Robust Stereo Analysis

K. Palaniappan, Yan Huang†, Xinhua Zhuang†, and A. Frederick Hasler‡

Univesities Space Research Association
‡Lab. for Atmospheres, Code 912
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

†Dept. of Electrical and Computer Eng.
Univ. of Missouri
Columiba, MO 65211

Abstract

1 Introduction

One of the most difficult aspects of developing computational algorithms for stereopsis that match the intrinsic capabilities of human vision is the correspondence problem; that is locating the same point, if it exists, in multi-viewed time-varying sensor measurements. Correspondences have been determined using feature-based or region-based matching algorithms with bottom-up or top-down implementations [3]. The bottom-up or low-level approach for stereo analysis includes: i) extracting feature points or area measures in both views, ii) matching the feature points or area measures under certain geometric, illumination, reflectance and object constraints, and iii) computing a depth or height map using the disparity values from correspondences using sensor geometry and scanning configuration. Most stereo algorithms invariably produce errors due to noise, low image or feature content, geometric distortion, depth discontinuities, occlusion, illumination and reflectance changes across the scene and between views, transparency effects leading to multiple matches, and instability of the cameras and sensors during image formation. Such model violations are difficult to handle in a comprehensive fashion. Robust statistical methods have recently been applied to a variety of computer vision problems including motion estimation [10] [11][6], surface recovery from range data [9], and image segmentation [2]. Robust methods offer a powerful alternative to smoothness and regularization constraints to mitigate the effects of model errors. A new multistage adaptive robust (MAR) algorithm combined with a multiresolution coarse-to-fine matching model is developed for robust stereo analysis.

Stereo analysis of remotely sensed images is useful for a variety of applications including cloud height measurement and digital terrain models. Stereopsis both in human vision and in remote sensing relies on the same principal of parallax but the matching problem is made more difficult due to the extremely long baselines with satellite geometries. The estimation of cloud-top structure using multiple satellite views (both geosynchronous and low earth orbiting) is an extremely challenging problem due to the time-dependent dynamics of the satellite-based imaging instruments and complex fractal-like surface properties of clouds [7]. Robust estimation methods have the advantage of possessing adaptive properties to local depth changes in the prescence of outliers. Time sequential stereoscopic observations of clouds from meteorological satellites provide a basic analysis tool for a broad spectrum of applications [4] including numerical weather prediction, cloud modeling, and global climate understanding.

A two-step robust stereo analysis algorithm is developed and applied to the satellite-based cloud height estimation problem. The first step involves estimating initial disparities (depth map or height field) provided by a hierarchical coarse-to-fine stereo analysis algorithm that uses regularization and image warping and is described in the next section. For the second step the matching model using to derive robust estimators is described in Section 3, followed by a description of the multi-stage robust estimation process for handling outliers and irregularities in the initial stereo disparity field. The incorporation of a multistage robust statistical process using a general matching model leads to improved performance.

2 Parallel Automatic Stereo Analysis Algorithm

The first step involves estimating an initial set of dense disparities using a stereo pair of images. An Automatic Stereo Analysis (ASA) algorithm [5] [8] has been developed and implemented on the commercially available massively parallel supercomputer MasPar MP-2 at the NASA Goddard Space Flight Center. The parallel implementation enables the estimation and visualization of cloud surfaces interactively in realtime [7]. The ASA uses a hierarchical approach with a coarse-to-fine resolution implementation based on varying the size of the template windows (matching blocks) rather than filtering the image. The ASA can search for both horizontal and vertical disparities though for the experiments in this paper the disparities have been constrained to be along horizontal epipolar lines only.

Suppose that the stereo image pair consists of a reference image f_r and a test image f_t. The ASA algorithm includes the following steps:

0-8186-7190-4/95 $4.00

- *Preprocessing.* The pixel brightness in both the test and reference images are rescaled to the full dynamic range typically 0 to 255. The two images are compared to determine the maximum and minimum offsets in the x and y directions which is used to shift the test image and center the offsets with respect to the reference image if required. The starting resolution (matching template size) is determined automatically using information about local image texture. The neighborhood search area for block matching, is a parameter that can be estimated using maximum image offset information as a guideline.

- *Determining matches.* For each reference image pixel $\boldsymbol{z}_n$, a matching block $\mathcal{M}_n$ centered at $\boldsymbol{z}_n$ is defined and its best match $\boldsymbol{z}'_n$ within the search area $\mathcal{R}$ of the test image is found using the normalized cross-correlation criterion. Denoting the disparity $\boldsymbol{z}'_n - \boldsymbol{z}_n$ by $\boldsymbol{d}_n$, then $\boldsymbol{d}_n = \arg\min_{\boldsymbol{d}\in\mathcal{R}} \frac{C_{f_r f_t}(\boldsymbol{d})}{\sqrt{C_{f_r f_r}(\mathbf{0})\cdot C_{f_t f_t}(\mathbf{0})}}$, where $C_{f_1 f_2}(\boldsymbol{d}) = \sum_{\boldsymbol{z}_i\in\mathcal{M}_n}(f_1(\boldsymbol{z}_i) - \overline{f_1(\boldsymbol{z}_n)})(f_2(\boldsymbol{z}_i - \boldsymbol{d}) - \overline{f_2(\boldsymbol{z}_n - \boldsymbol{d})})$, with $\overline{f_k(\cdot)}$ computed as the local sample mean.

- *Removal of "bad match" areas in the disparity map.* Under the constraint that surfaces are locally continuous the disparity function must also be continuous. So large disparity gradients greater than a threshold are considered to be due to bad matches. Such outlier disparities are interpolated using an iterative approach [5].

- *Smoothing the disparity map.* After interpolation, the disparity function may need to be smoothed to reduce the effects of noise and outlier disparities that could not be interpolated to obtain a more continuous warping function.

- *Hierarchical warping.* The disparity map estimated at each resolution is added cumulatively and is used as a two-dimensional distortion function to *geometrically correct* the original test image for stereopsis effects. The warped image values are obtained by resampling the test image using linear interpolation.

The above algorithm is iteratively applied to obtain a more refined disparity map. The disparities estimated at each resolution should decrease to zero and the warped test image should closely match the reference image. The ASA algorithm has good performance for a variety of satellite-based stereo cloud data [5] [8]. However, the ASA algorithm has difficulty capturing the fine scale surface structures, handling occlusion, noise, and other model errors. In the second step, the resultant ASA-based disparity map is used to initialize the proposed multi-stage adaptive robust (MAR) stereo algorithm for refining the depth map.

3 A Constrained Image Matching Model

Pixel-wise correspondence matching is known to be inherently ambiguous and an ill-posed problem. Consequently, many different smoothness constraints have been proposed to overcome ill-conditioning [1]. The selection of a smoothness constraint largely determines the nature and performance of the algorithm and is application dependent. In the ASA algorithm, an implicit smoothness constraint is imposed in that all the pixels within a matching template in the reference image are assumed to share the same disparity, which implies that the reference matching block is mapped to a corresponding block in the test image. This local uniform disparity constraint ignores small scale structure leading to systematic biases and causes outlier problems when disparity discontinuities exist within the matching template. An improved matching model is first developed to determine a robust estimator for disparity estimation.

3.1 Geometric Constraint

The true disparity function is linearly related to the depth function of the scene and may be arbitrarily complex. Viewing the disparity function as a 3-D surface, block based matching algorithms approximate each local patch of the disparity surface with a horizontal planar patch. Apparently, this type of approximation fails to account for the surface orientation of each local patch. A better model of the disparity surface is to use an affine transform-based geometric constraint.

Suppose two image patches I_r and I_t are the corresponding matching areas in the reference and the test images, respectively. Let $\boldsymbol{z}_i = (x_i, y_i)^T \in I_r$ and $\boldsymbol{z}'_i = (x'_i, y'_i)^T \in I_t$. The smoothness constraint between I_r and I_t can be described by the following affine transformation:

$$\boldsymbol{z}'_i = \boldsymbol{A}\boldsymbol{z}_i + \boldsymbol{a}, \qquad i = 1, \cdots, N, \tag{1}$$

where N is the size of I_r and the 2×2 matrix $\boldsymbol{A}$ and the 2×1 vector $\boldsymbol{a}$ together characterize a geometric transformation from $\boldsymbol{z}_i$ to its match point $\boldsymbol{z}'_i$. Assuming a horizontal epipolar stereopsis constraint, (1) can be more simply described by,

$$\begin{aligned} x'_i &= A_1 x_i + A_2 y_i + a, \\ y'_i &= y_i. \end{aligned} \tag{2}$$

Thus, the disparity d_i between the i-th pixel pair (x_i, y_i) and (x'_i, y'_i) is given by $d_i = x'_i - x_i$ assuming horizontal epipolar lines, or

$$d_i = (A_1 - 1)x_i + A_2 y_i + a. \tag{3}$$

Notice that the disparity function is now modeled as an arbitrary 3-D plane parameterized by A_1, A_2 and a. Therefore, a meaningful interpretation of this geometric model is that the disparity function is approximated by using arbitrary plane patches instead of horizontal ones. This certainly provides a better modeling of the cloud depth surface as the changing cloud

depth can now be better depicted. To prevent over-modeling, however, we still take the pixel-wise modeling approach. That is, for every pixel, the model is established over a surrounding patch I_r centered at that pixel and the evaluation result applies only to the center pixel.

3.2 Intensity Constraint

As the primitive property, matching areas in the reference and test images should have a high intensity correlation. Let $f_r(\cdot)$ and $f_t(\cdot)$ denote the intensity functions of the reference and test images, respectively. The intensity correlation between I_r and I_t is imposed by

$$f_t(x'_i, y'_i) = cf_r(x_i, y_i) + b, \tag{4}$$

where c and b characterize the contrast and brightness difference between the two image patches, respectively.

3.3 Matching Measure

Based on the foregoing geometric constraint and intensity constraint, an appropriate matching measure is defined by

$$s_i = f_t(A_1x_i+A_2y_i+a, y_i)-(cf_r(x_i, y_i)+b), \quad i = 1, \cdots, N. \tag{5}$$

If the two image patches I_r and I_t matches well, the s_i values should be close to zero and can be modeled as a Gaussian distribution.

3.4 Parameter Optimization

To find the best match for pixel (x_n, y_n), its surrounding neighborhood of appropriate size, e.g., 5×5, is chosen as the match patch I_r, based on which the parameter set (A_1, A_2, a, c, b) is to be evaluated and used to find the disparity d_n. The geometric parameters are intended to approximate the local disparity surface and should be locally determined but it is more reasonable to determine the intensity parameters on a larger scale. For example, a uniform intensity patch can be matched to another uniform patch by an infinite number of choices of c and b. When the image contrast is not uniform over the whole image domain due to regional luminance difference, the image can be sub-divided into large blocks and a separate set of (c, b) for each block can be estimated.

Suppose the size of the whole image or a large image block is M. Suppose the correspondences $((x_i, y_i), (x'_i, y'_i))$, $i = 1, \cdots, M$ have been estimated in the first step using the ASA algorithm. Then a good estimation of the intensity parameters c and b is determined by the least-squares solution to (4). The intensity parameters c and b are then fixed in determining the geometric parameters.

To minimize the sum of matching errors $\sum_{i=1}^{N} s_i^2$, nonlinear optimization methods need to be used as seen from (5) since $f_r(\cdot)$ and $f_t(\cdot)$ are nonlinear image intensity functions. To reduce the local extrema problems often associated with nonlinear optimization methods, we compute the initial values for the geometric parameters A_1, A_2 and a by using the matching results from the ASA algorithm, as detailed in the following. The specific optimization methods are presented in the next section.

After all the parameters are appropriately determined, the disparity d_i for pixel (x_i, y_i) is computed from (3). The above model-based optimization approach is a generalization of the block matching algorithm used by the ASA algorithm as the initial step of the robust stereo estimation process. When $A_1 = 1, A_2 = 0$, the geometric model reduces to the block matching model with a being the variable measuring the translational component. In the block matching algorithm, the parameter (translation a) is exhaustively searched and compared within a certain disparity range to avoid local extrema problems but the searching parameter does not have sufficient precision without using an interpolation scheme. In the model-based optimization approach, the parameters are determined using nonlinear optimization of an objective function which may suffer from local extrema problems. However, very high precision of the parameters can be achieved to match the local image statistics. The two approaches are combined to take advantages of their respective merits. The ASA algorithm is used to provide the required good initialization for the proposed multi-stage adaptive robust (MAR) stereo algorithm.

4 Multi-Stage Adaptive Robust Disparity Estimation

The constrained matching model provides a better approximation of the true disparity surface but surfaces are still assumed to be locally continuous. Local surface or disparity discontinuities, irregularities, shape distortions, partial occlusions, noisy image intensities and other model violations are accommodated using a multi-stage adaptive robust algorithm. A practical matching algorithm needs to address a wide variety of almost arbitrary model violations using a well-managed regular model.

4.1 Adaptive Refinement Scheme

When the matching patch I_r contains depth boundaries, shape distortions or occlusions, I_r can be considered as multi-structured, consisting of multiple smaller patches, each of which contains a portion of the original data set supporting a distinct matching model. However, the shape of each small patch is unknown and may not even be decided by spatial proximity. Therefore, it is almost impossible to artificially partition the patch I_r in advance.

Under such a condition, the matching errors in (5) are no longer Gaussian distributed under any set of model parameters; instead, they form a mixture of Gaussian distributions and possibly outliers under appropriate sets of (unknown) model parameters. In particular, each valid model is supported by a portion of data whose matching errors appear to be Gaussian distributed under an appropriate set of model parameters. Therefore, without advanced knowledge about the patch shapes and the parameter values, it is required to concurrently perform the the data grouping and parameter estimation. To achieve this task, we propose to use robust estimation methods that

are able to accomplish correct parameter estimation without much influenced by irrelevant data or outliers. Specifically, the MF-estimator [11] will be used since it can effectively handle a mixture of different models in a single (outlier-contaminated) data set. However, since the MF-estimator is computationally demanding, we propose an multi-stage adaptive approach where increasingly sophisticated methods are employed only when necessary. It offers a good balance between sophisticated modeling and computation efficiency.

For each pixel (x_n, y_n) in the reference image f_r, a 5×5 *working window* $\mathcal{M}_n$ centered at (x_n, y_n) is defined. This working window is treated as the image patch I_r in the foregoing matching model (4) and (2). The model parameters are to be determined based $\mathcal{M}_n$ and used to compute the disparity d_n for pixel (x_n, y_n) as in (3).

As discussed in the foregoing section, the intensity parameters c and b are globally determined and then fixed. The geometric parameters A_1, A_2 and a are estimated by using one of three estimators, i.e., least-squares estimator, bi-weight estimator and MF-estimator [11], embedded in a multi-stage adaptive approach as follows.

4.2 Least-Squares Estimator

The least-squares estimate of the matching parameters A_1, A_2 and a within the work window $\mathcal{M}_n$ is determined by

$$\min Q_{ls} = \sum_{i=1}^{N} {s_i}^2, \tag{6}$$

where N is the size of $\mathcal{M}_n$. To achieve the minimization, we basically follow the gradient-descent iterative method. Let A_1 be θ_1, A_2 be θ_2 and a be θ_3. Denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$. It can be easily derived that

$$\frac{\partial Q_{ls}}{\partial \theta_k} = 2 \sum_{i=1}^{N} s_i \frac{\partial s_i}{\partial \theta_k}, \quad k = 1, 2, 3, \tag{7}$$

where

$$\begin{aligned} \frac{\partial s_i}{\partial \theta_1} &= \frac{\partial f_t(A_1 x_i + A_2 y_i + a, y_i)}{\partial x} x_i, \\ \frac{\partial s_i}{\partial \theta_2} &= \frac{\partial f_t(A_1 x_i + A_2 y_i + a, y_i)}{\partial x} y_i, \\ \frac{\partial s_i}{\partial \theta_3} &= \frac{\partial f_t(A_1 x_i + A_2 y_i + a, y_i)}{\partial x}. \end{aligned} \tag{8}$$

To minimize Q_{ls}, we let

$$\sum_{k=1}^{3} \frac{\partial Q_{ls}}{\partial \theta_k} \Delta\theta_k \leq 0, \tag{9}$$

which leads to

$$\sum_{i=1}^{N} s_i \left(\sum_{k=1}^{3} \frac{\partial s_i}{\partial \theta_k} \Delta\theta_k \right) \leq 0, \tag{10}$$

which, in turn, can be sufficiently satisfied if we let

$$\left(\frac{\partial s_i}{\partial \theta_1}, \frac{\partial s_i}{\partial \theta_2}, \frac{\partial s_i}{\partial \theta_3} \right) \Delta\boldsymbol{\theta} = -s_i, \quad i = 1, \cdots, N. \tag{11}$$

Let

$$\boldsymbol{X} = \begin{pmatrix} \frac{\partial s_1}{\partial \theta_1} & \frac{\partial s_1}{\partial \theta_2} & \frac{\partial s_1}{\partial \theta_3} \\ \frac{\partial s_2}{\partial \theta_1} & \frac{\partial s_2}{\partial \theta_2} & \frac{\partial s_2}{\partial \theta_3} \\ \vdots & \vdots & \vdots \\ \frac{\partial s_N}{\partial \theta_1} & \frac{\partial s_N}{\partial \theta_2} & \frac{\partial s_N}{\partial \theta_3} \end{pmatrix} \tag{12}$$

Then, the $\boldsymbol{\theta}$ can be updated by

$$\Delta\boldsymbol{\theta} = -\boldsymbol{X}^{\dagger} (s_1, \cdots, s_N)^T. \tag{13}$$

where $\boldsymbol{X}^{\dagger}$ is the psedo-inverse of matrix $\boldsymbol{X}$. Thus, starting for certain initial values (see next section), A_1, A_2 and a can be iteratively solved.

When the above iteration converges, we compute model error variance

$$\sigma^2 = \sum_{i=1}^{N} \frac{1}{N} (f_t(A_1 x_i + A_2 y_i + a, y_i) - (c f_r(x_i, y_i) + b))^2 \tag{14}$$

If the true disparities within $\mathcal{M}_n$ are uniform, the matching model (2) and (4) should hold well and thus σ should be small. In contrast, when disparity discontinuities, shape distortions or occlusions occur in $\mathcal{M}_n$, the matching model is violated and a large σ results. Therefore, we define a threshold U to measure the reliability. If $\sigma > U$, the above estimation is considered unreliable and will be re-estimated using the following bi-weight estimator. Otherwise, the disparity d_n for pixel (x_n, y_n) is obtained by using (3).

4.3 Bi-weight Estimator

The bi-weight estimator is one of conventional robust estimation methods intended to suppress the influence of outliers in parameter estimation. Typically, the bi-weight estimator can tolerate up to $30 - 40\%$ percent of outliers. Therefore, it can be used to handle cases where the disparities within $\mathcal{M}_n$ are partially irregular. In practice, there exist many such cases since the moving window $\mathcal{M}_n$ can partly cover disparity discontinuities, shape distortions or occlusions.

The bi-weight estimator is essentially a weighted least-squares estimation method. It can be derived as minimizing the following objective function:

$$\min Q_{bw} = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i {s_i}^2 \tag{15}$$

where the weight function w_i is given by

$$w_i = \begin{cases} (1 - e_i^2)^2 & |e_i| \leq 1 \\ 0 & |e_i| > 1. \end{cases} \tag{16}$$

where

$$e_i = \frac{s_i}{c \cdot s_{med}}, \tag{17}$$

with s_{med} being the median of $|s_i|$, $i = 1, \cdots, N$, and c being a tuning constant typically between 2 and 10.

The minimization of Q_{bw} can be achieved similarly as in the least-squares estimator. With w_i being considered as constant at each iteration step, it can be easily derived that the parameter $\boldsymbol{\theta}$ should be updated by

$$\Delta\boldsymbol{\theta} = -\boldsymbol{X}^{\dagger}(w_1 s_1, \cdots, w_N s_N)^T. \quad (18)$$

And the weight w_i, $i = 1, \cdots, N$, are updated as in (16) and (17).

After convergence, we compute model error variance

$$\sigma^2 = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i (f_t(A_1 x_i + A_2 y_i + a, y_i) - (c f_r(x_i, y_i) + b))^2. \quad (19)$$

When a small percentage of the true disparities within $\mathcal{M}_n$ are irregular, the bi-weight estimator will reject them as outliers by assigning small or zero weights and yield a valid matching model (indicated by a small σ). However, there are two cases that more sophisticated MF-estimator needs to be further applied: (1). σ is large. This indicates that the disparities are more complex structured and the resultant estimation is still unreliable. (2). σ is small but the center pixel (x_n, y_n) is rejected as an outlier. This indicates that $\mathcal{M}_n$ may be multi-structured and the center pixel belongs to a patch other than the extracted one.

4.4 MF-estimator

The MF-estimator is developed as a generic robust estimator to handle multi-structure data set [11]. Basically, it can estimate a valid model from a mixture data set without much influenced by the irrelevant data (data associated with other models) and outliers. In fact, all other irrelevant data appear to be outliers with respect to one particular valid model. Once a valid model is established by the MF-estimator over the nonoutliers area, the same procedure can be applied over the rest of the outlier areas to find the second matching model and continue in such a recursive fashion. As a typical algorithm behavior, a matching model supported by a larger portion of data will be detected earlier in the recursion.

The MF-estimator is formulated as the maximum likelihood estimator of a general regression model. It maximizes the following objective function at each fixed partial model $t \geq 0$ [11]:

$$Q_{mf}(A_1, A_2, a, \sigma; t) = \sum_{i=1}^{N} ln(g_i + t), \quad (20)$$

where

$$g_k = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{s_i^2}{2\sigma^2}). \quad (21)$$

Let A_1 be θ_1, A_2 be θ_2 and a be θ_3. It can be derived that

$$\frac{\partial Q_{mf}}{\partial \theta_k} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \lambda_i s_i \frac{\partial s_i}{\partial \theta_k}, \quad k = 1, 2, 3,$$

$$\frac{\partial Q_{mf}}{\partial \sigma} = \frac{1}{\sigma} \left\{ \frac{1}{\sigma^2} \sum_{i=1}^{N} \lambda_i s_i^2 - \sum_{i=1}^{N} \lambda_i \right\}, \quad (22)$$

where

$$\lambda_i = \frac{g_i}{g_i + t}. \quad (23)$$

Following the gradient-descent method, it can be similarly derived that the parameter $\boldsymbol{\theta}$ should be updated by

$$\Delta\boldsymbol{\theta} = -\boldsymbol{X}^{\dagger}(\lambda_1 s_1, \cdots, \lambda_N s_N)'. \quad (24)$$

And σ is updated by

$$\sigma^2 = \frac{1}{\sum_{i=1}^{N} \lambda_i} \sum_{i=1}^{N} \lambda_i s_i^2. \quad (25)$$

The iteration continues until convergence.

The foregoing maximization is performed for each increasingly larger t starting at $t = 0$ until the result passes a normality test. The normality test is conducted as follows.

- *Size verification.* The set of nonoutliers is collected as

$$G = \{i : g_i > t,\ i = 1, \cdots, N\} \quad (26)$$

  To make sure the extracted matching model has enough data support, we require the number of nonoutliers to be large than a predefine threshold L. If $\#(G) < L$, the normality test fails.

- *Reliability verification.* The resultant σ is a measure of the goodness of the detected matching model. If $\sigma > U$, the normality test fails.

If the normality test fails, the maximization is continued at the next larger t until a predefined upper bound for t is reached, at which stage, no solution is declared. If the normality test passes, a valid matching model is detected. In this case, if the associated non-outlier data G set contains the center pixel (x_n, y_n), then the estimation is completed. However, if the center pixel does not belong to the detected nonoutlier patch G, we need to continue applying the MF-estimator on the rest of the outlier patch until a valid model containing the center pixel is found or no solution is declared. See the algorithm description in the following for details.

4.5 Post Processing

Using the same approach as in the first ASA step, two post processing techniques known as line interpolation and neighborhood smoothing are applied to remove the occasional bad matches and perform disparity smoothing.

Occasionally, bad matches result due to numerical instabilities in the above estimators, such as divergence, singular matrix inversion, etc. Thus, we apply the line interpolation technique [8] to detect such disparity values. Specifically, for each pixel, a line is constructed using the disparity values of its left and right

neighbors. If the disparity of the current pixel is off the line by a predefined number of pixels, it is considered as a bad match and replaced by the interpolation value.

Disparity smoothing is performed by averaging the disparity values over a mask consisting of the current pixel and its four immediate neighbors.

5 Algorithm Description

The multi-stage adaptive robust (MAR) stereo matching algorithm is summarized as follows:

- *Preprocessing*

  Interpolate the test image f_t by using the cubic spline interpolation techniques.

  Correlate the contrast and intensity difference between the two stereo image pair f_r and f_t. For each large block of the reference image, a contrast factor c and a intensity difference b are computed using (4) and stored.

- *Model Estimation and Disparity Computation*

  Select the thresholds $U, L, t, \Delta t$. For each valid pixel (x_n, y_n) in the reference image f_r, establish its working window $\mathcal{M}_n$ of size 5×5 and centered at (x_n, y_n). Collect the needed data values within $\mathcal{M}_n$ for the following multi-stage estimation procedure.

  ◇ *Initialization.*

  Let A_1 be θ_1, A_2 be θ_2 and a be θ_3. Denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$. Compute the initial values for the geometric matching model $\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}$ by using the least-squares solutions to,

  $$x_i' = \theta_1 x_i + \theta_2 y_i + \theta_3, \quad i = 1, \cdots, N, \qquad (27)$$

  where x_i and y_i are the x-coordinate and y-coordinate of the i-th pixel in $\mathcal{M}_n$, respectively, and $x_i' = x_i + d_i^0$ with d_i^0 being the initial disparity provided by the ASA step.

  ◇ *Stage I. Least-squares estimation*

  Compute the matrix $\boldsymbol{X}$ according to (12) and (8). and the update term $\Delta\boldsymbol{\theta}$ by (13). Once the update term is small compute the model reliability measure σ by (14). If $\sigma < U$, go to stage IV. Otherwise, proceed to stage II.

  ◇ *Stage II. Bi-weight estimation*

  Compute the matrix $\boldsymbol{X}$ according to (12) and (8), the weights w_i, $i = 1, \cdots, N$, by (16) and (17) and the update term $\Delta\boldsymbol{\theta}$ by (18). Once the update term is small compute the model reliability measure σ by (19). If $\sigma < U$, go to stage IV. Otherwise, proceed to stage III.

  ◇ *Stage III. MF-estimation*

  Use the initial model $\lambda_i^{(0)} = 1$, $i = 1, \cdots, N$ and compute the initial $\sigma^{(0)}$ by (25). Compute the matrix $\boldsymbol{X}$ according to (12) and (8), the λ_i, $i = 1, \cdots, N$, by (16) and (17) and the update term $\Delta\boldsymbol{\theta}$ by (24). Once the update term is small compute the model reliability measure σ by (25) and the non-outlier data set G by (26).

  ◇ *Stage IV. Disparity computation*

  If a valid matching model containing the center pixel (x_n, y_n) is found in the foregoing stages, compute the disparity d_n by (3). If no solution is declared in the foregoing stages, compare all the extracted valid matching models if any, and the original matching obtained by the ASA step, and choose the one that leads to the smallest matching measure s_n.

- *Post Processing*

  Removal of bad matches using local fitting and local smoothing to reduce the effects of noise.

6 Performance Evaluation

The ASA and MAR algorithms are evaluated using satellite multispectral data from the NOAA/AVHRR instrument for Hurricane Andrew. Fig. 1(a) is a visible image of Hurricane Andrew after passing over Florida on August 25, 1992. Fig. 1(c) shows an independent infrared channel information of the same scene, which can be used as the true height map. Taking Fig. 1(a) as the reference image and Fig. 1(c) as the true disparity map, we create a synthetic visible stereo image (test image) as shown in Fig. 1(b), with the maximum disparity being 25 pixels.

The resultant disparity map output from the ASA step implemented on the massively parallel Maspar computer is shown in Fig. 1(d). The resultant disparity map is then used to warp the reference image. Ideally, the warped image should perfectly resemble the test image. The difference between the warped image and the test image can also be used to assess performance. The results from the second MAR estimation step of stereo analysis are given in Fig. 1(b) and (e), respectively. In the MAR algorithm, at $U = 2.0$, about 38% pixels are handled by the least-squares estimator, 17% by the bi-weight estimator and 45% by the MF-estimator.

For quantitatively assessing the improvement using the MAR estimation step, we analyze the matching quality in terms of the two error measurements, i.e., absolute disparity error, and absolute warped image error. The absolute disparity error is the absolute value of the difference between the resultant disparity map and the true disparity map. The absolute warped image error is the absolute value of the difference between the warped image constructed using the estimated disparity map and the test image. The difference from the true disparity is shown in Figure 1(f). The absolute disparity errors after applying the MAR estimation step show an improvement of around 20% over the initial disparity estimates from the ASA step.

It is seen that the ASA step can reproduce the general cloud topography. The MAR estimation step fur-

ther recovers many fine scale cloud structures and multiple cloud tracers. Statistically, the MAR algorithm improves the means of both the absolute disparity errors and the absolute warped errors by more than 20%.

References

[1] J.K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images – A review," *Proc. of IEEE*, Vol.76, no.8, pp.917-935, 1988.

[2] T. Darrell and A. P. Pentland, "Cooperative robust estimation using layers of support," *IEEE Trans. Pattern Anal. Mach. Intell.*

[3] L. L. Grewe and A. C. Kak, "Stereo Vision," *Handbook of Pattern Recognition and Image Processing: Computer Vision*, Chp. 8, pp. 239-317, 1994.

[4] A.F. Hasler, "Stereoscopic measurements," *Weather Satellites: Systems, Data and Environmental Applications, Section VII-3*, P.K. Rao, S.J. Holms, R.K. Anderson, J. Winston, and P. Lehr, Eds, Amer. Meteor.Soc., Boston MA, pp. 231-239, 1990.

[5] A.F. Hasler, J. Strong, R.H. Woodward and H. Pierce, "Automatic analysis of stereoscopic satellite image pairs for determination of cloud-top height and structure," *J. Applied Meteorology*, Vol. 30, March, 1991.

[6] Y. Huang, K. Palaniappan, X. Zhuang and J. Cavanaugh, "Optic flow field segmentation and motion estimation using a robust genetic partitioning algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, To appear, 1995.

[7] K. Palaniappan, C. Kambhamettu, A. F. Hasler and D. B. Goldgof, "Structure and semi-fluid motion analysis of stereoscopic satellite images for cloud tracking," *Proc. Int. Conf. Computer Vision*, 1995.

[8] H.K. Ramapriyan, J.P. Strong, Y. Hung and C.W. Murry, "Automated matching of pairs of SIR-B images for elevation mapping," *IEEE Trans. Geoscience Remote Sensing*, Vol.24, no.4, pp.462-472, 1986.

[9] C. V. Stewart, "A new robust operator for computer vision: Application to range data," *Proc. IEEE Conf. Computer Vision and Patt. Recog.*, pp. 167-173, 1994.

[10] X. Zhuang, Y. Zhao and T. S. Huang,"Residual-based robust estimation and image-motion analysis," *Int. J. Imaging Syst. Tech.*, Vol. 2, pp. 371-379, 1990.

[11] X. Zhuang, T. Wang and P. Zhang,"A highly robust estimator through partially likelihood function modeling and its application in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.14, No. 1, pp. 19-35, 1992.

[12] X. Zhuang and Y. Huang, "Robust 3D-3D pose estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.16, no.8, pp.818-824, 1994.

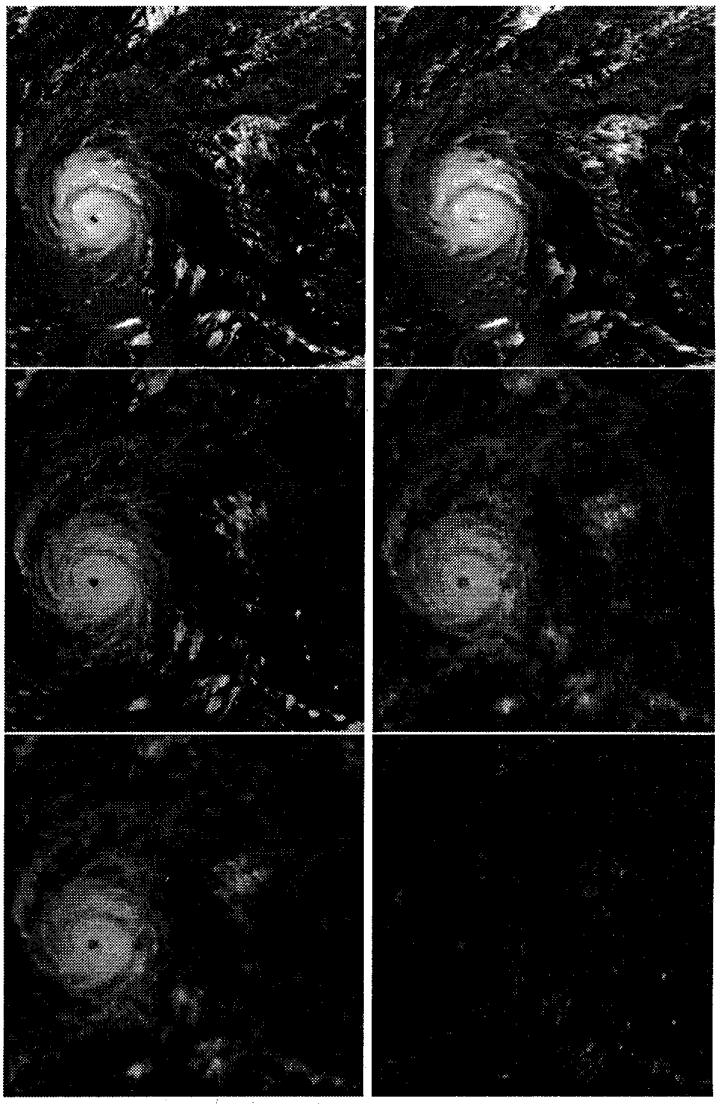

Figure 1: Hurricane Andrew stereo cloud data: (a) the original reference image, (b) the synthetic test image, (c) the true height (disparity) map, (d) computed cloud disparities using ASA algorithm, (e) disparities using MAR algorithm (f) difference from true disparities using MAR algorithm.