Z. Zhang and Y. Shan, "A Progressive Scheme for Stereo Matching", Springer LNCS 2018: 3D Structure from Images - SMILE 2000, M. Pollefeys et al. (eds.), pages 68-85, Springer-Verlag, 2001.

A Progressive Scheme for Stereo Matching

Zhengyou Zhang and Ying Shan

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
zhang@microsoft.com,
WWW home page: http://research.microsoft.com/˜zhang/

Abstract. Brute-force dense matching is usually not satisfactory because the same search range is used for the entire image, yielding potentially many false matches. In this paper, we propose a progressive scheme for stereo matching which uses two fundamental concepts: the disparity gradient limit principle and the least commitment strategy. The first states that the disparity should vary smoothly almost everywhere, and the disparity gradient should not exceed a certain limit. The second states that we should first select only the most reliable matches and therefore postpone unreliable decisions until enough confidence is accumulated. Our technique starts with a few reliable point matches obtained automatically via feature correspondence or through user input. New matches are progressively added during an iterative matching process. At each stage, the current reliable matches constrain the search range for their neighbors according to the disparity gradient limit, thereby reducing potential matching ambiguities of those neighbors. Only unambiguous matches are selected and added to the set of reliable matches in accordance with the least commitment strategy. In addition, a correlation match measure that allows rotation of the match template is used to provide a more robust estimate. The entire process is cast within a Bayesian inference framework. Experimental results illustrate the robustness of our proposed dense stereo matching approach.

Keywords: Stereo vision, Stereo matching, Disparity gradient limit, Least commitment, Progressive matching, Bayesian inference, Correlation, Image registration.

1 Introduction

Over the years numerous algorithms for image matching have been proposed. They can roughly be classified into two categories:

Feature matching. They first extract salient primitives from the images, such as corners and edge segments, and match them across two or more views. An image can then be described by a graph with primitives defining the nodes and geometric relations defining the links. Matching becomes finding the mapping of graphs: subgraph isomorphism. Some heuristics such as assuming affine transformation between images are usually introduced to reduce the complexity. These methods are fast because only a small subset of the image pixels are used, but may fail if the chosen primitives cannot be reliably detected in the images. They only produce a very coarse 3D model of the actual scene. The following list of references is by no means exhaustive: [9, 13, 15, 1, 5]

Template matching. They attempt to correlate image patches across views, assuming that they present some similarity [8, 10, 7, 18, 20]. The underlying assumption appears to be a valid one for relatively textured areas and for image pairs with small difference; however it may be wrong at occlusion boundaries and within featureless regions. Although these algorithms produce a dense 3D reconstruction of the actual scene, brute-force matching is usually not satisfying because of potentially many false matches.

All above stereo matching algorithms suffer from the difficulty in specifying an appropriate search range and the inability to adapt the search range depending on the observed scene structure.

In this paper, we propose a progressive scheme that, to some extent, combines these two approaches. It starts with a few reliable point matches obtained automatically via feature correspondence or through user input. It then tries to find progressively more pixel matches based on two fundamental concepts: disparity gradient limit principle and least commitment strategy. The disparity gradient limit principle states that the disparity should vary smoothly almost everywhere, and the disparity gradient should not exceed a certain value. This defines the search range for candidate matches. The least commitment strategy states that we should first select only the most reliable matches and therefore postpone an unreliable decision until enough confidence is accumulated. New matches are progressively added during an iterative matching process. At each stage, the current reliable matches constrain the search range for their neighbors according to the disparity gradient limit, thereby reducing potential matching ambiguities of those neighbors. Only unambiguous matches are selected and added to the set of reliable matches in accordance with the least commitment strategy.

Lhuillier and Quan recently reported a matching algorithm using a similar idea [11]. They also start with a few reliable point matches, but the technique to find more matches is very different from ours. They first choose the best match, and look for additional matches in their 5×5 neighborhood. Therefore, they only consider one match each time and propagate it in a very small area, while we consider all current matches simultaneously and do not restrict the propagation within a very small area. Chen and Medioni [3] uses a very similar strategy to that of Lhuillier and Quan, but work with a volumetric representation.

The paper is organized as follows. Section 2 presents the disparity gradient limit principle and the least commitment strategy, and introduces a scheme for progressive matching. Section 3 describes the implementation details on how disparities are predicted and estimated, which is formulated within a Bayesian inference framework. Section 4 proposes a new correlation technique designed for cameras in general position. Section 5 provides experimental results, including intermediate ones, with two sets of real data. Section 6 concludes the paper with a discussion on future work.

2 A Progressive Scheme

We first describe the two fundamental concepts, namely the disparity gradient limit principle and the least commitment strategy. We then present a simple progressive scheme which starts a few seed matches and then tries to find progressively more pixel matches based on these two concepts.

2.1 Disparity Gradient Limit Principle

Disparity is directly related to depth. Disparity changes coincide with depth changes. The disparity gradient limit principle states that the disparity should vary smoothly almost everywhere, and the disparity gradient should not exceed a certain value. Psychophysical studies have provided evidence that in order for the human visual system to binocularly fuse two dots of a simple stereogram, the disparity gradient (ratio of the disparity difference) between the dots to their cyclopean separation must not exceed a limit of 1 [2, 16]. Objects in the world are usually bounded by continuous opaque surfaces, and disparity gradient can be considered as a simple measure of continuity. The disparity gradient limit principle provides a constraint on scene jaggedness embracing simultaneously the ideas of opacity, scene continuity, and continuity between views [14]. It has been used in several successful stereo matching algorithms including the PMF algorithm [15] to resolve matching ambiguity.

The disparity gradient limit principle is used differently in our work, as we will explain in details in Section 3.1. It is exploited to estimate the uncertainty of the predicted disparity for a particular pixel, and the uncertainty is then used to define the search ranges for candidate matches.

2.2 Least Commitment Strategy

The least commitment strategy states that we should first select only the most reliable decisions and therefore postpone an unreliable decision until enough confidence is accumulated. It is a powerful strategy used in Artificial Intelligence, especially in action planning [22, 19]. Since no irreversible decision is made (i.e. all decisions made are reliable), this principle offers significant flexibility in avoiding locking search into a possibly incorrect step where an expensive refinement such as backtracking has to be exploited.

The least commitment strategy is explored in our algorithm in four ways (abbreviated as STAB):

Search range. Matching criterion such as correlation is local and heuristic. If the match of a pixel has to be searched in a wide range, there is a high probability that the found match is not a correct one. It is preferable to defer matching of these pixels as late as possible because the search range may be reduced later after more reliable matches are established.

Texture. A pixel is more discriminating in a highly textured neighborhood than others. It is difficult to distinguish pixels in the same neighborhood having similar intensity. Therefore, we can expect to have more reliable matches for pixels in areas with strong textures, and thus try to match them first.

Ambiguity. We may find several candidate matches for a pixel. Rather than using expensive techniques such as dynamic programming to resolve the ambiguity, we simply defer the decision. Once more reliable matches are found in the future, the ambiguity will become lower because of a better disparity estimate with smaller uncertainty.

Bookkeeping. If a pixel does not have any candidate match, it is probably occluded by others or is not in the field of view of the other camera, then we do not need to search for its match in the future. Similar, if a pixel has already found a match, further search is not necessary. We bookkeep both types of pixels for efficiency.

2.3 A Progressive Stereo Matching Algorithm

We can now outline the proposed progressive algorithm. Details will be given in the following sections.

A pixel in the first image has three labels: `MATCHED` (already matched), `NOMATCH` (no candidate matches found), and `UNKNOWN` (not yet decided). All pixels are initially labeled as `UNKNOWN`.

For a pixel which is labeled `UNKNOWN`, we compute a list of candidate pixels in the second image which satisfy the epipolar constraint and disparity gradient limit constraint. We use the normalized cross correlation as our matching criterion. For a pair of pixels between two images, we compute the normalized cross correlation score between two small windows, called *correlation windows*, centered at the pixels. The correlation score ranges from -1, for two correlation windows which are not similar at all, to $+1$, for two correlation windows which are identical. The pair of pixels are considered as a potential match if the correlation score is larger than a predefined threshold T_C. The list of candidate pixels are ordered on the epipolar line, and the correlation scores form a curve. If there is only one peak on the correlation curve exceeding the threshold T_C, then the pixel at the peak is considered as the match of the given pixel in the first image, and the given pixel is labeled as `MATCHED`. If there is no peak exceeding the threshold T_C, we label the given pixel as `NOMATCH`, as we mentioned earlier. If there are two or more peaks exceeding T_C, the matching is ambiguous, and according to the least commitment principle, we simply leave it as is. We iterate this procedure until no more matches can be found or the maximum number of iteration is attained.

As we described earlier, pixels in highly textured areas are considered first. Textureness is measured as the sample deviation of the intensity within a correlation window. In order for a pixel in the first image to be considered, its sample deviation must be larger than a threshold T_{σ_I}. The threshold T_{σ_I} evolves with iteration. It is given by a monotonic function *ThresholdSigmaIntensity* which never increases with iteration.

Similarly, if a given pixel in the first image has a large uncertainty of its disparity vector, this pixel should be considered as late as possible. In order for a pixel to be considered, the standard deviation of its predicted disparity vector must be smaller than a threshold T_{σ_D}. The threshold T_{σ_D} evolves with iteration. It is given by a monotonic function *ThresholdSigmaDisparity* which never decreases with iteration. That is, we, at the beginning, only considered pixels that have a good prediction of the disparity vector.

Please note that the above description is outlined only to present the essential ideas. The actual implementation of several components such as correlation computation is different, as we will describe in the next section.

The pseudo C++ code of the algorithm is summarized in Figure 1.

The above algorithm has a number of important properties:

Progressiveness. Because of bookkeeping, the number of pixels examined in each iteration becomes smaller. Also, as we will show later, the search range for a pixel

```
iteration = 0;
while (the maximum number of iterations is not reached)
    and (more matches are found) {
    $T_{\sigma_I}$ = ThresholdSigmaIntensity(iteration);
    $T_{\sigma_D}$ = ThresholdSigmaDisparity(iteration);
    for (every pixel labeled UNKNOWN in the first image) {
        estimate the disparity vector and its uncertainty;
        if ($\sigma_I$_of_the_pixel < $T_{\sigma_I}$)
            continue;  // not enough textured
        if ($\sigma_D$_of_the_pixel < $T_{\sigma_D}$)
            continue;  // too much uncertainty for its match
        compute the list of candidate pixels in the second image;
        compute the correlation score $C$ for each candidate pixel;
        if (there is one peak on the correlation curve)
            and (its $C$ > $T_C$) {
            update its disparity vector;
            label the pixel as MATCHED.
        }
        else if (there is no candidate whose $C$ > $T_C$) {
            label the pixel as NOMATCH.
        }
    }
}
```

Fig. 1. Pseudo C++ code of the progressive stereo matching algorithm.

is reduced when we update the disparity with more matched pixels. This property guarantees that the iterative procedure is actually making some progress and that the search space is being reduced.

Monotonicity. Because of the monotonicity of functions *ThresholdSigmaIntensity* and *ThresholdSigmaDisparity*, threshold T_{σ_I} is getting smaller and threshold T_{σ_D} is getting larger with the progress of the algorithm. This means that the probability that a pixel labeled as `UNKNOWN` is selected for matching test becomes higher, eventually resulting more `MATCHED`/`NOMATCH` pixels. Together with the update of disparity vectors and their uncertainty, this property guarantees that the set of `UNKNOWN` pixels considered is truly different from that prior to refinement, “different” in the sense of the actual pixels considered and also of their candidate pixels to match in the other image.

Completeness. This property says that adding more `MATCHED`/`NOMATCH` pixels will not lose any potential matches. This is desirable because it means that an expensive refinement such as backtracking is never performed. The above proposed algorithm clearly satisfies this property because of the least commitment strategy, provided that the disparity gradient limit constraint is satisfied over the entire observed scene.

The completeness property of our algorithm does not imply that as the final result each pixel must be labeled either `MATCHED` or `NOMATCH`. Indeed, pixels within a uniform color region may still be labeled as `UNKNOWN`. However, from the neighboring matched pixels, these pixels have an estimate of their disparity vectors that can be used if necessary, for example, for image-based rendering.

3 Implementation Details

In this section, we provide the details in implementing the progressive algorithm described in the last section. Basically, for each pixel labeled `UNKNOWN`, we need to do two things: prediction the disparity and its uncertainty, based on the information provided by the neighboring matched pixels; estimation of its disparity based on the information contained in the images.

If we formulate the problem in terms of Bayesian inference (see e.g. [21]), the first corresponds to the prior density distribution of the disparity, $p(d|\mathbf{m}, B)$, where d is the disparity of the given pixel $\mathbf{m}$, and B denote the relevant background information at hand such as the epipolar geometry and the set of already matched pixels. The second corresponds to the sampling distribution $p(I'|d, \mathbf{m}, B)$, or the likelihood of the observed data (i.e., the second image I') given d, $\mathbf{m}$ and B. Bayes' rule can then be used to combine the information in the data with the prior probability, which yields the posterior density distribution

$$p(d|I', \mathbf{m}, B) = \frac{p(I'|d, \mathbf{m}, B)p(d|\mathbf{m}, B)}{p(I'|\mathbf{m}, B)} \ , \tag{1}$$

where $p(I'|\mathbf{m}, B)$ does not depend on d and can be considered as a constant because the second image I' is fixed. We can thus omit the factor $p(I'|\mathbf{m}, B)$ and work on the unnormalized posterior density distribution $p(I'|d, \mathbf{m}, B)p(d|\mathbf{m}, B)$, still denoted by $p(d|I', \mathbf{m}, B)$ to abuse the notation. Appropriate computations to summarize $p(d|I', \mathbf{m}, B)$ are finally performed in order to decide whether the pixel under consideration should be labeled `MATCHED` or `NOMATCH`, or kept as `UNKNOWN` for future decision.

3.1 Prediction of the Disparity and its Uncertainty

Before introducing our work, it is helpful to define disparity and disparity gradient and summarize the related results obtained by others.

Disparity is well defined for parallel cameras (i.e., the two image planes are the same) [6]. Without loss of generality, the horizontal axis is assumed to be aligned in both images. Given a pixel of coordinates (u, v) in the first image and its corresponding pixel of coordinates (u', v') in the second image, disparity is defined as the difference $d = v' - v$. Disparity is inversely proportional to the distance of the 3D point to the cameras. A disparity of 0 implies that the 3D point is at infinity.

Consider now two 3D points whose projections are $\mathbf{m}_1 = [u_1, v_1]^T$ and $\mathbf{m}_2 = [u_2, v_2]^T$ in the first image, and $\mathbf{m}'_1 = [u'_1, v'_1]^T$ and $\mathbf{m}'_2 = [u'_2, v'_2]^T$ in the second image ($u'_1 = u_1$ and $u'_2 = u_2$ in the parallel cameras case). Their disparity gradient is defined to be the ratio of their difference in disparity to their distance in the cyclopean image.[1] In the first image, the disparity gradient is given by

$$DG = \left| \frac{d_2 - d_1}{v_2 - v_1 + (d_2 - d_1)/2} \right| \ . \tag{2}$$

Experiments in psychophysics have provided evidence that human perception imposes the constraint that the disparity gradient DG is upper-bounded by a limit K. That is, if a point on an object is perceived, neighboring points having $DG > K$ are simply not perceived correctly. The limit $K = 1$ was reported in [2]. The theoretical limit for opaque surfaces is $K = 2$ to ensure that the surfaces are visible to both eyes [14]. Although the range of allowable surfaces is large with $K = 2$, disambiguating power is weak because false matches receive and exchange as much support as correct ones. Another extreme limit is $K \approx 0$, which allows only nearly front-parallel surfaces, and this has been used locally in the stereogram matching algorithm described in [12]. In the PMF algorithm, the disparity gradient limit K is a free parameter, which can be varied over range $(0, 2)$. An intermediate value, e.g., between 0.5 and 1, allow selection of a convenient trade-off point between allowable scene surface jaggedness and disambiguating power because it turns out that most false matches produce relatively high disparity gradients [14]. Again, as reported in [14], less than 10% of world surfaces viewed at more than 26cm with 6.5cm of eye separation will present with disparity gradient larger than 0.5. This justifies use of a disparity gradient limit well below the theoretical value (of 2) without imposing strong restrictions on the world surfaces that can be fused by the stereo algorithm.

When the cameras are in general position, it is not reasonable to hope to define a scalar disparity as a simple function of the image coordinates of two pixels in correspondence [6]. In this work, we simply use a vector $\mathbf{d} = [u' - u, v' - v]^T$, called the disparity vector. This is the same as the flow vector used in optical flow computation. If a scalar value is necessary, we use $d = \|\mathbf{d}\|$ and call it the disparity. If we look at objects that are smooth almost everywhere, both $\mathbf{d}$ and d should vary smoothly. Similar to (2), for two points $\mathbf{m}_1$ and $\mathbf{m}_2$ in the first image, we define the disparity gradient as

$$DG = \frac{\|\mathbf{d}_2 - \mathbf{d}_1\|}{\|\mathbf{m}_2 - \mathbf{m}_1 + (\mathbf{d}_2 - \mathbf{d}_1)/2\|} \ . \tag{3}$$

Imposing the gradient limit constraint $DG \leq K$, we have

$$\|\mathbf{d}_2 - \mathbf{d}_1\| \leq K \|\mathbf{m}_2 - \mathbf{m}_1 + (\mathbf{d}_2 - \mathbf{d}_1)/2\| \ .$$

Using inequality $\|\mathbf{v}_1 + \mathbf{v}_2\| \leq \|\mathbf{v}_1\| + \|\mathbf{v}_2\|$ for any vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we obtain

$$\|\mathbf{d}_2 - \mathbf{d}_1\| \leq K \|\mathbf{m}_2 - \mathbf{m}_1\| + K \|(\mathbf{d}_2 - \mathbf{d}_1)/2\|$$

which leads immediately, for $K < 2$, to

$$\|\mathbf{d}_2 - \mathbf{d}_1\| \leq \frac{2K}{2 - K} D \ , \tag{4}$$

where $D = \|\mathbf{m}_2 - \mathbf{m}_1\|$ is the distance between $\mathbf{m}_1$ and $\mathbf{m}_2$. We immediately have the following result:

[1] For a pair of pixels in correspondence with coordinates (u, v) and (u', v'), the cyclopean image point is at $((u + u')/2, (v + v')/2)$

Lemma 1. *Given a pair of matched points $(\mathbf{m}_1, \mathbf{m}'_1)$ and a point $\mathbf{m}_2$ in the neighborhood of $\mathbf{m}_1$, the corresponding point $\mathbf{m}'_2$ that satisfies the disparity gradient constraint with limit K must be inside a disk centered at $\mathbf{m}_2 + \mathbf{d}_1$ with radius equal to $\frac{2K}{2-K}D$, which we call the* continuity disk.

In other words, in absence of other knowledge, the best prediction of the disparity of $\mathbf{m}_2$ is equal to $\mathbf{d}_1$ with the continuity disk defining its uncertainty.

We may want to favorite the actual disparity to be at the central part of the continuity disk. We may also want to consider a small probability that the actual disparity is outside of the continuity disk, due to occlusion or surface discontinuity. We therefore model the uncertainty as an isotropic Gaussian distribution with standard deviation equal to half of the radius of the continuity disk. More precisely, given a pair of matched points $(\mathbf{m}_i, \mathbf{m}'_i)$, the disparity of a point $\mathbf{m}$ is modeled as

$$\mathbf{d} = \mathbf{d}_i + D_i \mathbf{n}_i \ , \tag{5}$$

where $\mathbf{d}_i = \mathbf{m}'_i - \mathbf{m}_i$, $D_i = \|\mathbf{m} - \mathbf{m}_i\|$, and $\mathbf{n}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I})$ with $\sigma_i = K/(2-K)$. Note that disparity $\mathbf{d}_i$ also has its own uncertainty due to limited image resolution. The density distribution of $\mathbf{d}_i$ is also modeled in our work as a Gaussian, i.e., $p(\mathbf{d}_i) = N(\mathbf{d}_i | \bar{\mathbf{d}}_i, \sigma_{d_i}^2 \mathbf{I})$. It follows that the density distribution of disparity $\mathbf{d}$ is given by

$$p(\mathbf{d} | (\mathbf{m}_i, \mathbf{m}'_i), \mathbf{m}) = N(\mathbf{d} | \bar{\mathbf{d}}_i, (\sigma_{d_i}^2 + D_i^2 \sigma_i^2) \mathbf{I}) \ . \tag{6}$$

If we are given a set of point matches $\{(\mathbf{m}_i, \mathbf{m}'_i) | i = 1, \ldots, n\}$, we then have n independent predictions of disparity $\mathbf{d}$ as given by (6). The prior density distribution of the disparity, $p(\mathbf{d}|\mathbf{m}, B)$, can be obtained by combining these predictions with the minimum variance estimator, i.e.,

$$p(\mathbf{d}|\mathbf{m}, B) = N(\mathbf{d}|\bar{\mathbf{d}}, \sigma^2 \mathbf{I}) \ , \tag{7}$$

where

$$\bar{\mathbf{d}} = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_{d_i}^2 + D_i^2 \sigma_i^2} \Big)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_{d_i}^2 + D_i^2 \sigma_i^2} \bar{\mathbf{d}}_i$$
$$\sigma^2 = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_{d_i}^2 + D_i^2 \sigma_i^2} \Big)^{-1} \ .$$

A more robust version is first to identify the Gaussian with smallest variance, and then to combine it with those Gaussians whose means fall within two or three standard deviations.

It remains the problem of choosing σ_i, which as mentioned earlier is related to the disparity gradient limit K. In the PMF algorithm, K is set to a value between 0.5 and 1, which is equivalent to a value between 1/3 and 1/2 for our σ_i. Considering that the disparity gradient constraint is still a local one, it should become less restrictive when the point being considered is away from a matched point. Hence, we specify a range $[\sigma_{\min}, \sigma_{\max}]$, and σ_i is given by

$$\sigma_i = (\sigma_{\max} - \sigma_{\min})(1 - \exp(-D_i^2/\tau^2)) + \sigma_{\min} \ . \tag{8}$$

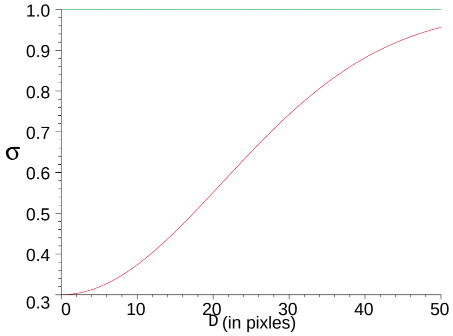

Fig. 2. Function of σ_i (related to the disparity gradient limit) w.r.t. the distance to a matched pixel. See (8).

When $D_i = 0, \sigma_i = \sigma_{\min}$; when $D_i = \infty, \sigma_i = \sigma_{\max}$. The parameter τ controls how fast the transition from $\sigma_{\min}$ to $\sigma_{\max}$ is expected. In our implementation, $\sigma_{\min} = 0.3$ pixels, $\sigma_{\max} = 1.0$ pixel, and $\tau = 30$. This is equivalent to $K_{\min} = 0.52$ and $K_{\max} = 1.34$. Figure 2 displays how σ_i varies with respect to the distance D_i. From many images we have tried, this strategy works well.

3.2 Computation of the Disparity Likelihood

We now proceed to compute the sampling distribution $p(I'|d, \mathbf{m}, B)$, or the likelihood of the observed data (i.e., the second image I') given d, $\mathbf{m}$ and B.

Because of the epipolar constraint, we do not need to compute the density for each pixel in I'. Furthermore, we do not even need to compute the density for each pixel on the epipolar line of $\mathbf{m}$ because of the prior density computed in (7). The list of pixels of interest, called the *candidate pixels* and denoted by $Q(\mathbf{m})$, is the intersection of the epipolar line of $\mathbf{m}$ with the continuity disk defined in Lemma 1.

The densities are related to the correlation scores C_j between $\mathbf{m}$ in the first image and each candidate pixel $\mathbf{m}'_j \in Q(\mathbf{m})$ in the second image. Instead of using the standard correlation technique based on two rectangular windows, we have developed a new one which is well adapted for two images in general position. We defer its presentation to Section 4. For the moment, it suffices to say that the correlation score C is between -1 (when they are not similar at all) and $+1$ (when they are identical). Finally, correlation scores are mapped to densities by adding 1 followed by a normalization. More precisely, the correlation score C_j of a pixel $\mathbf{m}'_j$ is converted into a density as

$$p(I'(\mathbf{m}'_j)|\mathbf{d}^{(j)}, \mathbf{m}, B) = \frac{C_j + 1}{\sum_{k \in Q(\mathbf{m})}(C_k + 1)} \; , \tag{9}$$

where $\mathbf{d}^{(j)} = \mathbf{m}'_j - \mathbf{m}$.

3.3 Inference From the Posterior Density

The posterior density distribution $p(d|I', \mathbf{m}, B)$ is simply multiplication of $p(I'(\mathbf{m}'_j)|\mathbf{d}^{(j)}, \mathbf{m}, B)$ in (9) with $p(\mathbf{d}^{(j)}|\mathbf{m}, B)$ in (7) for each candidate pixel $\mathbf{m}'_j$.

Based on $p(d|I', \mathbf{m}, B)$, we can do a number of things. If there is only one prominent peak, the probability that this is a correct match is very high, and we thus make the decision and label the pixel in the first image `MATCHED`. If there are two or more prominent peak, the matching ambiguity is high, i.e., the probability of making a wrong decision is high. Following the least commitment principle, we leave this pixel to evolve. If there is no prominent peak at all, the probability that the corresponding point in the second image is not visible is very high (either occluded by others or out of the field of view), and we label the pixel in the first image `NOMATCH`.

In order to facilitate the task of choosing an appropriate threshold on the posterior density distribution, and since anyway we are working with the *unnormalized* posterior density distribution, we normalize the prior and likelihood functions differently. The prior in (7) is multiplied by $\sigma\sqrt{2\pi}$ so that the maximum is equal to one. The likelihood in (9) is changed to $(C_j + 1)/2$ so that it is equal to 1 for identical pixels and 0 for completely different pixels. A peak in the posterior density distribution is considered as a prominent one if its value is larger than 0.3, which corresponds to, e.g., the situation where $C_j = 0.866$ and the disparity is at 1.5σ.

4 A New Correlation Technique

The correlation technique described in this section is designed for stereo cameras in general position.

Consider a pair of points $\mathbf{m}$ and $\mathbf{m}'$ as shown in Fig. 3, where the corresponding epipolar lines l and l' are also drawn. We can easily compute a Euclidean transformation

$$\mathbf{m}'_i = \mathbf{R}(\theta)(\mathbf{m}_i - \mathbf{m}) + \mathbf{m} \ , \tag{10}$$

where $\mathbf{R}(\theta)$ is a 2D rotation matrix with rotation angle equal to θ, the angle between the two epipolar lines. It sends $\mathbf{m}$ to $\mathbf{m}'$ and a point on l to a point on l'.

Choose a rectangular window centered at $\mathbf{m}$ with one side parallel to the epipolar line. A point $\mathbf{m}_i$ corresponds to a point $\mathbf{m}'_i$ given by (10). Point $\mathbf{m}'_i$ is usually not on the pixel grid, and its intensity is computed through bilinear interpolation from its four neighboring pixels. Correlation score is then computed between points $\mathbf{m}_i$ in the correlation window and points $\mathbf{m}'_i$ according to (10). We use the normalized cross correlation [6] which is equal to 1 for two identical sets of pixels and -1 for two completely different sets.

If two epipolar lines are both horizontal or vertical, the new technique will be equivalent to the standard one.

An even more elaborate way to compute the correlation is to weight differently each point: Pixels in the central part have more weights than those near the border. In our

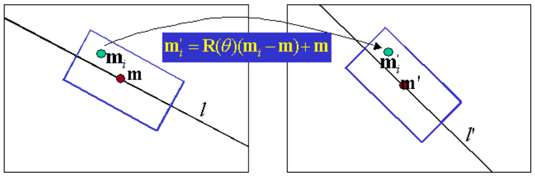
$\mathbf{m}_i$
$\mathbf{m}$
l
$\mathbf{m}'_i = \mathbf{R}(\theta)(\mathbf{m}_i - \mathbf{m}) + \mathbf{m}$
$\mathbf{m}'_i$
$\mathbf{m}'$
l'

Fig. 3. The new correlation technique for stereo cameras in general position.

Table 1. Number of matched pixels in each iteration

iteration	0	1	2	3	4	5	6
T_{σ_I}		7	6	5	4	3	2
T_{σ_D}		12	14	16	18	20	20
Books	141	455	712	939	1239	1440	1500
NMars	153	421	1249	2036	2360	2651	2741

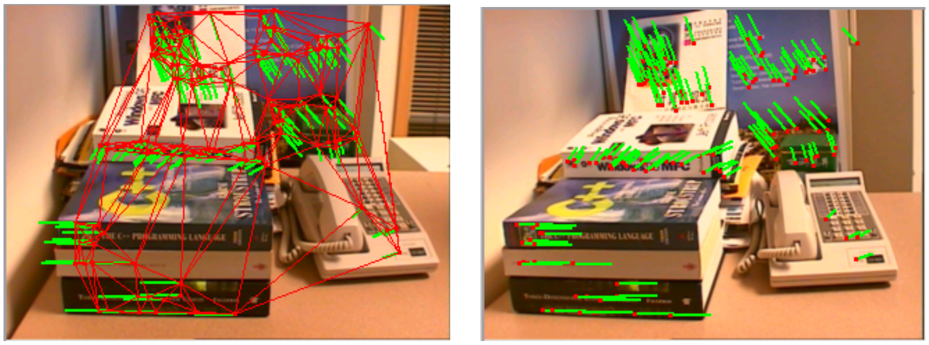

Fig. 4. Scene `Books`: Initial point matches indicated by the disparity vectors together with the Delaunay triangulation in the first image.

implementation, the size of correlation window is 11 pixels along the epipolar line and 9 pixels in the other direction. The pixels are weighted by a 2D Gaussian with standard deviation equal to 11 pixels along the epipolar line and 9 pixels in the other direction.

5 Experimental Results

We have conducted experiments with several sets of real data, and very promising results have been obtained. In this section, we report two of them: one is an office scene with books, called Scene `Books` (see Fig. 4); another is a scene with rocks from INRIA, call Scene `NMars` (see Fig. 9). Although the images in Scene `Books` are color, only black/white information is used. The image resolution is 740×480 for Scene `Books`, and 512×512 for Scene `NMars`.

To reduce computation cost, instead of using all previously found matches in predicting disparities and their uncertainties, we only use three neighboring points defined by the Delaunay triangulation [17]. The dynamic Delaunay triangulation algorithm described in [4] is used because of its efficiency in updating the triangulation when more point matches are available. It is reasonable to use only three neighboring points because other points are usually much farther away, resulting in a larger uncertainty in its predicted disparity, hence contributing little to the combined prediction of the disparity given in (7).

The initial set of point matches, together with the fundamental matrix, were obtained automatically using the robust image matching technique described in [23]. All param-

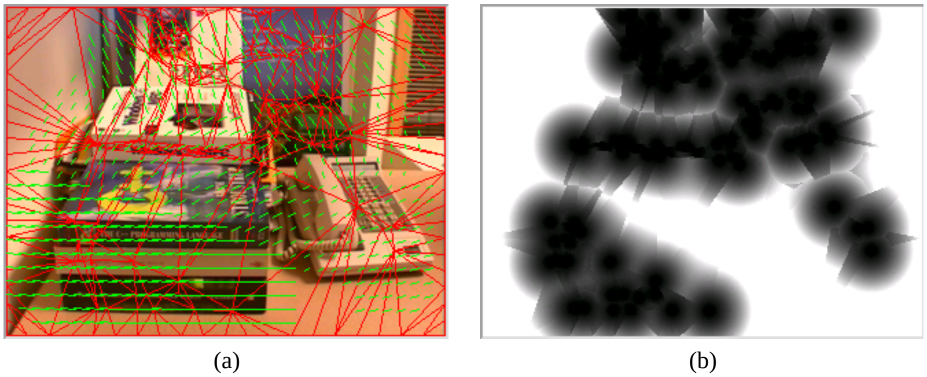

Fig. 5. Scene `Books`: Results with the initial point matches. (a) Delaunay triangulation and the predicted disparity vectors; (b) Predicted deviation of the disparity vectors.

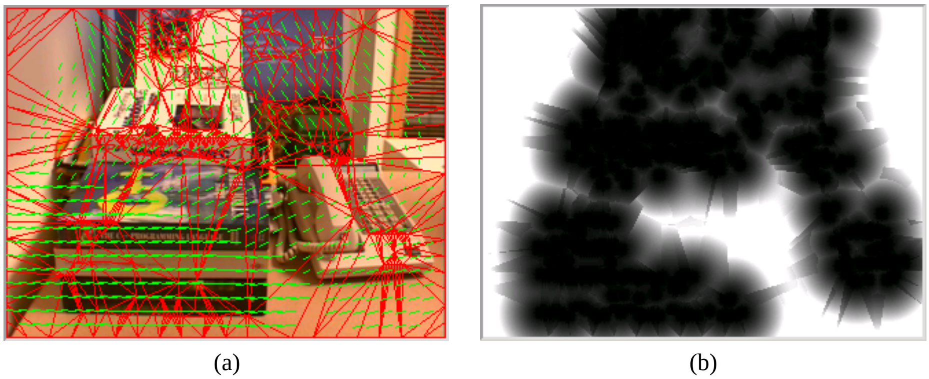

Fig. 6. Scene `Books`: Results after the second iteration. (a) Delaunay triangulation and the predicted disparity vectors; (b) Predicted deviation of the disparity vectors.

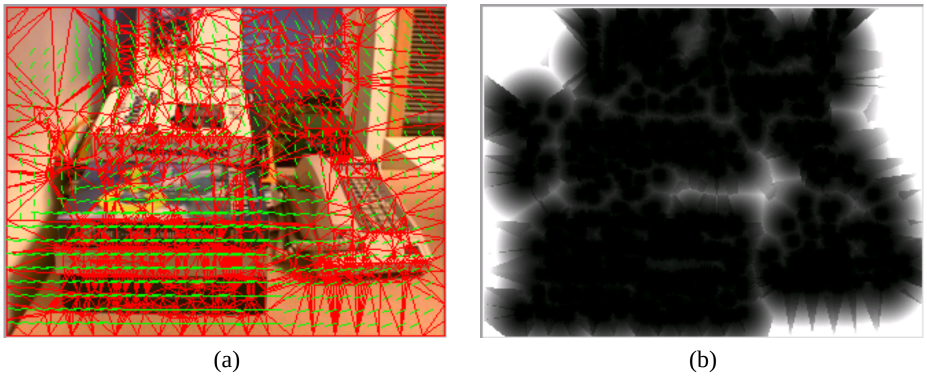

Fig. 7. Scene `Books`: Results after the sixth iteration. (a) Delaunay triangulation and the predicted disparity vectors; (b) Predicted deviation of the disparity vectors.

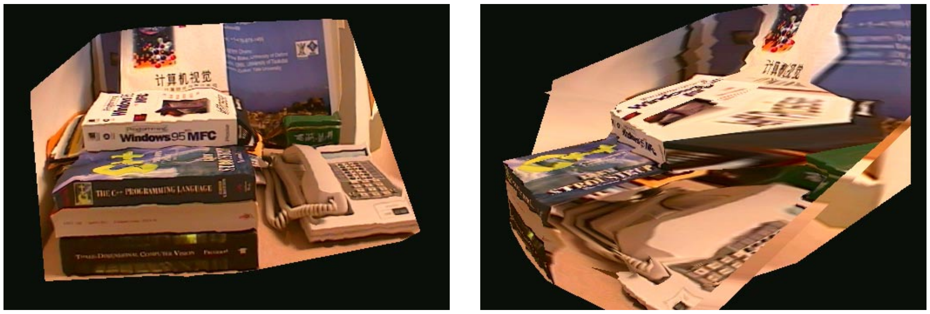

Fig. 8. Scene `Books`: Views of the 3D reconstruction with texture mapped from the 1st image.

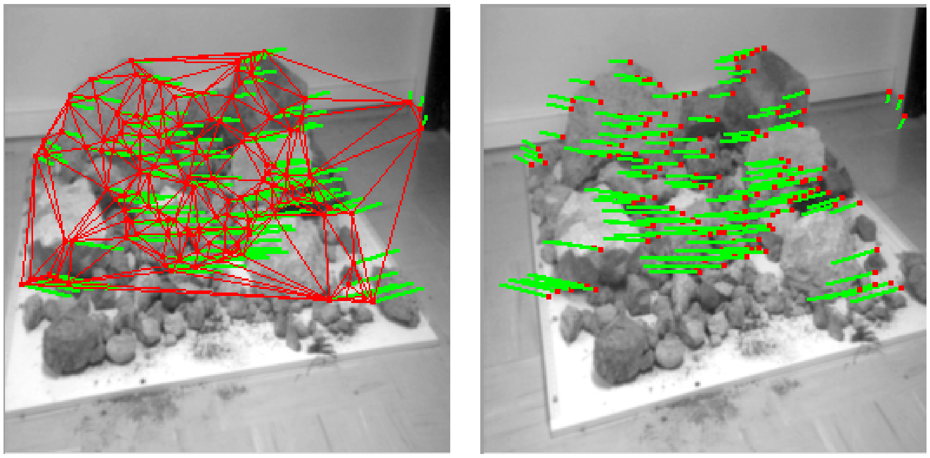

Fig. 9. Scene `NMars`: Initial point matches indicated by the disparity vectors together with the Delaunay triangulation in the first image.

eters are the same for both data sets. The search range was $[-60, 60]$ (pixels) for both horizontal and vertical directions.

All parameters in our algorithm are the same for both data sets. In particular, the values of functions *ThresholdSigmaIntensity* and *ThresholdSigmaDisparity* with respect to the iteration number are given in the second and third rows of Table 1. For example, for iteration 4, $T_{\sigma_I} = 4$ and $T_{\sigma_I} = 18$. In Table 1, we also provide the number of matches after each iteration. The number of matches for iteration 0 indicates the number of initial matches found by the robust matching algorithm. Note that instead of working on each pixel, we actually consider only one every four pixels because of the memory limitation in our Delaunay triangulation algorithm.

The initial set of point matches for Scene `Books` is shown in Fig. 4. Based on these, the disparity and its uncertainty were predicted, which are shown in Fig. 5. On the left, the disparity vectors are displayed for every 10 pixels and their lengths are half of their actual magnitudes. On the right, the standard deviation of the predicted disparities

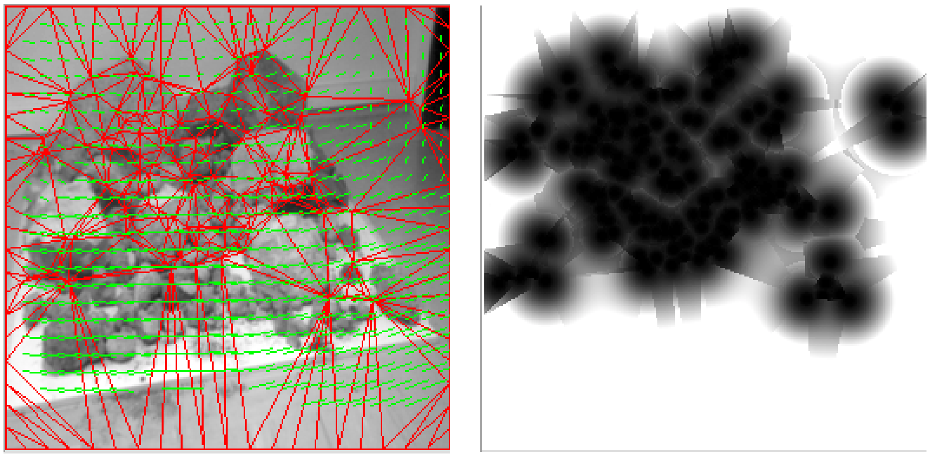

Fig. 10. Scene `NMars`: Results with the initial point matches. (left) Delaunay triangulation and the predicted disparity vectors; (right) Predicted deviation of the disparity vectors.

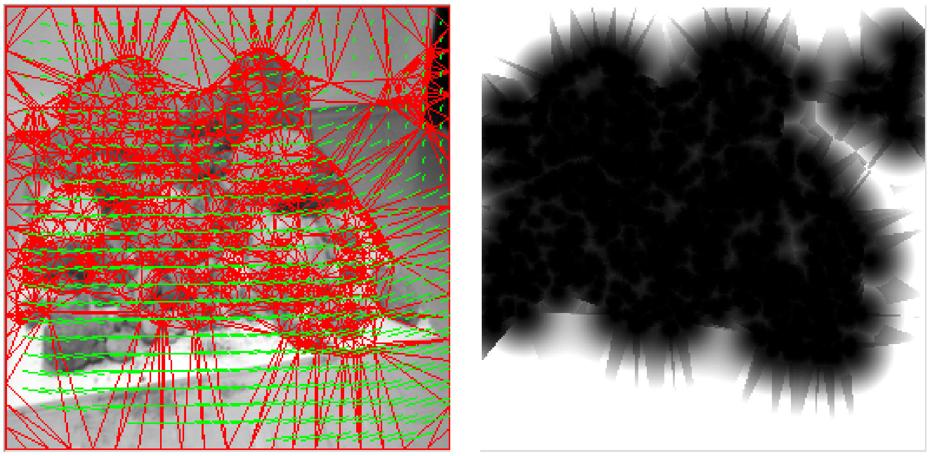

Fig. 11. Scene `NMars`: Results after the third iteration.

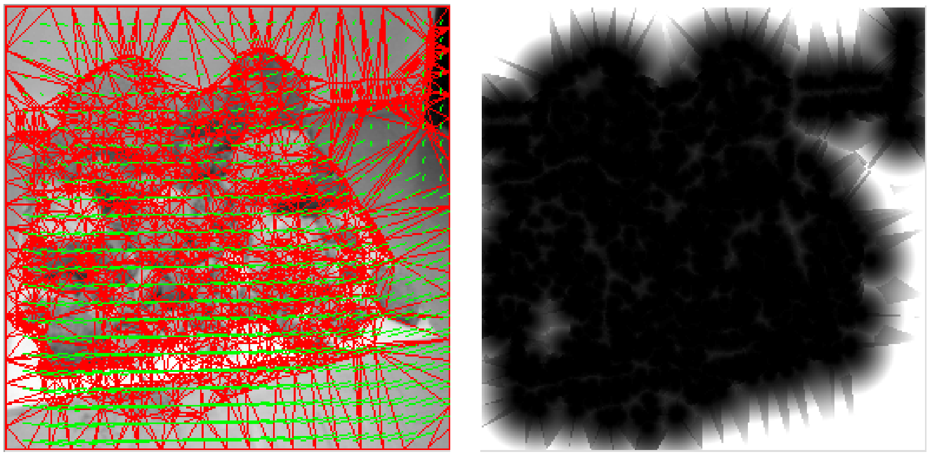

Fig. 12. Scene `NMars`: Results after the sixth iteration.

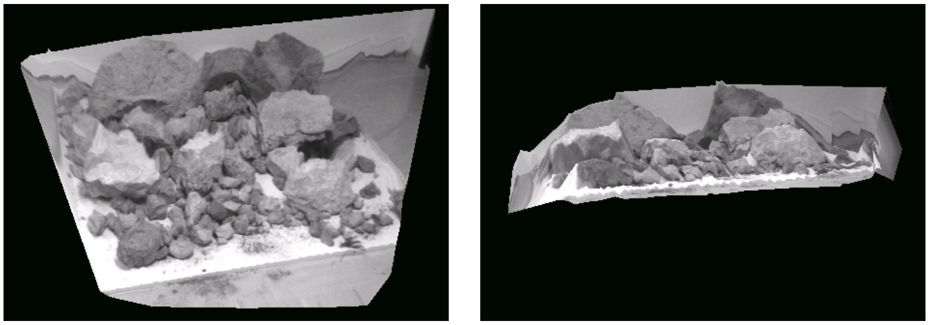

Fig. 13. Scene NMars: Views of the 3D reconstruction with texture mapped from the 1st image.

is shown in gray levels after having multiplied by 5 and truncated at 255. Therefore, “black” pixels in that image mean that the predicted disparities are quite reliable, while “white” pixels implies that the predicted disparities are very uncertain. The intermediate results after iteration 2, and 6 are shown in Fig. 6, and Fig. 7. We can observe clearly the fast evolution of the matching result. The uncertainty image becomes darker quickly. As we know the intrinsic parameters of the camera with which the images were taken, 3D Euclidean reconstruction can be obtained, two views of which are shown in Fig. 8. We can see that the book structure has been precisely recovered.

Similar results have been obtained with Scene NMars. As can be observed from Fig. 9, the lower part of the scene cannot be matched because the disparity is larger than the prefixed range (plus/minus a quarter of the image width). The predicted disparity vectors and their uncertainty computed from the initial set of matches are shown in Fig. 10, while those after iteration 6 are shown Fig. 12. It is clear that our progressive stereo algorithm is capable of finding matches with large disparity, the lower part of the scene in our case, even if the initial search range is large enough. 3D Euclidean reconstruction was also computed, two views of which are shown in Fig. 13.

6 Conclusions

In this paper, we have proposed a progressive scheme for stereo matching. It starts with a few reliable point matches obtained either manually from user input or automatically with feature-based stereo matching. It then tries to find progressively more pixel matches based on two fundamental concepts: disparity gradient limit principle and least commitment strategy. Experimental results have proven the robustness of our proposed dense stereo matching approach.

We have also cast the disparity estimation in the framework of Bayesian inference, and have developed a new correlation technique well adapted for cameras in general position.

There are a number of ways to extend the current algorithm. For example, the current implementation only estimate disparities with pixel precision. One of our future work consists in produce disparities with subpixel precision. We will also investigate in an even more efficient implementation.

References

1. N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *The International Journal of Computer Vision*, 1(2), April 1987.
2. P. Burt and B. Julesz. A gradient limit for binocular fusion. *Science*, 208:615–617, 1980.
3. Q. Chen and G. Medioni. A volumetric stereo matching method: Application to image-based modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 29–34, Colorado, June 1999. IEEE Computer Society.
4. O. Devillers, S. Meiser, and M. Teillaud. Fully dynamic Delaunay triangulation in logarithmic expected time per operation. *Comput. Geom. Theory Appl.*, 2(2):55–80, 1992.
5. Umesh R. Dhond and J.K. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.
6. Olivier Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
7. Pascal Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1):35–49, Winter 1993. Available as INRIA research report 1369.
8. A. Goshtasby, S. H. Gage, and J. F. Bartholic. A two-stage cross correlation approach to template matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(3):374–378, May 1984.
9. W.E.L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, 1985.
10. M.J. Hannah. A system for digital stereo image matching. *Photogrammetric Engeneering and Remote Sensing*, 55(12):1765–1770, December 1989.
11. M. Lhuillier and L. Quan. Image interpolation by joint view triangulation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 139–145, Colorado, June 1999. IEEE Computer Society.
12. D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.
13. Gérard Medioni and Ram Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31:2–18, 1985.
14. S. Pollard, J. Porrill, J. Mayhew, and J. Frisby. Disparity gradient, lipschitz continuity, and computing binocular correspondance. In O.D. Faugeras and G. Giralt, editors, *Robotics Research: The Third International Symposium*, pages 19–26. MIT Press, 1986.
15. S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby. PMF : a stereo correspondence algorithm using a disparity gradient constraint. *Perception*, 14:449–470, 1985.
16. K Prazdny. On the disparity gradient limit for binocular fusion. *Perception and Psychophysics*, 37(1):81–83, 1985.
17. F. Preparata and M. Shamos. *Computational Geometry*. Springer-Verlag, New-York, 1985.
18. L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In Bernard Buxton, editor, *Proceedings of the 4th European Conference on Computer Vision*, Cambridge, UK, April 1996.
19. S. Russel and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, New Jersey, 1995.
20. D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. *The International Journal of Computer Vision*, 28(2):155–174, 1998.
21. D.S. Sivia. *Data Analysis: a Bayesian tutorial*. Oxford University Press, 1996.
22. D. Weld. An introduction to least commitment planning. *AI Magazine*, 15(4):27–61, 1994.
23. Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, October 1995.