MISR Stereoscopic Image Matchers: Techniques and Results

Jan-Peter Muller, Athula Mandanayake, Catherine Moroney, Roger Davies, David J. Diner, *Associate Member, IEEE*, and Susan Paradise

Abstract—The Multi-angle Imaging SpectroRadiometer (MISR) instrument, launched in December 1999 on the NASA EOS Terra Satellite, produces images in the red band at 275-m resolution, over a swath width of 360 km, for the nine camera angles 70.5°, 60°, 45.6°, and 26.1° forward, nadir and 26.1°, 45.6°, 60°, and 70.5° aft (hereafter, referred to as Df, Cf, Bf, Af, An, Aa, Ba, Ca, and Da, respectively). A set of accurate and fast algorithms was developed for automated stereo matching of cloud features to obtain cloud-top height and motion over the nominal six-year lifetime of the mission. Accuracy and speed requirements necessitated the use of a combination of area-based and feature-based stereo-matchers with only pixel-level acuity. Feature-based techniques are used for cloud motion retrieval with the off-nadir MISR camera views, and the motion is then used to provide a correction to the disparities used to measure cloud-top heights which are derived from the innermost three cameras. Intercomparison with a previously developed "superstereo" matcher shows that the results are very comparable in accuracy with much greater coverage and at ten times the speed. Intercomparison of feature-based and area-based techniques shows that the feature-based techniques are comparable in accuracy at a factor of eight times the speed. An assessment of the accuracy of the area-based matcher for cloud-free scenes demonstrates the accuracy and completeness of the stereo-matcher. This trade-off has resulted in the loss of a reliable quality metric to predict accuracy and a slightly high blunder rate. Examples are shown of the application of the MISR stereo-matchers on several difficult scenes which demonstrate the efficacy of the matching approach.

Index Terms—Cloud-top height, Multi-angle Imaging Spectro-Radiometer (MISR), operational applications of photogrammetry, stereo matching.

I. INTRODUCTION

THE operational production of reliable global cloud-top heights and motion has long been a goal for the application of satellite imagery to numerical weather forecasting and climate studies. Although operational techniques have been developed for cloud-top pressure assignment of cloud motion

J.-P. Muller is with the Department of Geomatic Engineering, University College London, London WC1E 6BT, U.K. (e-mail: jpmuller@ge.ucl.ac.uk).

A. Mandanayake is with 3Q Technologies, Unit 6, Harefield, UB9 6JA, U.K.

C. Moroney, R. Davies, D. J. Diner, and S. Paradise are with the Jet Propulsion Laboratory, Pasadena, CA 91109-8099 USA.

vectors using radiative transfer methods such as brightness temperature, CO_2 slicing, and Oxygen A-band, all of these current techniques suffer from the need to have accurate radiometric calibration and external temperature-pressure profiles derived either from radiosonde measurements or objective analyses.

Stereo photogrammetry provides a purely geometric technique which can be used from both geostationary satellites and polar orbiters. Hasler [1] first demonstrated the successful application of stereo from geostationary satellites using simultaneous Geosynchronous Operational Environmental Satellites (GOES) imagery to produce stereo red/green anaglyphs to show qualitatively the changing heights of clouds, in response to and the cause of different atmospheric dynamical regimes and synoptic weather patterns. This work culminated in the successful demonstration of the variable height and motion field around a hurricane [2] from stereo images taken at 3-min intervals. The first demonstration of automated stereo matching to retrieve quantitative cloud-top heights and motion was presented in [3], and the application to understanding the dynamic evolution of a hurricane eye was described in [4]. However, the lack of any suitable long-term datasets and the difficulty in having to use 6-bit gray-level 1-km visible imagery has resulted in very limited use by the meteorological community. Wylie *et al.* [5] recently showed much improved results using GOES-S and GOES-9, whereas [6] demonstrated that one can combine thermal IR-based brightness temperatures to reduce the computational time necessary for stereo matching. A constellation of MISR-like instruments on polar-orbiting platforms could provide extremely useful wind and cloud-top height data for the operational weather forecasting community. If a fleet of overlapping geostationary satellites could, in future, include instruments with 1-km, or preferably, higher resolution thermal imagery, with higher bit quantization and associated dynamic range, operational applications of stereo photogrammetric determinations of cloud-top heights and motion could supersede existing radiance-based sensors, providing more reliable and independent measurements to the operational weather forecasting community.

MISR is the first spaceborne instrument on a polar platform to provide operational and simultaneous retrieval of cloud-top heights and motion [7], [8]. This goal has been very challenging given that there are some six million measurements per orbit with some 16 orbits per day resulting in a requirement for real-time analysis to match some 65 000 measurements per minute. Crucial to this goal of providing an operational technique able to meet this challenge has been the development of a suitable suite of fast, robust, accurate, and reliable

automated image matchers. The different stereo-matchers are first described together with a "superstereo" technique used as a benchmark for assessing different stereo matching techniques. The stereo-matchers were developed using simulated cloud images and tested on a variety of cloudy scenes from a lower resolution conical scanning instrument the Along-Track Scanning Radiometer (ATSR2) (on board the European ERS-2 satellite). So far, as the authors are aware, this is the first time that digital stereo photogrammetric techniques have ever been used to produce continuous measurements from a spaceborne platform. An intercomparison of the "superstereo" technique with the area-based technique developed for stereo matching clouds is then shown and discussed in the context of developing a robust technique capable of being used operationally. An assessment is also shown of the accuracy of the area-based matcher by comparing the height field derived from MISR with an independently derived set of ground digital terrain elevation data (DTED) over two mountainous regions. How these stereo-matchers are employed within the MISR cloud-top height and motion retrieval system is outlined in Section II and is discussed in more detail in [9], [10] and overall in [7]. Stereo matching is performed as part of the MISR level 2 top-of-the-atmosphere cloud algorithm (known as "L2TC") with the Df-Bf-An-Ba-Da cameras for wind (i.e., cloud motion vectors) retrieval and the Af-An-Aa cameras for height retrieval. See [7] for definitions of the cameras and more details about the camera choice. Although the base-to-height ratio is not optimal (thus, resulting in poorer height resolution) for the choice of the Af-An-Aa near-nadir cameras, they are useful to minimize distortions due to foreshortening, which is increasingly present with the off-nadir cameras. Section II describes the overall retrieval system and Section III, the stereo-matchers themselves. Section IV describes the results of an intercomparison of the operational area-based stereo-matcher against the "superstereo" matcher, and Section V describes a qualitative and quantitative assessment of the stereo-matcher with ground DTED heights for cloud-free MISR scenes of two mountainous regions. Section VI presents discussion and conclusions.

II. MISR Cloud-Top Wind and Height Retrieval Scheme

The MISR L2TC retrieves cloud-top motion on a 70.4-km grid and uses these motion vectors (also referred to as the winds in the following discussion) to provide a correction to the apparent disparities in cloud feature locations which are then used to retrieve cloud-top heights. The mathematical basis of the photogrammetric techniques which allows the unique viewing geometry of the nine MISR cameras to be used to retrieve cloud-top motion and heights simultaneously is given in [9]. A short summary is included here to aid the reader.

Some 7 min elapses between the forward-most (Df) camera and the aftmost (Da) camera, during which time cloud advection is likely to have taken place due to wind. This motion appears in the stereo-matched disparities of the chosen cameras as a height error if not properly accounted for. For areas close to jet streams, fronts, or severe storms, this error can be very large.

The L2TC system uses a combination of the extreme off-nadir cameras that are most susceptible to wind effects with less

oblique cameras in a pairwise combination of camera triplets (Df-Bf-An) and (An-Ba-Da) to determine motion, assuming that the atmosphere contains just two principal cloud layers ("high" and "low"). These retrievals are performed by matching the 275-m red channel radiance values from the aforementioned camera pairs using radiance data (which have already been projected onto a surface ellipsoid and resampled onto an equal area space oblique mercator (SOM) grid). This step uses the two-dimensional (2-D) nested maximum feature-based matcher (described in Section III). This feature-based matcher was employed to drastically reduce the computational time needed to calculate the cloud-tracked winds compared with an area-based matcher. Each set of disparities is used to calculate a motion vector, and these individual vectors are binned into a 2-D histogram. The two most populated bins in this histogram are said to be the retrieved motion vectors for this 70.4-km domain. Two search windows in the comparison image are created based on retrieved winds. The search window that contains the final match determines the appropriate cloud motion value to use for calculating the cloud-top height. If the best match occurs in the overlap region of the two search windows, the average of the two motion vectors is used for the correction.

Image pixels on the 275-m SOM grid from the innermost three cameras are then matched in a pairwise fashion using the An-Af and An-Aa camera pairs with the M2/M3 stereo-matchers described in Section III. The matching is performed on every fourth pixel (equivalent to every 1.1 km), and the cloud-top heights are compared from the An-Af pair with those matched from the An-Aa pair. The higher of these two values is chosen within a prespecified threshold. The larger height value is chosen because this height field is later used to provide a common reference altitude to reproject the radiances to, in preparation for calculating the albedo. Using the maximum height helps to minimize the obscuration encountered during this reprojection process [10]. Normalization of the radiance values is used within the matcher to try to minimize effects due to the bidirectional reflectance distribution function (BRDF) of the clouds. These matched pixel disparities are then used together with the camera geometry [9] and the wind error correction to calculate cloud-top heights, and these are reported in the MISR L2TC product.

III. MISR Stereo Matching Techniques

Over the last 40 years, literally thousands of stereo matching algorithms have been developed. In the world of machine vision and image understanding, these techniques have tended to emphasize matching features extracted from images, such as edges or corners, in order to maximize speed and image throughput, whereas in the field of photogrammetry the emphasis has been on trying to maximize completeness, accuracy, and reliability within the framework of an estimate of the quality based on least squares adjustment.

Several papers, including [11] and [12], grouped stereo matching algorithms into two main categories:

1) area (patch)-based techniques that rely upon the concept of smooth surfaces so that adjacent pixels will generally represent contiguous points in space;

2) feature-based techniques which can be further subdivided into the following:

    a) semantic features that have known physical properties and/or spatial geometry, which can be subdivided further into subclasses:

- generic (e.g., vertices of linear structures, occlusion edges, and prominent surface markings);
- domain-specific types (e.g., corner of building, road surface marking)

    b) intensity anomaly features which include zero crossings of the difference-of-Gaussian type, as well as anomalous image patches (e.g., bright corner of field).

Although cloud features appear to have distinct edges, these edges are often not well defined and appear different when viewed from different angles, particularly with optically thin clouds. Prelaunch experimentation showed that cloud edges would not be reliable features for matching. Similarly, corners are unreliable matching features for analogous reasons. Therefore, a feature-based matcher that uses maximum brightness values which usually appear within a cloud were employed here in order to provide feature-based image disparities. Area-based matching techniques using cross correlation on image patches are also inappropriate, due to the highly variable size of cloud features and the requirement with cross correlation to work with sufficiently large enough windows for fast Fourier transforms to pick up a large enough signal. Cross correlation has been used for matching geostationary satellite images for the retrieval of cloud motion vectors, but here the emphasis is on matching cloud patterns or fields rather than individual clouds (see recent reviews in [13] and [14]).

A particular challenge with cloud stereo matching is the presence of height discontinuities, occlusions, transparent layers, and shadows, especially at the 275-m resolution of the MISR pushbroom images. Fortunately, the 14-bit radiance images from MISR usually have sufficient contrast for patterns to be found that can be matched. However, this high contrast and dynamic range also mean that the aforementioned challenges are more noticeable than in lower bit quantized images typical of current meteorological satellites.

The challenge of providing a global set of stereo-matched points necessitated the development of a simple scheme, which, at the same time would be sufficiently fast to permit global processing, accurate enough to meet the scientific requirements, and robust enough to be able to cope with the difficulties posed by the nature of cloud images. After extensive experimentation with both simulated and coarser scale (ATSR2) images [15] and the testing of a large variety of different algorithms including cross correlation and adaptive least squares correlation, a two-pronged approach was developed and implemented.

In the first approach, a feature-based matcher, named nested maxima (NM), was developed to find bright maxima within clouds. This was used to find cloud targets that could be employed for the retrieval of cloud motion vectors within a mesoscale domain (70.4 km). In the second approach, an area-based matcher [called multipoint matcher using means

(M2)] and a variant [called multipoint matcher using medians (M3)] were developed. Both M2 and M3 are used for the higher resolution height retrieval. In both cases, only pixel-level acuity was chosen, with no patch resampling, due to speed and efficiency considerations.

MISR data are first resampled from their original pushbroom pole-to-pole swath into a SOM projected onto the WGS84 ellipsoid. This swath is further subdivided into a set of equally sized image blocks (140.8 km along-track by 563.2 km cross-track), each one of which is processed separately in along-track order. In this way, near epipolar imagery is formed, which considerably simplifies and consequently speeds up stereo matching.

A. Nested Maxima Stereo-Matcher

NM is a feature-based matcher designed to sort through large amounts of data at a very high speed. Its coverage is sparse with typically only 1%–2% of the data being matched. Most of its matches are accurate to within 1–2 pixels (see Section IV for more detail on this), but the distribution of returned disparities does exhibit a long tail. Feature-matching techniques are used to quickly select a small set of candidate matches, and then an area-matcher (M2 in the current implementation) is used to select the best match from among a short list of possibilities. This is much quicker than applying the computationally expensive area-matcher to each pixel in the scene.

First, all the local maxima are identified in each along-track string of data. Only the neighboring points in the along-track (vertical) direction are considered—it is immaterial whether a point is a local maxima in the across-track direction or not. The local maxima of the local maxima are then found, and this process is repeated up to five times, with each new set of local maxima forming a different level. When looking at the original radiance data, there has to be a monotonic decrease in value for two pixels on both sides of the maxima: at the higher levels, this search radius is reduced to one.

Once all the local maxima on all the levels have been identified, matching begins on a top-down basis, processing the "top" layer (or the last level of local maxima) first and then proceeding downward until one is working on the local maxima gathered from the original radiance data. Each local maximum in the target image is considered in turn, and the corresponding search window is drawn around the comparison image. Each along-track string within the comparison search window is examined for local maxima. If the string contains only one local maximum on this level within the search window bounds, then it is considered a preliminary candidate match. The "backward" (comparison to target) search window is then superimposed on the target string, and a tally is made of how many target local maxima exist on this level within the search window bounds. If only the original target point is found, the comparison point is promoted to an "official" candidate match. The above process is repeated independently for each comparison along-track string within the search window. In this manner, all comparison pixels that could be a match for more than one target point are rejected, and only the unique matches are retained.

Once the list of candidate matches is complete, the M2 matcher is run on each target-comparison pair, and then the pair

with the lowest metric value (see Section IV) is chosen. The same ambiguity test that is used with the M2 and M3 matchers (see below and [10] for details) is also applied here.

This matching process is applied to all unused target local maxima on this level, starting at the top level and ending at the second lowest. After each level has been processed, all the matches that have been found thus far are marked as "used" and are removed from further consideration.

The search windows mentioned above are calculated by considering the maximum possible wind and height ranges (0–100 m/s and 0–20 km, respectively) and by taking into account the view angles and time lags between the two cameras. These search windows are tilted by the same angle that the spacecraft ground-track makes with respect to the SOM grid. The angle of the ground-track is determined by reprojecting a selected point to the surface ellipsoid from a set target height (repeating this process for both the target and comparison cameras) and setting the tilt of the window to the ratio of the across-track offset (differenced between cameras) and the along-track offsets.

B. Multipoint Matcher Using Means and Multipoint Matcher Using Medians

M2 and M3 are area-matching techniques that aim for almost complete coverage (70%–90% is typical for MISR). They are slower than NM, because they try to match every pixel and because they also check a greater proportion of the search window. A small pixel patch (10×6 is used by MISR) is extracted around each target pixel chosen for matching (every fourth pixel on a regular grid in the An camera) and around each high-resolution 275-m comparison pixel in the search window. The M2/M3 metric is computed for each pair, and the one with the lowest metric value is chosen as the best match (providing the metric is less than a preset threshold value). This is then subjected to an ambiguity test to weed out those matches with very similar metric values but widely different disparities. The thresholds used to accept or reject an M2 or M3 match are fixed numbers that have been derived empirically from studying match quality versus metric value for a number of different scenes. Metric values of 0.75 and 1.0 are used for M2 and M3, respectively. These may be subject to change in the future as we analyze more data. They do not vary according to scene type but are specific to all MISR scenes, irrespective of the types of clouds present in the scene. Choosing the proper patch size to be used involves a trade-off between matcher accuracy and execution speed. A small patch is faster, but a larger patch decreases the blunder rate and improves the accuracy of the matcher. The wind retrieval search windows are determined by calculating the maximum distance a given comparison pixel could be displaced from the target pixel, based on the range of wind and height values being searched for and taking into account the difference between view angles and imaging time of the pair of cameras. The height retrieval search window is calculated by displacing the comparison camera image the proper distance, as indicated by the retrieved cloud motion values and then increasing the bounds of the search window to accommodate the proper height range (0–20 km). It is tilted to align with the spacecraft ground-track (see [10] for a full mathematical description). The disparity due

to height is assumed to be a factor in the along-track direction only.

In the interests of minimizing the computational load, two techniques have been developed to decrease the size of the search window that is scanned each time. The first one (termed "previous match") looks for successful adjacent matches in both the horizontal and vertical directions. For a match to be considered as a good starting or seed point, it must pass a tighter-than-usual threshold test of its own, where these new thresholds are set to one half of their preset values of 0.75 for M2 and 1.0 for M3. A small search window is drawn around the union of any previous matches. If a successful match is found (one that passes both the threshold and ambiguity tests), it is accepted without further ado.

The second method ("image pyramid") works by reducing the resolution of the image by a specified factor and then matching these averaged images using the full-size search window. A smaller search window is then drawn around the results of this procedure, and the full-resolution images are matched. In the operational software, there are only two levels to the pyramid, and the image is reduced by a factor of two in size.

The previous match method is used first, followed by the image pyramid if necessary. In both cases, M2 is attempted first with a fall-back to M3. This is done in the interests of time, since M2 is a factor of three times faster than M3.

At the last level of the pyramid, if a final successful match resulted from M2, we also apply M3 to this single patch location to determine whether M3 verifies the point as a valid match as well. If so, the match is given higher weight in the assignment of confidence levels to the stereoscopic heights.

For a stereo image pair, M2 takes the target patch in the reference image and a set of comparison patches within a search window in the comparison image and computes a matching metric. This metric is computed by taking all the bidirectional reflectance factor (BRF) values in each patch, subtracting the mean BRF within the patch from each pixel, and then normalizing by the difference in the maximum and minimum BRFs. Then, the absolute difference between these values in the target patch and the corresponding values in the comparison patch, summed over the area of the patches and normalized by an uncertainty estimate, is tested against a threshold. The M2 metric is defined as follows:

$$S_{M2} = \frac{\sum\limits_{i,j} \left| \left[\frac{R(x_i,y_i)-\langle R \rangle}{R_{\max}-R_{\min}} \right] - \left[\frac{C(x_i,y_i)-\langle C \rangle}{C_{\max}-C_{\min}} \right] \right|}{\sigma_{M2}} \tag{1}$$

where

$R(x_i,y_i)$	reference pixel values at (i,j);
$C(x_i,y_i)$	corresponding value in the comparison image;
$R_{\max}, R_{\min}$	maximum and minimum values within the reference image, respectively;
$C_{\max}, C_{\min}$	maximum and minimum values within the comparison patch, respectively;
$\langle R \rangle$	average value within the reference patch;
$\langle C \rangle$	average value within the comparison patch;
i, j	relative indices within the patches for summation;

and where

$$\sigma_{M2} = \sum_{i,j} \left| \left[\frac{R(x_i, y_i) - \langle R \rangle}{R_{\max} - R_{\min}} \right] \right|. \tag{2}$$

The quantity σ_{M2} is an estimate of the average uncertainty in the numerator of (1). For a reference patch, the x and y values of disparity are those for which S_{M2} is smaller than or equal to threshold T_{M2}. If $S_{M2} > T_{M2}$ the disparity is discarded. If multiple matches from M2 satisfy the threshold criterion, the best match for a patch is defined as the one that minimizes S_{M2}.

M3 is applied in a similar manner as M2, except that medians rather than means as the matching metric is given by

$$S_{M3} = \frac{\text{median}_{ij} \left\{ \left| \frac{R(x_i, y_i)}{\text{median}(R)} - \frac{C(x_i, y_i)}{\text{median}(C)} \right| \right\}}{\sigma_{M3}} \tag{3}$$

where the inclusion of σ_{M3} follows the same rationale as with M2 but is defined differently, i.e., the BRF values in each patch are normalized by dividing by the median value of each patch. A distribution of the absolute values of the differences in the corresponding pixels between the patches is then formed (i.e., the absolute value of the difference of the top right-hand pixels in each patch, etc.), and the final metric is the median of this distribution, normalized by σ_{M3} as shown in (3).

The derivation of σ_{M3} uses similar methodology as in the derivation of σ_{M2}, with the exception that instead of estimating the value in (2) using the mean BRF in the patch, we use the median. With this assumption, we obtain

$$\sigma_{M3} = \text{median}_{ij} \left\{ \left| \frac{R(x_i, y_j)}{\text{median}(R)} - 1 \right| \right\}. \tag{4}$$

A match with $S_{M3} \leq T_{M3}$ is classified as a success. T_{M3} and T_{M2} may be different. As with M2, the match must pass both the threshold and ambiguity tests to be accepted. Briefly, the ambiguity test looks at the spatial distribution of all those points with a metric value within 10% of the minimum. If any such pixel has a difference in returned disparity of more than three pixels in either direction from the original retrieval, the match is deemed to be ambiguous and is rejected.

In a series of experiments (not shown here) it was demonstrated that the M2/M3 metric could not be employed to predict the accuracy of an individual match. Although currently this is a serious drawback in that no internal accuracy estimate can be given, the quality of the matches and resulting heights (see Section IV and [7] for examples) appear to be good enough to meet the scientific requirements. Despite this failing, however, the M2 and M3 metrics do serve as a coarse indicator of match quality.

Fig. 1 shows the results of a study comparing metric value and height accuracy. A "truth" field was constructed by running the superstereo "Gotcha" matcher (see Section IV). A series of threshold cuts on the M2 results were made, and the error distribution between M2 and Gotcha was calculated for all points with the metric value *greater* than the threshold. The means and medians of the height error increase sharply with the threshold value, indicating that the threshold cut does remove a lot of the blunders present in the M2/M3 results.
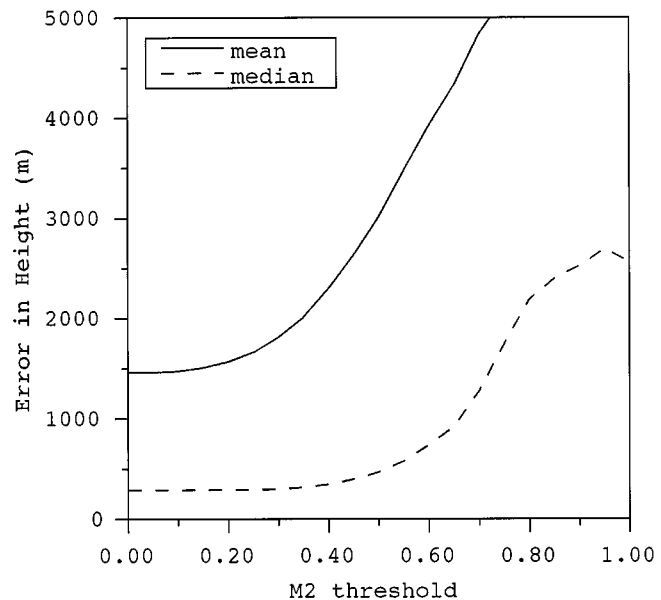

Fig. 1. Difference between cloud-top heights from M2 and Gotcha (MISR path 090, orbit 3708, blocks 82–86). Note the sharp increase in errors as a function of the M2 threshold. Further comparisons are shown in Figs. 5–7.

C. Pyramidal Gruen–Otto–Chau Adaptive Least Squares Correlation (P-Gotcha)

The third matcher employed is referred to elsewhere in this paper as "superstereo." It was originally developed at University College London (UCL) within the framework of the Alvey "real-time 2.5D stereo vision" project [16]. The initial algorithm was described by [17], based on the development of an adaptive patch-based correlation system leaning heavily on least squares adjustment theory proposed in [18]. It was nicknamed Gotcha after the three authors of these papers, Gruen, Otto, and Chau. Subsequently it was rewritten in C++, and then a pyramidal version was implemented to try to deal with height discontinuities and provide speedups. Mathematical details of the underlying theory are given in [19]. This algorithm does not use any predefined seed points, but assumes that at a sufficiently high level of an image pyramid there will be a zero disparity. Random points are chosen, and these are matched using the adaptive least squares correlation (ALSC). The points with the highest precision (i.e., eigenvalue of the variance–covariance matrix) are then used as starting points for a region-growing control strategy that selects the next points on the basis of their precision on the periphery of any previously matched point. P-Gotcha can also be supplied with seed points from algorithms such as NM.

Originally, it was selected as the stereo-matcher for MISR, due to the following:

1) its subpixel acuity (typically to 0.3 pixels);
2) ability to provide a precision matrix that could be used to predict an estimated accuracy based on the eigenvalue of the variance–covariance matrix;
3) robustness against blunders.

The pull-in radius for searches, however, is limited to ± 3 pixels, due to its extreme sensitivity and potential to find
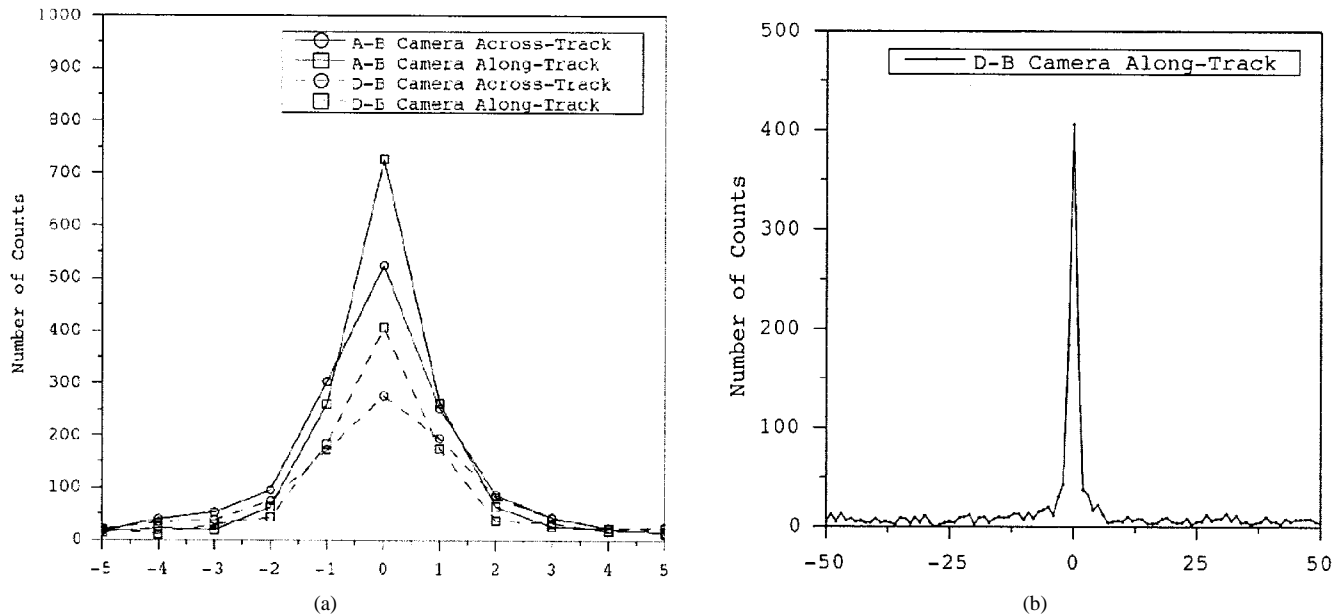
Fig. 2. Pixel-by-pixel comparisons of disparities retrieved using NM with those obtained from M2/M3 (path 201, orbit 3846, blocks 43–51). Only those points with the B camera location in common are included. (a) Blowup of the central peak displaying all four components of the comparison. (b) Complete error distribution for the D/B Along-Track Camera alone. The three other components look very similar.

incorrect matches owing to the use of least squares. Consequently, P-Gotcha requires the selection of seed points and a control strategy based on a "best first" pruning coupled with region-growing [17]. Unfortunately, P-Gotcha is exceedingly slow, due to the 10^5 floating-point operations per match. It was decided not to proceed with this algorithm as the operational matcher even on a parallel implementation, due to the uncertainty surrounding the multiple-instruction multiple-data and single-instruction multiple-data computers at the time [19]. Day and Muller [16] showed that Gotcha had much higher accuracy than the best feature-based stereo-matchers available at the time. Following this, in independent tests conducted over the last decade, it was determined that M2/M3 and NM had similar accuracy to Gotcha at pixel-level acuity and performed much better than any other algorithm in existence at that time. P-Gotcha is, therefore, used here to provide a benchmark to judge the performance of other (faster) algorithms such as M2 and M3. In the cases shown here, NM is used to provide seed points within clouds for P-Gotcha.

IV. INTERCOMPARISON OF STEREO MATCHERS

A number of experiments have been conducted to intercompare stereo-matched disparities from within MISR:L2TC and against those from P-Gotcha. First, we describe experiments to assess how NM matches compare against M2/M3.

The NestedMax and M2/M3 matchers have been compared and found to yield very similar results, as can be seen in Fig. 2(a) and (b). These results were obtained by using the named matchers as part of the MISR wind retrieval (see the accompanying paper [7] for the results of the stereo height retrievals), so separate difference distributions were obtained for the An/B and B/D camera combinations in both the across- and along-track directions. A combination of M2 and M3 was

used to obtain the results, although it should be noted that M2 and M3 have been determined in the past to be very similar. The point is to compare the results from the much sparser and faster feature-matcher (NM) with the area-matchers of M2 and M3 (these are further assessed against P-Gotcha later on in this section).

Fig. 2(a) shows a blowup of the central peak for all four distributions, and Fig. 2(b) shows the complete error distribution for the B/D camera pair along-track matching. Clearly, there is excellent agreement overall between the NM and M2/M3 results for all four components of the comparison. As is expected, the An/B matching shows slightly better agreement than the D/B does, and this is in keeping with the increased difficulty of matching the more oblique angles. All the error (difference) distributions show a long tail that is typical of the NM results, but since each wind calculation uses a large number of individual NM matches, this tail is not considered to be important in the overall MISR retrievals.

When M2/M3 is run on every 16th point to yield comparable coverage to that of NM, the overall wind results appear very similar for a difficult scene with discontinuous wind vectors. Yet, M2/M3 took eight times as long to obtain equally sparse results of similar quality. This shows that using NM is a very good substitute for the area-matchers if only sparse coverage is desired. When full coverage is needed, it is necessary to run M2/M3.

The second series of experiments was to find difficult (or even pathological) examples to intercompare the performance of M2/M3 with P-Gotcha. Two multilayer cloud systems were chosen, which are referenced via their EOS Terra orbit number. The orbit 1010 scene was chosen, as it contains thin as well as high-, middle-, and low-level clouds whereas the orbit 3708 scene was chosen to illustrate the comparison of the MISR data with Gotcha in a test case where the retrievals seemed to be working well.
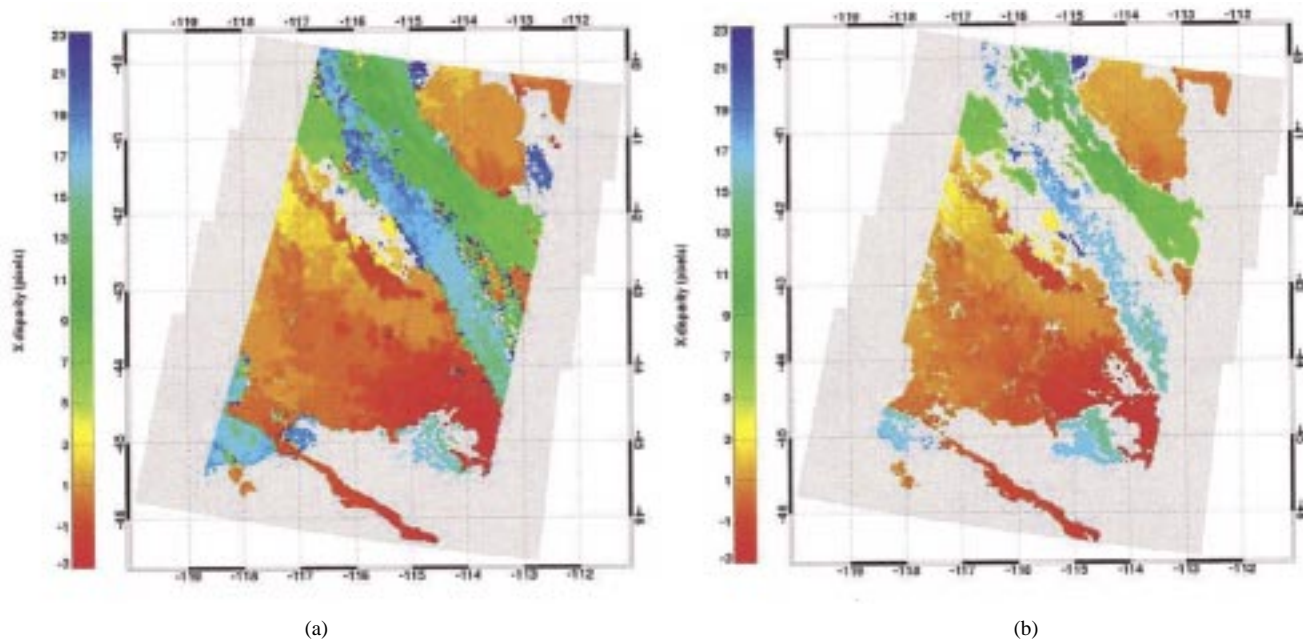
Fig. 3. Comparison between the operational stereo-matched disparities from (a) L2TC and (b) P-Gotcha along-track x disparity. Note the better coverage of the MISR operational product (MISR path 027, orbit 1010, blocks 123–127).

A. Orbit 1010

Fig. 3 compares the x (along-track) disparities from MISR against those from P-Gotcha. Note the much better coverage of MISR and the similarity in the derived disparities. In Fig. 4, a 2-D scatterplot of P-Gotcha versus MISR confirms this visual impression with the vast majority of disparities lying close to the perfect correlation diagonal. A difference statistic indicates that the differences are 0.04 ± 0.89 pixels, which translated into height is equivalent to being in the range 22–489 m. This is well within the overall height resolution of ± 562 m [7]. The height resolution is said to be equal to the height jump caused by a difference in disparity of a single pixel.

B. Orbit 3708

In a different region, this time in the Southern hemisphere and only over ocean, a comparison is shown in Fig. 5 of P-Gotcha with MISR:L2TC that indicates better coverage from P-Gotcha compared to M2/M3, but again the morphology of the cloud-top heights is very similar. The difference histogram shown in Fig. 6 indicates that the P-Gotcha heights are slightly lower (see Section V for a further discussion). A histogram taken of the cloud-top heights in Fig. 7 shows this bias, as well as the close agreement in the shape of the histograms between the two.

V. QUALITY ASSESSMENTS

It is completely unfeasible and unrealistic to use standard photogrammetric approaches to evaluate automated stereo-matcher performance, such as the manual measurement of individual matches, when millions of matches are performed each data day. Qualitative assessments can be made through the creation of stereo anaglyphs and the superimposition of stereo-matched disparities as floating marks. An example of
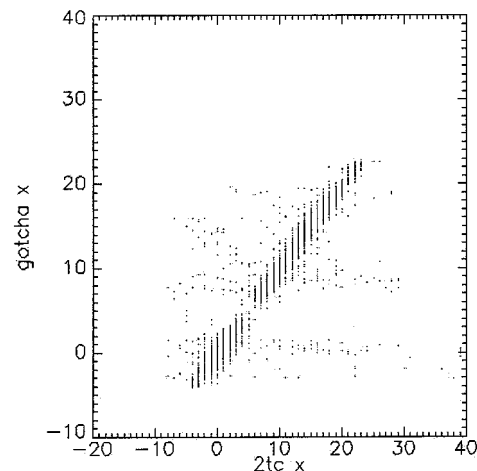

Fig. 4. Two-dimensional scatterplot of the along-track disparity from P-Gotcha versus M2/M3. Note the very high correlation between the two, showing that the 2TC scheme which takes ten times less processing time is producing comparable disparities to one of the most accurate stereo-matchers in existence today.

this of a block produced from the nadir and forward $(+26°)$ view is shown in Fig. 8. When viewed with suitable red-green anaglyph glasses, it can be seen that the matches from MISR fit closely to the observed surface.

The current implementation of MISR:L2TC uses trinocular stereo in a pairwise approach with derived heights from the AnAf compared against those from AnAa. The strategy is to compare these heights and assess whether they are within a predefined threshold. The mean and standard deviation of the AnAf–AnAa height difference distribution is calculated, and all those pairs of heights whose difference is more than two standard deviations away from the mean are rejected. This test is performed on 70.4-km domains. This algorithm was deemed to be more adaptable to current conditions and capable of dealing
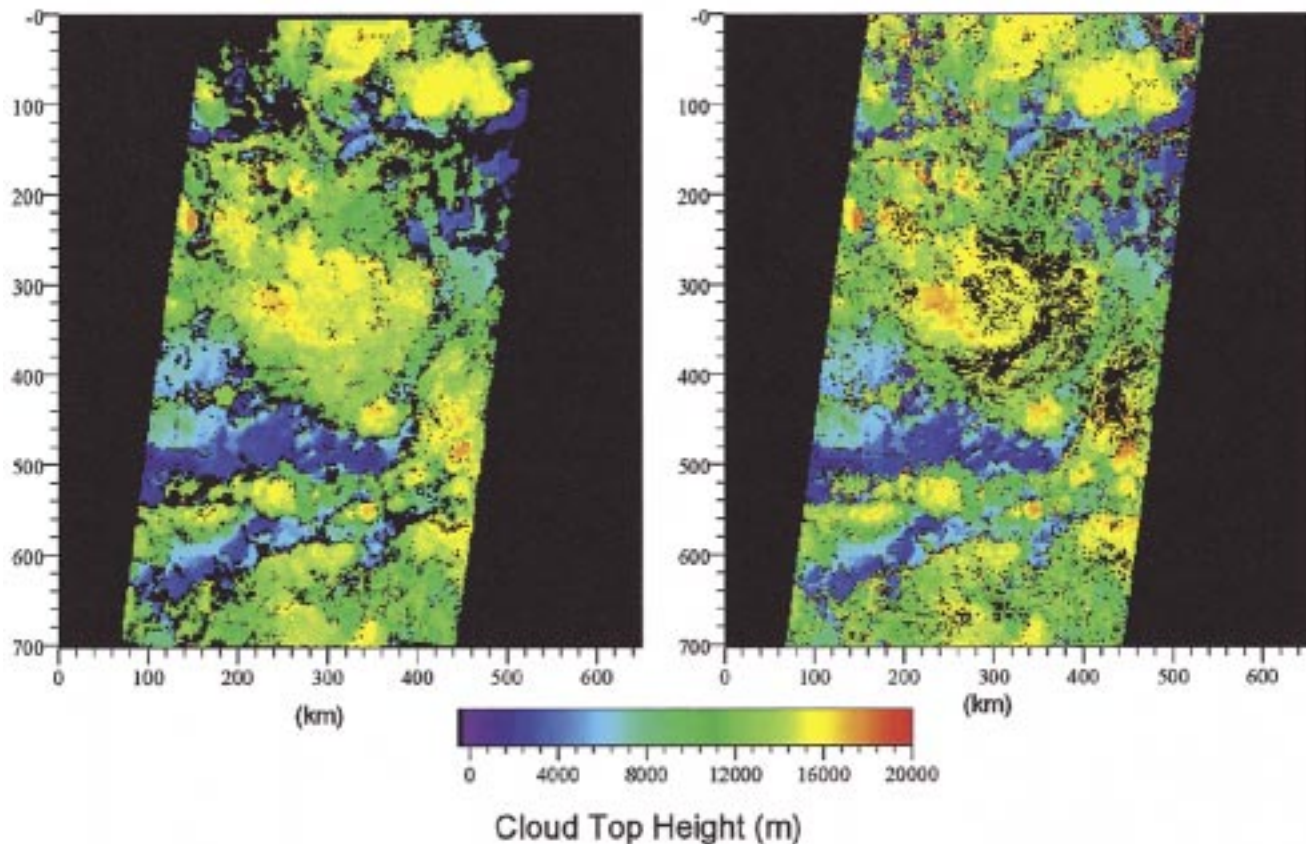
Fig. 5. Comparison of along-track disparity from P-Gotcha (left) and M2/M3 (right) over a multilayer cloud (MISR path 090, orbit 3708, blocks 82–86). Note the similar morphology in both sets of along-track disparities.

with a possible bias in the forward scattering height difference rather than setting a blanket threshold value. When the Af- and Aa-derived heights agree with each other, the highest of the two heights is chosen to minimize possible obscuration problems in the reprojection of the BRFs up to the reflecting layer reference altitude (RLRA) (see [10]).

Although it is not possible to use manual measurements of pixel disparities to assess the accuracy, reliability, or completeness of the stereo-matcher, due to the lack of any meaningful statistical basis (as millions of manual measurements would be required over hundreds of scenes, which are unfeasible), it is possible to assess quantitatively the accuracy of the M2/M3 stereo-matchers used within L2TC using cloud-free areas through an intercomparison of "cloud-top" heights derived from MISR with those from independent ground surface terrain models. This method was first demonstrated in [15].

Two areas were chosen for this study: both in mountainous regions where altitudes range from 0–5 km with mountains primarily oriented north–south in the U.S. Rocky Mountains (MISR path 140, orbit 11 837) and 0–9 km with mountains primarily oriented east–west in the India–China Himalaya Mountains (MISR path 37, orbit 5525).

The standard MISR:L2TC processing scheme was then used to derive cloud-top heights which, owing to the lack of clouds, represent the ground surface. Digital terrain elevation data (DTED) that has been provided by the National Imagery and Mapping Agency on a 30-arcsec grid (so-called DTED-0) was

obtained from http://geoengine.mil. Muller *et al.* [20] showed that the accuracy of DTED-0 over these regions using both independent heights and ERS-1 radar-altimetry-derived heights [21] was around 25–m root mean square. This is more than an order of magnitude greater than what is theoretically possible with MISR [7].

Fig. 9 shows the pixel disparities from M2/M3 for the trioscopic pairs An-Af–An-Aa in the across- and along-track directions for the MISR data over the Rockies, whereas in Fig. 10 a comparison is shown of the MISR digital elevation model (DEM) compared against the corresponding area from the DTED-0 ground surface DEM. Notice the high visual correlation between the terrain morphology in both. In Fig. 11, the difference in heights is shown both as 2-D scatterplot and a corresponding histogram. The overall difference in heights is 0.19 ± 1.11 km. These height values should be compared against the height resolution of 0.562 km. These results indicate that MISR is producing heights close to this value for the Rockies' scene.

Much larger height differences (0.99 ± 2.11 km) (not shown here) were observed for the Himalayas scene, which, although the DEM morphologies are similar, are dominated by 70.4-km domain block artifacts. These are further described and illustrated in [7]. These domain block artifacts are likely responsible for the much larger height differences observed for this scene. It should be noted that the retrieved M2/M3 disparity field itself is smooth, indicating that the problems lie with the difficulty of

Gotcha vs. M2/M3 Intercomparison for MultiLayer System
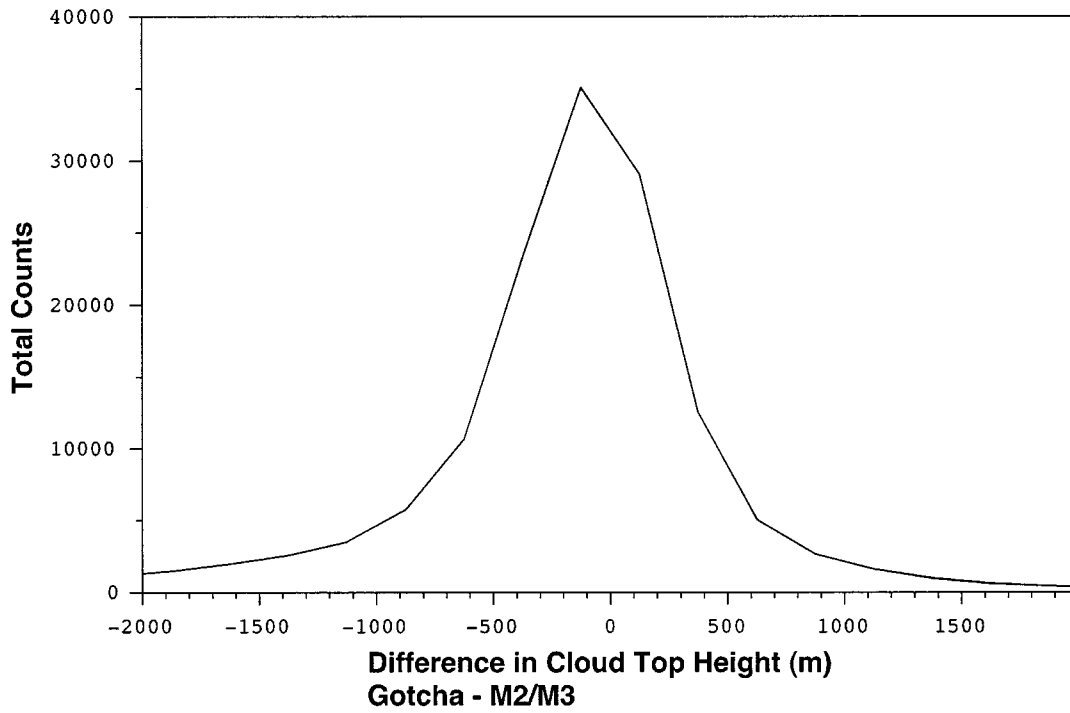### Path 090, Orbit 3708, blks 82-86

Fig. 6. Histogram of P-Gotcha minus M2/M3 along-track disparities (for MISR path 090, orbit 3708, blocks 82–86). Note that the P-Gotcha heights are slightly lower than MISR:L2TC.

Gotcha vs. M2/M3 Intercomparison for MultiLayer System
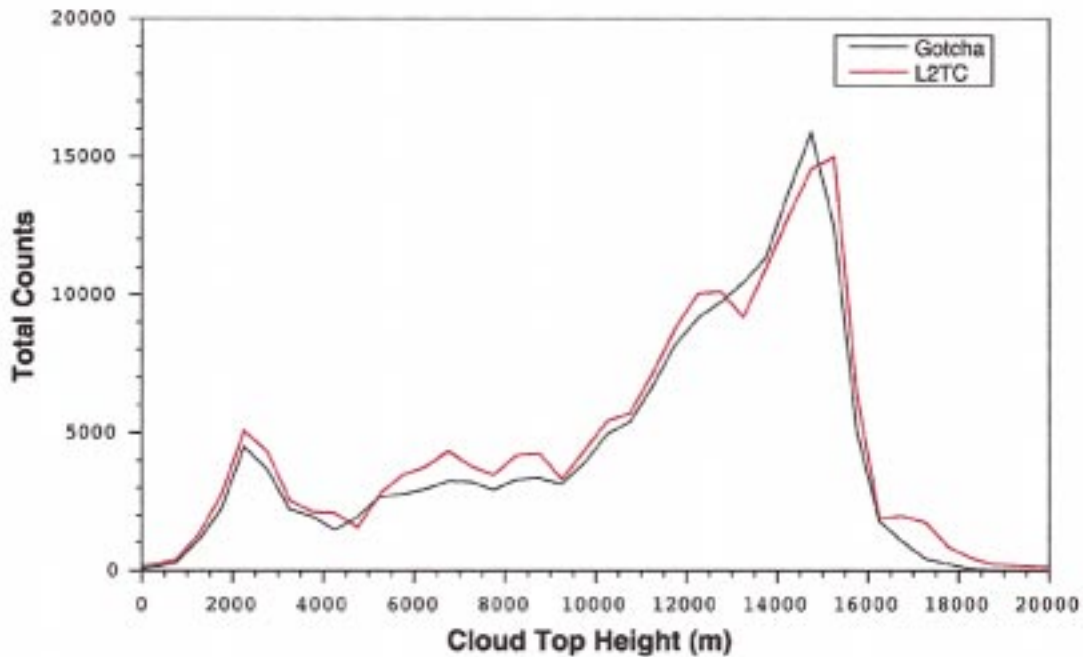### Path 090, Orbit 3708, blocks 82-86

Fig. 7. Profile of along-track disparities from P-Gotcha versus M2/M3 (for MISR path 090, orbit 3708, blocks 82–86). Note the good agreement between the two profiles.
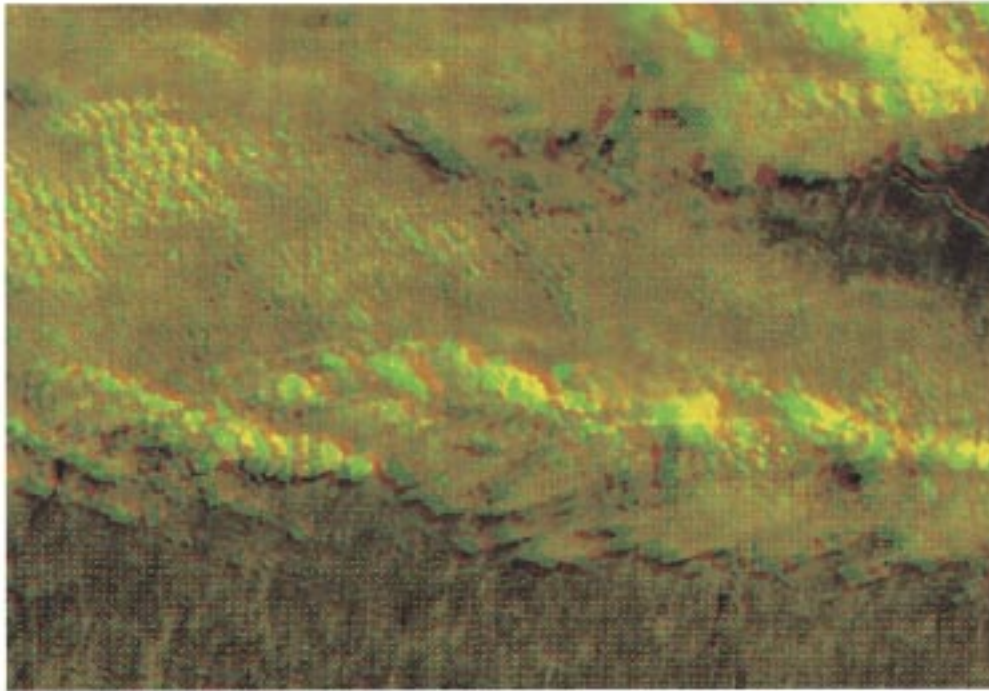
Fig. 8. Example of a stereo anaglyph of a two-level cloud created using An nadir (red) and Af forward off-nadir (green) red MISR images for orbit 1155. Superimposed on this are the matches produced by M2/M3. The observable surface from stereo viewing appears to match well the M2 stereo-matched points.
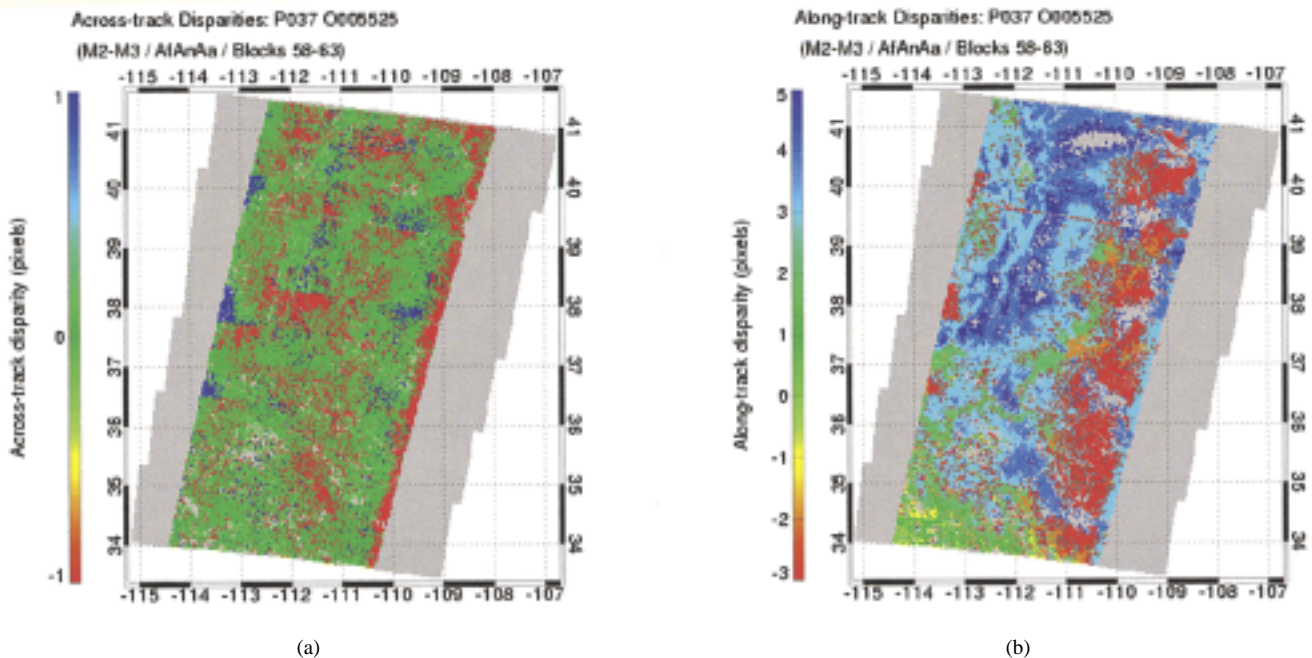

(a)   (b)

Fig. 9. Pixel disparities from M2/M3 for the trioscopic pairs An-Af–An-Aa in the across-track (a) and along-track (b) directions for the MISR data for a clear scene over the Rockies (MISR path 37, orbit 5525, blocks 58–63). Notice the good coverage and the lack of any obvious artefacts, with the exception of the effect of the missing line in the along-track disparities.

retrieving consistent wind fields rather than in applying M2 and M3 themselves.

VI. DISCUSSIONS AND CONCLUSIONS

The preliminary examples shown here demonstrate that a reliable and robust stereo-matcher algorithm has been devel-oped for the production of the world's first set of operational cloud-top heights and motion derived using automated stereo photogrammetry. A comparison of the M2/M3 matches with a "superstereo" algorithm shows close agreement fit (≤ 1 pixel) between the matchers. It should be remembered that the M2/M3 combination is some ten times faster than P-Gotcha. A comparison of heights from cloud-free MISR scenes with
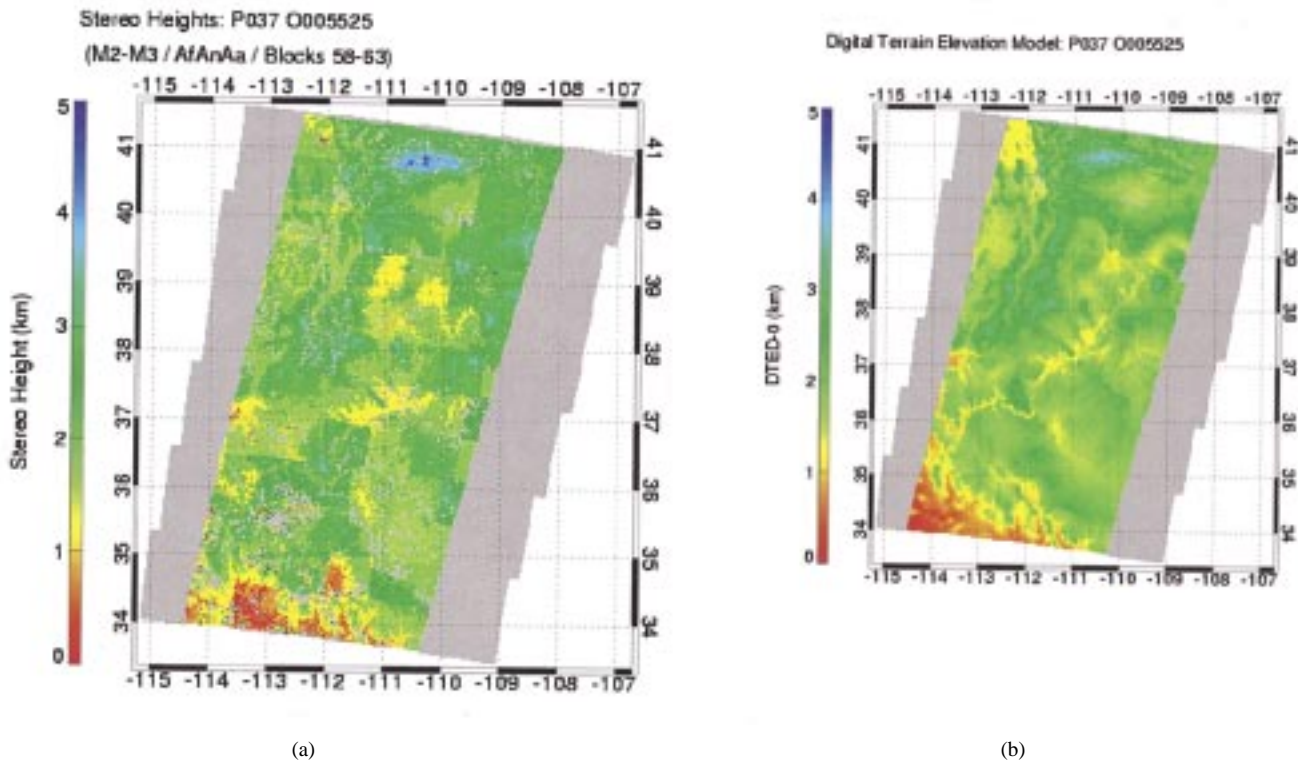
(a)   (b)

Fig. 10.   Comparison of MISR 2TC heights (a) for a clear scene with ground surface heights from DTED-0. (b). Note the similar morphology.

(a)   (b)
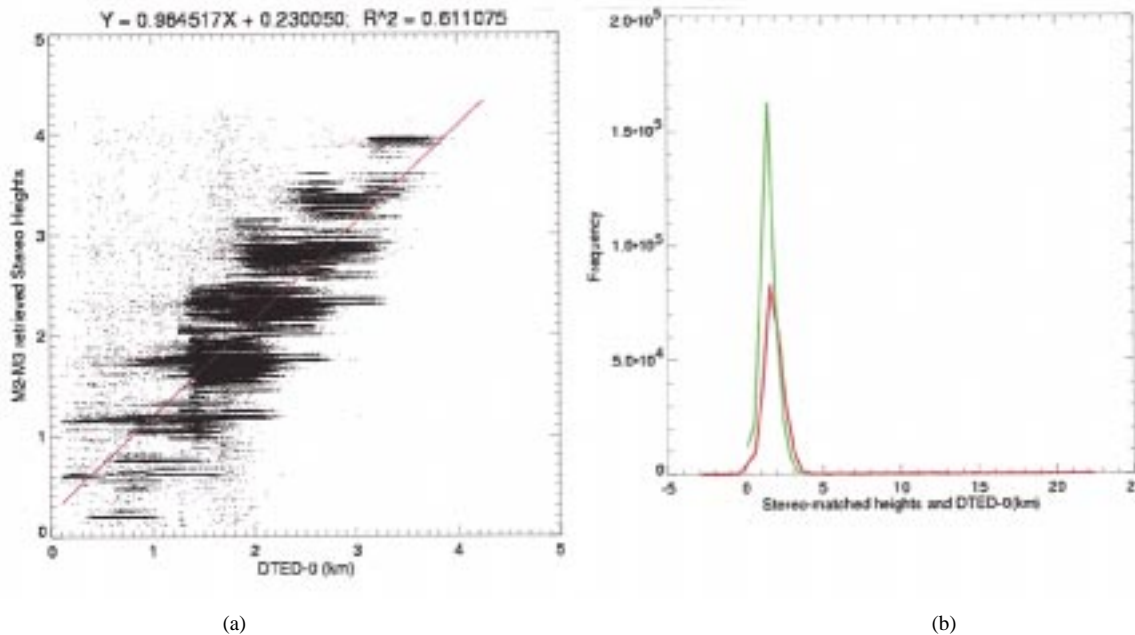
Fig. 11.   Comparison of 2-D scatterplot (a) of MISR 2TC versus DTED-0 and the histogram of the heights (b) for both 2TC and DTED. Note that the scatterplot, which includes the line of best fit and the associated correlation coefficient, shows that there are clusters of points which are biased high from MISR. The histogram shows that there is an equivalent long tail due to blunders (0.78%), which will cause a small bias in the derived statistics.

ground surface terrain models shows height differences close to a height resolution of 562 m.

Noise effects are observed (due to match failures), and occasionally the matcher fails to find suitable matches (due to the lack of contrast) or finds wholly incorrect matches. Currently, it is not possible to state on a pixel-by-pixel basis when this will occur nor provide an estimated accuracy for each derived terrain height estimate.

The current computational cost of the stereo matching techniques is still very high and represents some 23% of the total computational budget for the MISR data products, 80% of which is equally partitioned between M2/M3 and NM. Future research will focus on improving their performance, providing an internal accuracy metric value, and developing a suitable technique to detect and eliminate blunders from the final product.

REFERENCES

[1] A. F. Hasler, "Stereographic observations from satellites: An important new tool for the atmospheric sciences," *Bull. Amer. Meteorol. Soc.*, vol. 62, pp. 194–212, 1981.

[2] A. F. Hasler and K. R. Morris, "Hurricane structure and wind fields from stereoscopic and infrared satellite observations and radar data," *J. Clim. Appl. Meteorol.*, vol. 25, pp. 709–727, 1986.

[3] A. F. Hasler, "Automatic analysis of stereoscopic image pairs from GOES satellites," in *Proc. 3rd Satellite Meteorol. Ocean.*, Anaheim, CA, Feb. 1–5, 1988.

[4] A. F. Hasler, K. Palaniappan, C. Kambhammetu, P. Black, E. Uhlhorn, and D. Chesters, "High-resolution wind fields within the inner core and eye of a mature tropical cyclone from GOES 1-min images," *Bull. Amer. Meteorol. Soc.*, vol. 79, pp. 2483–2496, 1998.

[5] D. P. Wylie, D. Santek, and D. O. Starr, "Cloud-top heights from GOES-8 and GOES-9 stereoscopic imagery," *J. Appl. Meteorol.*, vol. 37, pp. 405–413, 1998.

[6] S. E. Mahani, X. G. Gao, S. Sorooshian, and B. Imam, "Estimating cloud top height and spatial displacement from scan-synchronous GOES images using simplified IR-based stereoscopic analysis," *J. Geophys. Res.—Atmosph.*, vol. 105, pp. 15597–15608, 2000.

[7] C. Moroney, R. Davies, and J.-P. Muller, "Operational retrieval of cloud-top heights using MISR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1532–1540, July 2002.

[8] A. Horvath and R. Davies, "Simultaneous retrieval of cloud motion and height from polar-orbiter multiangle measurements," *Geophys. Res. Lett.*, vol. 28, pp. 2915–2918, 2001.

[9] J. Zong, R. Davies, J.-P. Muller, and D. Diner, "Photogrammetric retrieval of cloud advection and cloud-top height from the Multi-angle Imaging Spectro-Radiometer (MISR)," *Photogram. Eng. Remote Sens.*, 2002, to be published.

[10] D. J. Diner, R. Davies, L. Di Girolamo, A. Horvath, C. Moroney, J.-P. Muller, S. R. Paradise, D. Wenkert, and J. Zong, "MISR level 2 cloud detection and classification algorithm theoretical basis," Jet Propulsion Lab., Pasadena, CA, JPL Tech. Doc. D-11399, Rev. D, 1999.

[11] S. T. Barnard and M. A. Fischler, "Computational stereo," *ACM Comput. Surv.*, vol. 14, pp. 553–572, 1982.

[12] T. Ozanian, "Approaches for stereo matching," *Model. Identif. Contr.*, vol. 16, pp. 65–94, 1995.

[13] W. P. Menzel, "Cloud tracking with satellite imagery: From the pioneering work of Ted Fujita to the present," *Bull. Amer. Meteorol. Soc.*, vol. 82, pp. 33–47, 2001.

[14] M. Rohn, G. Kelly, and R. W. Saunders, "Impact of a new cloud motion wind product from Meteosat on NWP analyses and forecasts," *Mon. Weather Rev.*, vol. 129, pp. 2392–2403, 2001.

[15] A. J. Prata and P. J. Turner, "Cloud-top height determination using ATSR data," *Remote Sens. Environ.*, vol. 59, pp. 1–13, 1997.

[16] T. Day and J.-P. Muller, "Digital elevation model production by stereo-matching spot image-pairs: A comparison of algorithms," *Image Vis. Comput.*, vol. 7, pp. 95–101, 1989.

[17] G. P. Otto and T. K. W. Chau, "Region-Growing algorithm for matching of terrain images," *Image Vis. Comput.*, vol. 7, pp. 83–94, 1989.

[18] A. W. Gruen, "Adaptive least squares correlation: A powerful image matching technique," *S. Afr. J. Photogramm., Remote Sens. Cartogr.*, vol. 14, pp. 175–187, 1985.

[19] M. Holden, M. J. Zemerly, and J.-P. Muller, "Parallel stereo and motion estimation," in *Parallel Architectures and Algorithms for Computer Vision, Image Processing and Neural Networks*, I. Pitas, Ed. Chichester, U.K.: Wiley, 1993, pp. 175–232.

[20] J.-P. Muller, J.-R. Kim, and J. G. Morley, "Quality assessment of global cartographically-derived DEM's using spaceborne altimetry," in *Proc. 25th Annu. Roy. Statist. Soc. Conf.*, Cardiff, U.K., Sept. 8–10, 1999.

[21] J. L. Bamber and J. P. Muller, "Derivation of a global land elevation data set from satellite radar altimeter data for topographic mapping," *J. Geophys. Res. Atmos.*, vol. 103, pp. 32159–32168, 1998.

Jan-Peter Muller received the B.S. degree (with honors) in physics from Sheffield University, Sheffield, U.K., in 1976, the M.S. and D.I.C. degrees in atmospheric physics and dynamics from Imperial College, London, U.K., in 1977, and the Ph.D. degree in planetary meteorology from University College London (UCL), London, U.K., in 1982.

He is currently Professor of image understanding and remote sensing at UCL and has been on the UCL faculty since 1984. He has been a member of the MISR and MODIS science teams since 1990. His research interests include practical applications of machine vision to remote sensing, with a special focus on operational applications of stereo photogrammetry and SAR interferometry technology for applications in climate modeling and weather forecasting. He is currently Chairman of the CEOS Cal/Val Subgroup on Terrain mapping from satellites and Coordinator of the EU-CLOUDMAP2 project providing near real-time cloud products to weather forecasting centers in Europe.

Athula Mandanayake received the B.S. degree (with honors) in engineering from the University of Peradeniya, Sri Lanka, in 1984, and conducted graduate research in real-time pattern recognition in television images at University College London (UCL), London, U.K., from 1987 to 1990.

He is currently Director of Systems Development at 3Q Technologies, Ltd. (formerly, Tcti Plc), Harefield, U.K. He joined 3Q as a Development Engineer in 1997. 3Q designs, develops, and manufactures dynamic and static 3-D capture systems for medical (under 3DMd) and nonmedical (games and movies) markets. In 1991, he joined the Department of Geomatic Engineering (formally Photogrammetry and Surveying) at UCL as a Research Assistant, and from 1993 to 1997 he was a Research Fellow at UCL. His research interests include automated satellite image navigation and 3-D image reconstruction.

Catherine Moroney received the B.S. degree (with honors) in physics from McGill University, Montreal, QC, Canada, in 1990 and the M.S. degree in physics from the University of Toronto, Toronto, ON, Canada, in 1992.

She is currently the Cognizant Design Engineer for the MISR Level 2TC subsystem at the Jet Propulsion Laboratory (JPL), Pasadena, CA. She has been a member of the MISR team since 1994, first at McGill University, then the University of Arizona, Tucson, and now at JPL. Her research interests have focused on the implementation and design of the stereo-matchers used in MISR processing.

Roger Davies received the B.S. degree (with honors) in physics from Victoria University, Wellington, New Zealand, in 1970 and the Ph.D. degree in meteorology from The University of Wisconsin, Madison, in 1976.

He is currently a Principal Scientist at the Jet Propulsion Laboratory (JPL), Pasadena, CA. He joined JPL in 2002 after serving on the faculties of several universities, most recently the University of Arizona, Tucson, where he was Professor of atmospheric sciences and taught courses ranging from radiative transfer to climate theory. He has been a member of the MISR Science Team since its inception, specializing in the multiangle remote sensing of cloud properties.

Dr. Davies is an Honorary Member of the Hungarian Meteorological Society.

David J. Diner (A'01) received the B.S. degree in physics (with honors) from the State University of New York at Stony Brook in 1973 and the M.S. and Ph.D. degrees in planetary science from the California Institute of Technology, Pasadena, in 1977 and 1978, respectively.

He has been with the Jet Propulsion Laboratory, Pasadena, since 1981. He is currently a Principal Member of the Technical Staff and Leader of the Multi-Angle Imaging Science Element in the Earth and Space Sciences Division. He has been involved in numerous NASA planetary and Earth remote-sensing investigations, and is Principal Investigator of the MISR experiment and its airborne counterpart, AirMISR.

Dr. Diner is a member of the American Geophysical Union and the IEEE Geoscience and Remote Sensing Society.

Susan Paradise received the B.S. degree in computer science and mathematics from Purdue University, Lafayette, IN, and the M.S. degree in mathematics from the University of Southern California, Los Angeles.

She is currently with the Jet Propulsion Laboratory, Pasadena, CA. She has worked as a Lead Level 2 Software Engineer for the MISR project and has contributed to many of the algorithm refinements for the Top of Atmosphere/Cloud subsystem.