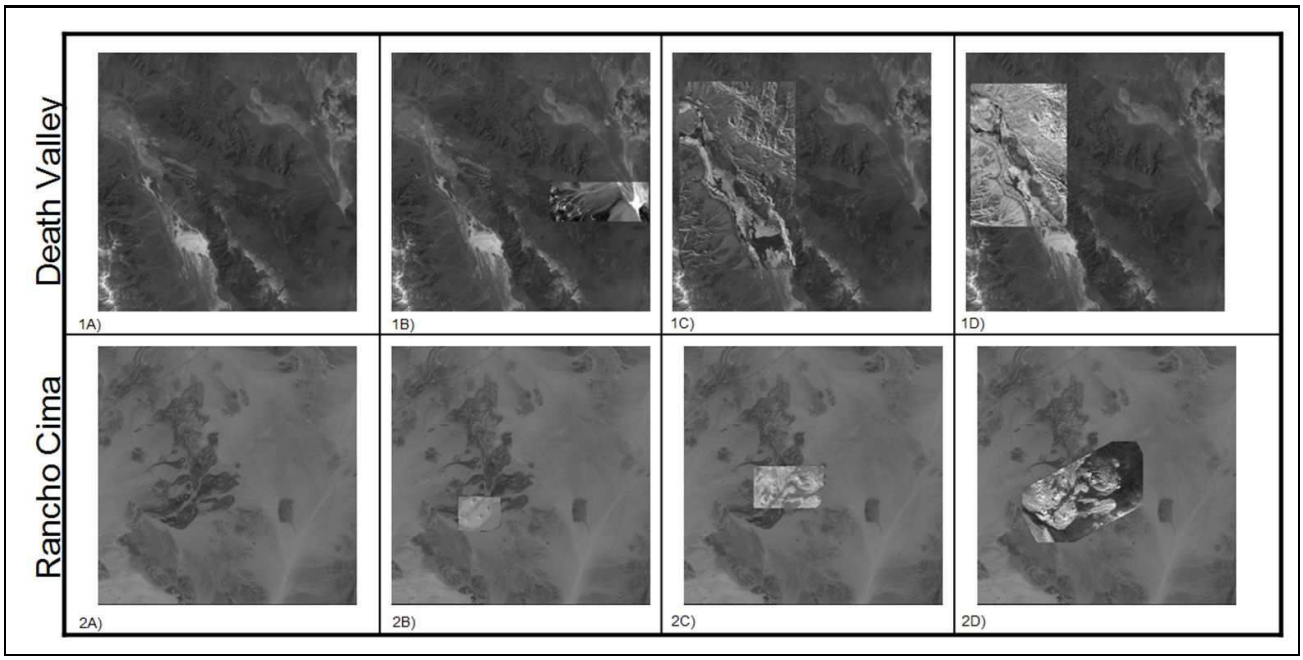

Fig. 5. Automatic registration of (1A) SPOT image to (1B) LANDSAT, (1C) SEASAT and (1D) SIR-C and of (2A) ASTER to (2B) TIMS, (2C) SIR-C and (2D) AirSAR. Matching succeeds in spite of wide appearance change.

In most cases, provided some image texture existed in the template region, the minimal template size could reliably be reduced to 200 x 200 (or 1.8% of the map size). For navigation purposes, we assume a good prior on position from the previous localized frame and the onboard state estimator. As a result, we anticipate that the effective map size that must be searched will be greatly reduced and that 11% is already pessimistic for a template to map ratio. This is significant in terms of computational complexity. Even with an unoptimized implementation of the match algorithm, we find that for the ASTER vs. SIR-C case and a template size of 400 x 240 that reducing the native map size to span 3 times the template size (Template was 35% of reduced map in area) reduces runtime by a factor of 18 from 9 sec. to 0.5 sec on a standard 2 GHz laptop. Note that in the two starred cases in Table II there is spectral overlap. However, in the SPOT vs. Landsat case, there was still significant appearance change, while in the ASTER vs. Landsat case, there was significant difference in contrast. Both cases failed with SAD correlation in regions where Mutual Information succeeded. An example of the former can be seen in (1B) of Fig. 5

C. Orientation and Scale

We tested the sensitivity of our match algorithm to errors in orientation and scale. All tests were conducted using the terrestrial data shown in Table II with 500 x 500 templates in regions of relatively high texture. Ground truth was taken as the match result with no input error. In each case, the full 1500 x 1500 ASTER or SPOT map was searched for a match, and success was defined as accuracy to within one pixel of ground truth. We tested orientation sensitivity by rotating the template in-plane in steps of 1 degree. We found that in all cases, matches were successful with rotations of less than 5 degrees. We tested scale sensitivity by scaling the template in steps of 1% of image dimensions. Again we found that within bounds of 5% image scale error, matches were successful. This is well within expected scale and orientation uncertainty [8],[1].

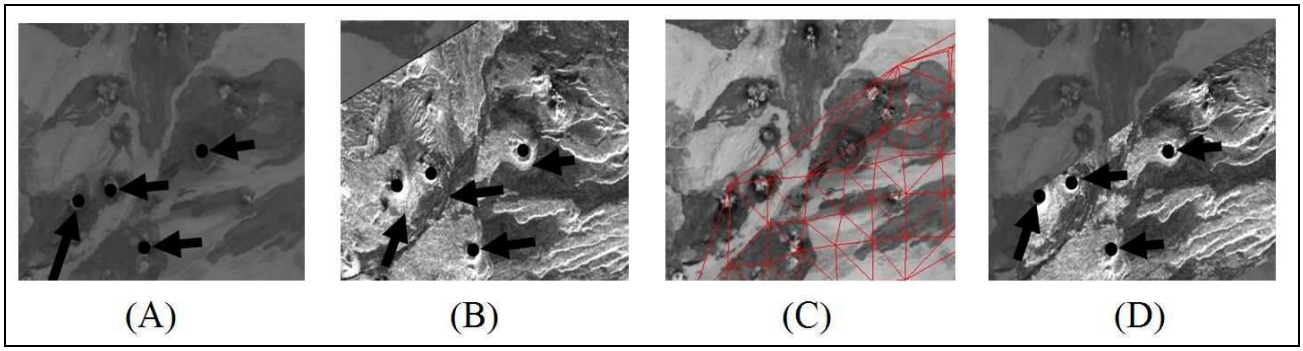

Fig. 6. Closeup of ASTER vs. AirSAR result. Landmarks are indicated by dots, and image correspondences by arrows. The initial match is correct on global scale but shows significant local mismatch. Match results improve dramatically after non-linear warping.

D. Local Distortion

Consider (2D) in Fig. 5. This is an example of two different sensing modalities for which the resulting imagery cannot be directly related by a linear transformation. Furthermore, the ASTER image is from a nadir viewpoint, while the AirSAR image is a sidelooking airborne RADAR, further complicating the local match. In (A) of Fig. 6, we see a closeup of the ASTER image with several obvious landmarks highlighted with black dots. In (B), the same landmarks are indicated with dots, while the correspondence from the match result is indicated with arrows. The two agree in the lower portion of the image, but there is significant error visible in the upper portion. This illustrates that while a large template produces an overall match driven by statistical similarity, local errors persist. Provided we can compensate at the local level, this is clearly a strength rather than a weakness. We apply the method described in §III-C to adjust the image. The Delaunay triangulation of the ASTER image is shown in (C) with image contrast artificially enhanced to make the triangulation more visible. Finally, (D) shows the warped result in the convex hull of the triangulated network. Observe that visible landmarks (dots) now match the image correspondence (arrows) throughout.

V. CONCLUSIONS

We propose a solution for localization of a balloon in Titan's atmosphere using multi-modal image registration between mosaicked balloon imagery and orbital imagery. We have shown that the basic approach works using both terrestrial data and the only directly relevant data available from Titan. Our analysis shows that an assumption of translation uncertainty alone suffices, given the expected fidelity of scale and orientation information. Finally, we have shown that in spite of differences in data return and imaging characteristics between widely disparate sensors, we can perform a local refinement. This serves as both an aid to localization and as a product in itself with applications to high-resolution mapping and scientific analysis.

VI. ACKNOWLEDGMENTS

We are indebted to various colleagues at the Jet Propulsion Laboratory for their assistance with portions of this research. Thanks to Bonnie Buratti and Steve Wall for valuable discussions regarding instrumentation and science requirements for Cassini and potential future Titan missions, to Tom Farr for terrestrial SAR/visible/IR data and to the Titan Mission Study Team for positive feedback on our balloon navigation concept. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] P. Antreasian et al. Cassini orbit determination performance during the first eight orbits of the saturn satellite tour. *AAS/AIAA Astrodynamics Conference*, 2005.
[2] L. Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, 1992.
[3] C. Elachi et al. Cassini radar views of the surface of titan. *Science*, 208(5724):970–974, 2005.
[4] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.
[5] A. Goshtasby. Piecewise cubic mapping functions for image registration. *Pattern Recognition*, 20:525–533, 1987.
[6] J. Inglada and A. Giros. On the possibility of automatic multisensor image registration. *IEEE Trans. on Geoscience and Remote Sensing*, 42(10), 2004.
[7] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. *IEEE International Conf. on Computer Vision*, 2, 2003.
[8] A. Lee and G. Hanover. Cassini spacecraft attitude control system flight performance. *AIAA Guidance, Navigation and Control Conference*, 2005.
[9] F. Maes. Multimodality image registration by maximization of mutual informaion. *IEEE Trans. on Medical Imaging*, 16(2), 1997.
[10] Solar system exploration roadmap for nasa's science mission directorate, 2006. NASA Document CL#06-1867-A.
[11] C. Porco et al. Imaging of titan from the cassini spacecraft. *Nature*, 434:159–168, 2005.
[12] C. Shannon. *The Mathematical Theory of Communication*. Univ. of Illinois Press, 1949.
[13] L. Soderblom. Exploring saturn's moon titan, an earth-like alien, 2006. Presentation: Cassini-Huygens Analysis and Results of the Mission.
[14] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. *IEEE Intl. Conf. on Computer Vision*, pages 836–846, 1998.
[15] Z. Zhu, A. Hanson, and E. Riseman. Generalized parallel-perspective stereo mosaics from airborne video. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26:226–237, 2004.
[16] B. Zitova and J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000, 2003.