DISPARITY ESTIMATION AND RECONSTRUCTION IN STEREO VISION

Salvador Gutiérrez and José Luis Marroquín

Comunicación Técnica No I-03-07/7-04-2003
(CC/CIMAT)

DISPARITY ESTIMATION AND RECONSTRUCTION IN STEREO VISION

Salvador Gutiérrez and José Luis Marroquín
Centro de Investigación en Matemáticas
Apdo. Postal 402, C.P. 36000, Guanajuato, Gto., México
801-422-4884, FAX: 801-422-0201, e-mail: salvador@et.byu.edu, jlm@cimat.mx

Abstract – **The matching approach to solve the so-called correspondence problem in static, binocular stereo vision has intrinsic limitations. Specifically, matching is of no use in occluded areas because there is nothing to match in those regions. Other kinds of problems, like large regions of the image with a very homogeneous texture result in erroneous matching in almost every case. Disparity in such regions can be determined with a different approach, based on well known facts and principles of stereo vision. One such method is proposed in this paper and its performance is compared to state of the art stereo algorithms. The proposed methodology is based upon diffusion of the most likely disparity hypotheses for pathological regions. This diffusion approach is founded upon well known principles of stereo vision.**

I. Introduction

Binocular stereo vision with fixed non-color cameras with planar retinas is a standard problem in many applications, mainly in robot navigation and passive range-finding applications. There are quite a few solutions available in the market and new algorithms are proposed regularly. Yet, in spite of more than 20 years of research, the results attainable with available methods tend to show some defects. Such are becoming more important as more applications of stereo-vision to computer graphics and virtual reality are developed. Moreover, these defects reveal basic mechanisms that have not been well understood in the process of stereoscopic vision. It still is difficult to obtain detailed and smooth disparity maps with good edge definition, although some reported methods approach this ideal under certain circumstances.

Some of the problems in matching show up as

- "disparity spills" (objects in the foreground appear fatter as they get closer to the cameras), and
- incorrect disparity edges.

The existence of these problems suggests that the matching (or *correspondence*) problem has not been appropriately approached. Matching is *impossible* in occluded areas because occluded points are inexistent in either image of a stereoscopic pair. Most existing algorithms come up with some disparity value assigned to those areas, while a few algorithms locate occluded areas and leave them blank. Whatever way these disparities are assigned it involves a guess, which in some algorithms is not even consciously made, although many use heuristics (justified to different degrees) to fill the gaps.

Matching large homogeneous areas is also an ill-posed problem, since all points of these regions would match almost all points of the corresponding region in the other image of the stereoscopic pair. These facts imply that stereo algorithms should provide disparity values for these regions, which are *not* found by a matching process. We propose an approach to fill those gaps with the most likely hypothesis based on a novel framework for Bayesian estimation and diffusion of disparity values.

Estimated disparity functions have a domain that corresponds to either image of a stereoscopic pair, so we should always get two disparity images, one referred to the left image and one to the right. Because of the definition of disparity, as the difference in relative position of the projections of a point on the left and right cameras, knowledge of the disparity assigned to one point in either camera allows finding the correspondent point on the opposite camera. Thus, if we go from one point in a camera to its corresponding point in the other camera and back, we should reach the same point from which we started this process. This is not always the case for three basic reasons:

- occluded points can not be mapped to the other side,
- it is very likely that there are mismatches, specially in areas of very homogeneous intensity, or where corresponding images are greatly distorted,
- disparities are only estimated to a certain precision.

We use these facts to evaluate the consistency of the two disparity images output by the algorithm, specially for pathological regions (occluded, homogeneous, and very distorted areas). The proposed approach locates pathological areas, assigns them disparity values which are the most likely according to the context, and evaluates the consistency of these assignements with respect to the basic relation between left and right images, given the disparities. When there are inconsistent points, the corresponding estimates are revised in accordance to the context and consistency is again assessed. If the stereoscopic pair of images is not greatly pathological, this process can be made to converge until an arbitrarily consistent estimate is obtained.

A brief review of the state of the art in static stereo algorithms is given in section II. The basic principles of stereo matching are reviewed in section III. Section IV describes the novel technique used to compute disparity

values, stating it within the framework of Bayesian estimation. Section V explains the concept of disparity consistency, which is later used to assess quality in the pathological regions reconstruction process. Then in section VI we describe the proposed methodology to find implicit occlusions and homogeneous areas, measuring consistency of disparity estimates and restoration of disparity in those regions. The results obtained from this method's application can be seen in section VII, where a comparison with some state of the art algorithms is shown.

II. State of the Art

Here we provide a brief review of the state of the art in stereo vision. An exhaustive survey of the literature is beyond the scope of this document. Some books cover the basics of the subject: [1], [2], [3], [4], [5], [6].

Bela Julesz was using computer synthesized stereoscopic pairs to elucidate binocular depth perception in the 1960's [7]. Some of the first computer algorithms to find depth from an arbitrary stereoscopic pair were devised in the 1970's [8], [9], [10], when researchers developed cooperative algorithms to investigate stereopsis.

The *correspondence problem* (stereo matching), has had a more or less continuous evolution with its ups and downs. From the beginning, the difficulty of the matching problem was recognized and a set of constraints and rules were proposed to limit the number of possible matchings [10]. Since good quality matchings occur only sparsely along a stereo pair many algorithms have concentrated on producing a *sparse* disparity map, [5]. Also, many algorithms have been devised to produce a *dense* disparity map. A review of the vast literature that has been published on stereo matching will not be attempted in this document but readers may refer to [11], and [12].

The many different approaches that have been developed differ in the kind of features or tokens used for matching, or in their conception of matching space, or in the nature of matching algorithms, or in the metrics used to judge similarity, their scope and of course, in the multitude of implementation details.

A variety of *features* have been used for matching by a number of authors including:

- Pixel to pixel stereo: Regardless of any interpolation procedure or any mode of interaction with neighboring pixels, or any support aggregation scheme, algorithms use intensity values of individual pixels to estimate disparity [13], [14].
- Window based (fixed 2D window): The basis for comparison of positions on different images is the result of a computation on the elements of a neighborhood of fixed size. Windows have been very popular and are traditional within the correlation approaches [5]. This approach has been made more robust by methods that work on a ranking of intensities of the window elements and use special metrics to compare candidate matchings [15]. An approach using more than one fixed window for each position is described by [16]. Other window-based features can involve the output of filters [17], or edge detectors [18].
- Variable 2D window: Some approaches adaptively increase the size of an initial window, depending on a threshold on a variance measure [19], [20], being more robust in large homogeneous areas of stereoscopic pairs. An advanced variable window method was proposed by [21] that finds the affine transformation that deforms the window in one of the images in such a way that a correlation measure is optimized.
- Arbitrary feature vector: A feature vector for each position is constructed with results of computations such as the output (magnitude and phase) of a bank of Gabor filters that sample all possible orientations, frequencies and scales, (as reported in [22]. Another example is a feature vector with three components: grey-level intensity in the first component, and derivatives along the x and y directions in the second and third components.
- Variable support aggregating region: Some support aggregation schemes using nonlinear diffusion implicitly work with 2D or 3D variable regions with sizes depending on the number of times the process is iterated [23], [24], [25].

All of the above choices may use color information for matching purposes increasing reliability significantly.

The *matching space* is the geometrical disposition of information useful for matching. It may be imagined as a continuous space, but algorithms work with sampled, discrete versions of it. For most matching procedures working with epipolar lines it is a 2D space with its axes corresponding to epipolar lines from the left and right images. For 3D approaches it usually is a 3D space where two of its coordinate axes are just the horizontal and vertical axes of one of the images, the third axis representing disparity or depth. Some approaches define and use a more sophisticated matching space where it is a projective 3D space, allowing conversion between disparity and depth, and transformation of information between different points of view [26]. Sometimes adding extra dimensions on top of this, which aids in decision making for disparity, color and transparency retrieval [25].

The *nature of the matching algorithm* can be very different from one approach to another:

- *Dynamic programming* which minimizes some sort of cost function is a popular choice because it allows natural statements of some constraints such as occlusions, continuity and monotonicity [16], ordering, exclusion of double occlusions [14], etc. See also [27], [28], [13], [29], [14].
- *Graph theoretical algorithms* play a role in approaches that state the matching problem as a problem in a graph [30]. When stating it particularly as a maximum flow

problem there is no need for explicit use of epipolar geometry, allowing use of multiple cameras with arbitrary geometries. The solution gives a minimum-cut that corresponds to disparity [26].

- The *Bayesian approach* allows a probabilistic statement of the matching problem, involving an imaging model that takes into account *a priori* information necessary to add constraints to possible solutions, and a prior model that reflects statistical properties of scenes where the theory is supposed to work. Bayes' Theorem combines these models giving a *posterior* distribution. Minimization of the expected value of a cost function computed with respect to the posterior distribution gives the MAP (*Maximum a Posteriori*) or the MPM (*Maximizer of Posterior Marginals*) estimator, depending on the cost function definition. The optimization problem may be solved using dynamic programming [28], [16] or when working with paradigms like Gauss-Markov-Measure-Fields [31], [32] the problem will be solvable using some other standard optimization techniques.
- *3D support region*: Using this kind of support region allows direct expression of conditions or constraints such as limited disparity gradient [33], or the coherence principle [34].
- *Phase-Based Methods*: Images are convolved with quadrature filters (v.g.; Gabor filters) and disparity is computed from the measured phase difference [35], [36]. The simplicity of these approaches is appealing and they automatically provide subpixel precision, however, the disparity range in which these methods are reliable is usually small (about one half the filter's wavelength) and it is difficult to obtain precise disparity edges. For these reasons they were not implemented for comparison and discussion in this document. This approach can be combined with motion cues to improve performance [35].
- A geometric approach using a *partial differential equations* (PDE) framework is described in [37]. It defines a variational principle that must be satisfied by the surfaces of the objects in the scene and their images (more than two). The derived Euler-Lagrange equations provide a set of PDE's which govern evolution of an initial surface towards the observed scene objects. When implemented with level sets surface evolution it can manage multiple objects. It assumes that scene objects are graphs of smooth functions and that they are perfectly lambertian. It can handle multiple views. It has been applied to simple synthetical objects.
- *Cooperative algorithms* were developed which operate on many "input" elements and reach global organization through local interaction constraints[10]. Two constraints were identified: C1, where ach point has a unique position in space at any time; and C2, where matter is cohesive. These constraints lead to two rules: R1, on uniqueness (each point from each image can be assigned at most one disparity value); R2, on continuity (disparity varies smoothly almost everywhere). These constraints and rules have been applied to random dot stereograms. Recently, Zitnick and Kanade have proposed a cooperative algorithm that works with 3D support to enforce or inhibit match values in a 3D disparity space ([38]).
- *Multi-frame* procedures use more than two images to strengthen the certainty of matches or simply provide a natural way to include information from more than two images [39], [40], [41], [42], [43], [37], [25].
- *Multi-resolution* approaches estimate disparity on a hierarchy of scales, processing large scales first and using these estimates to initialize matching procedures on smaller scales [44], [45], [46].
- *Neural Networks* have been used to infer dense disparity maps without iterative calculations through a special training method [47], and also sparse disparity maps [48].
- The stereo matching problem has been stated as a *nearest-neighbor* problem through the use of *intrinsic curves*, which are paths that a set of image descriptors trace as an image scanline is traversed from left to right [49] (reminding space phase trajectories in dynamical systems).

Metrics are those procedures used by matching algorithms to judge similarity between features. If features are pointlike, such as single pixel grey-scale values they may be compared using the absolute value of the difference of candidate points, as in [50]. Alternatively, squared differences can also be used [40]. If 1D, 2D or 3D features are used, the L_1, L_2 or L_∞ norms may be used, or alternatively a correlation measure [51], [15]. Some probabilistic approaches using Bayesian estimation employ *likelihoods* as metrics [52], [31], where greater likelihood values correspond to greater similarity.

The *scope* of most matching algorithms extends to a disparity map, though there has been some recent interest in reconstructing realistic 3D scenes mapping textures on a depth map (with applications to virtual reality in mind). These algorithms require interaction with graphics which poses new problems, since the quality of the output of most matching algorithms is not enough to meet the demands of these new applications [53], [54], [50], [55]. Some matching algorithms are now designed to retrieve disparity, color and transparency simultaneously [25].

Matching algorithms can be found in software or hardware *implementations*, sequential or parallel. Some general purpose stereo systems include two, three or more cameras, and allow video rate computation of disparities.

III. Constraints in Matching

A number of constraints have been mentioned in the literature to reduce the number of mismatches and increase the confidence of given matches. Since they are of very different nature from each other, it is not easy to imple-

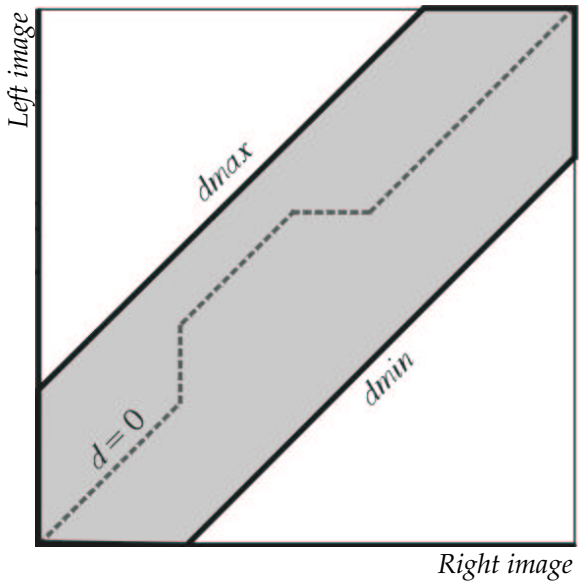

Fig. 1.
Matching space.

ment them all in a single algorithm. Although *a posteriori* checking for each constraint can always be done.

A. Uniqueness

Any valid matching should assign each point in the reference image to only one point in the search image, and viceversa. Most matching algorithms ensure uniqueness in the assignement from one image to the other, but not in the opposite direction [45].

B. Monotonicity and Ordering

The monotonicity constraint can be formulated in a right or left matching space, as defined above in section 1. A matching defines a curve in this space. Each point of a matching curve indicates that a feature from the left image is matched to a feature on the right image.This function should be monotonically increasing, otherwise the uniqueness constraint would be violated. When the monotonicity constraint is required to avoid simultaneous occlusion it is called the *ordering* constraint (see [16].

C. Disparity Gradient

When following a surface with a slope with respect to the image plane (in rectified images), the change in disparity with respect to the displacement of the projected point in the corresponding epipolar line is bounded. There is psychophysical evidence that this constraint is enforced by the human visual system [5].

D. Coherence

Disparity change within a small enough neighborhood should be limited. Most of the disparity values should be alike within the neighborhood. In other words, disparity surfaces should be piecewise continuous [34].

E. Disparity Consistency

In any two perfect disparity maps obtained from plain binocular stereo vision, corresponding points of each disparity map (left and right referred) have the same disparity values assigned, with opposite signs.

IV. Estimating Disparity

The target for reconstruction is a disparity map (referred to either the right or left image) which is modelled as a random scalar field f_L, $f_R \in \mathcal{Q} \subseteq \mathbb{R}$ (where $\mathcal{Q}$ is the set of all possible disparity values), defined over a discrete base space $\mathcal{S}$ (usually a 2D lattice coinciding with the pixel structure of an image). When not distinguishing between left and right referred disparity fields they will be denoted by f.

Reconstruction of the disparity field under these circumstances is an ill-posed problem (in the sense of Hadamard), because it does not have a unique solution. Tychonov's regularization theory allows introduction of *a priori* information necessary to solve this problem. The problem can be stated under the framework of Bayesian estimation.

If observed intensity values (images of the stereoscopic pair) are denoted by $g_L, g_R \in \mathcal{R} \subseteq \mathbb{R}$ an expression for the posterior probability of the reconstructed image given the observed image is

$$P_{f|g_L,g_R} = \frac{P_{g_L,g_R|f}\ P_f}{P_{g_L,g_R}} \tag{1}$$

by the Bayesian relation. When the necessary terms for this expression are found, those values of f maximizing this expression will provide an estimator for the reconstructed image (in fact, this would be the Maximum A Posteriori estimator). This and other estimators can be obtained by computing the expected value of suitable cost functionals with respect to this distribution.

The term P_f, denoted as the *prior* distribution, can be computed from the fact that f is a random scalar field that is defined over a lattice (as described above).

Hammersley-Clifford's theorem says that the state of the Markov Random Field can be described by a *Gibbsian* distribution

$$P_f = \frac{1}{Z_f} \exp\left[-\gamma \sum_C V_C\left(f\right)\right] \tag{2}$$

where Z_f is a normalizing constant, γ is a constant, $V_C\left(f\right)$ is a *potential* function that is evaluated and summed over

the cliques of the lattice. The specific potential function used in this case will be the Ising potential.

$$V_{\langle r,s\rangle}(f) = \begin{cases} \text{-1 if } f_r = f_s \\ \text{1 if } f_r \neq f_s \end{cases} \quad (3)$$

where f_r represents the value of field f in position $r \in \mathcal{S}$, in this case γ is a parameter controlling the granularity of the reconstructed field and $\langle r,s\rangle$ represents a clique consisting of nodes r and s.

The term $P_{g_L,g_R|f}$ can be computed from the prior knowledge that noise is additive and has a contaminated gaussian distribution $\Phi(z,\sigma,\varepsilon)$, (see equation 4).

$$\Phi(z,\sigma,\varepsilon) = \frac{1}{Z_\xi}\left[(1-\varepsilon)\exp\left(-\frac{z^2}{2\sigma^2}\right)+\varepsilon\right] \quad (4)$$

Then for each node (i,j) in the field

$$P_{g_L,g_R|f}(i,j) \quad = \quad \Phi(\Delta_{RL},\sigma,\varepsilon)$$

where

$$\Delta_{RL} = g_R(x_i,y_j) - g_L(x_i + f(x_i,y_j), y_j)$$

is the basic relationship between the left and right images of a 3D-point given by the disparity function $f(x_i,y_j)$. Since noise is identically, independently distributed we have

$$\begin{aligned} P_{g_L,g_R|f} &= \prod_{i,j} P_{g_L,g_R|f}(i,j) \\ &= \prod_{i,j} \Phi(\Delta_{RL},\sigma,\varepsilon) \end{aligned} \quad (5)$$

Note that P_{g_L,g_R} in the denominator of (eq. 1), can be regarded as a normalization constant Z_T, hence the posterior distribution can be written as

$$P_{f|g_L,g_R} = \frac{1}{Z_T}\exp(-U(f)) \quad (6)$$

where (after renaming and simplifying constants)

$$U(f) = \sum_{i,j}\rho(\Delta_{RL},\sigma,\varepsilon) + \lambda\sum_r\sum_{s\in\mathcal{N}(r)} V(f_r,f_s) \quad (7)$$

where

$$\mathcal{N}(r) = \{s \in \mathcal{S} | s \text{ is neighbor of } r\} \quad (8)$$

and

$$\rho(z,\sigma,\varepsilon) = -\ln(\Phi(z,\sigma,\varepsilon)). \quad (9)$$

The functional $U(f)$ is termed the *energy functional*. Maximizing $P_{f|g_L,g_R}$ is equivalent to minimizing $U(f)$.

A. Cost Functionals and Estimators

A cost functional $c\left(f,\hat{f}\right)$ measures how different an estimated configuration $\hat{f}$ is from the true one f. The estimation problem can be stated as finding $\hat{f}$ such that the expected value of a cost functional $c\left(f,\hat{f}\right)$ is minimized. If $c\left(f,\hat{f}\right)$ is such that

$$c\left(f,\hat{f}\right) = \begin{cases} 0 \text{ iff } f_r = \hat{f}_r,\ \forall r \in \mathcal{S} \\ 1 \text{ otherwise} \end{cases} \quad (10)$$

then minimizing the expected value of $c\left(f,\hat{f}\right)$ with respect to $P_{f|g_L,g_R}$ is equivalent to maximizing $P_{f|g_L,g_R}$ itself. Indeed, if the most probable configuration of f given g_L and g_R is chosen as $\hat{f}$ then $E\left(c\left(f,\hat{f}\right)\right) = \sum_f c\left(f,\hat{f}\right)P_{f|g_L,g_R}$ will have $P^*_{f|g_L,g_R}$ (the highest value of $P_{f|g_L,g_R}$) multiplied by a 0 coefficient, whereas choosing any other configuration will have this value multiplied by a coefficient of 1 and the 0 coefficient will multiply another $P_{f|g_L,g_R} \leq P^*_{f|g_L,g_R}$ (note that all expressions for $E\left(c\left(f,\hat{f}\right)\right)$ for all possible choices of $\hat{f}$ will have the same probability terms, one of them with coefficient 0 and all others with coefficient 1), this means that the expected cost for the most likely configuration will be less than or equal to the others. The corresponding maximizer is called the *maximum a posteriori* (or MAP) estimator[56]. When seen from this point of view the MAP estimator looks too conservative since the penalty for one mistake or for any number of mistakes is the same. Better estimators can be obtained in this way by using different cost functionals, for example, taking

$$c\left(f,\hat{f}\right) = \sum_{r\in\mathcal{S}}\left(1-\delta\left(f_r-\hat{f}_r\right)\right) \quad (11)$$

will lead to another optimal estimator by applying the general result stating that, if the posterior marginal distributions for every element of the field are known, then the optimal Bayesian estimator with respect to any additive, positive definite cost functional c may be found by independently minimizing the marginal expected cost for each element [57], [56]. Thus, in this case, knowledge of the marginals

$$P_r(q) = \sum_{f:f_r=q} P_{f|g_L,g_R}(f;g_L,g_R)$$

(where $P_r(q)$ represents the probability that a certain node $r \in \mathcal{S}$ has the disparity value q for all possible configurations of f, given g_L and g_R), allows expressing the marginal expected cost for an arbitrary element $r \in \mathcal{S}$

$$E\left(c\left(f_r,\hat{f}_r\right)\right) = \sum_{f_r\in\mathcal{Q}}\left(1-\delta\left(f_r-\hat{f}_r\right)\right)P_r(f_r)$$

By an argument similar to that one given for cost functional 10, it can be seen that minimizing the expected value is equivalent to maximizing the marginals. So, the optimal Bayesian estimator in this case is $f_r^* = q \in \mathcal{Q}$ such that $P_r(q) \geq P_r(x)$ for all $x \neq q$. This estimator is called the *maximizer of the posterior marginals* (MPM).

To compute optimal estimators of this kind (using cost functionals of the form of eqn. 11), the marginal distributions must be computed, or approximated. However, traditional approaches based on algorithms like the Gibbs Sampler (see [58]) may require a very large number of steps to get a good estimator of the true marginal distributions.

B. Gauss-Markov Measure Fields

Calculation of the estimators mentioned in the last section (particularly the MPM), can be approached by using Gauss-Markov Measure Fields (from now on they will be referred to as GMMF's, see [31], [32], [59]).

The marginal probability distributions mentioned in section IV-A can be seen as random variables that are to be estimated. In their discrete version they may be seen as vector valued random variables $p_r = (p_r(q_1), \ldots, p_r(q_m))^\top$, $m = |\mathcal{Q}|$, ($\mathcal{Q}$ is the set of disparity values considered for matching), defined on the nodes $r \in \mathcal{S}$ of an MRF (called F), with the property that $\sum_{k=1}^{m} p_r(q_k) = 1$ (from which they may be also referred to as discrete *measures*), where $p_r(q_i)$ represents the probability that the field F has the value $q_i \in \mathcal{Q}$ in node r. Marroquín has shown that the neighborhood structure of such a vector valued MRF has the same neighborhood structure as the single valued MRF used in section IV-A, (see [32]).

The GMMF approach tries to model the marginals from the posterior distribution (see eqn. 6). The proposed model is

$$P(p) = \frac{1}{Z} \exp(-U(p))$$

with

$$U(p) = \sum_r |p(r) - \hat{p}(r)|^2 + \lambda \sum_{\langle r,s \rangle} |p(r) - p(s)|^2 \quad (12)$$

where $\langle r, s \rangle$ are cliques of lattice $\mathcal{S}$. Here the likelihood $\hat{p}(r)$ is given by

$$\hat{p}(r) = \Phi((g_R(x_i, y_j) - g_L(x_i + q_k, y_j)), \sigma, \varepsilon) \quad (13)$$

Of course, there could be sites in the lattice $\mathcal{S}$ where there are no observations, or it is known a priori that the observations have a low confidence. In this case, the corresponding measure is a uniform measure

$$\hat{p}(r) = \frac{1}{m}, \quad (14)$$

see [59] for details.

Fig. 2.
Cross section of a GMMF. Darker regions correspond to disparities with higher probabilities.

A GMMF then, models disparity as a multilayered array on top of the image lattice $\mathcal{S}$, where the modes of the measures define the disparity surfaces that correspond to the 3D scene (see Fig. 2). An optimal estimator can be found by maximizing the posterior probability or equivalently, minimizing the energy functional.

A discrete GMMF can be regarded as a set W^{m-1} of m-dimensional vectors with real, positive components that add up to unity. In other words, W^{m-1} is the simplex defined by the intersection of the $(m-1)$-dimensional subspace $x_1 + x_2 + \ldots + x_m = 1$ with the region of points with positive coordinates. Its vertices are the m points $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$,...,$(0, 0, \ldots, 1)$. Thus, taking the Euclidean metric, the maximum possible distance between any couple of measures is $\sqrt{2}$. The most even distribution is represented by the barycenter of this simplex, and the most uneven distributions lie in the vertices. Unimodal distributions are points close to the vertices, multimodal distributions lie close to some of the barycenters of the faces and sides of ∂W^{m-1}, the boundary of W^{m-1}(see Fig. 3). Discrete measures may be gradually flattened or sharpened to control coupling between layers, though this should be done with care, since convolving blindly with a smoothing kernel would destroy the normalization property. Nevertheless, it is easy to see that the points of the rectilinear segment passing through the barycenter and the point corresponding to a measure, extended until it intersects one of the faces of the simplex, represent a family of discrete measures having the same modal structure but ranging from the sharpest (in one side) to the flattest in the center. This family will be referred to as the "scale space" of a discrete measure.

To compute the optimal estimators mentioned in section IV-A, the energy functionals (q.v., eqn. 7) are minimized with a simple iterative method obtained by equating the gradient to zero ($\nabla U = 0$) and solving for the components of the measure vector. This process can be interpreted as a discrete-time dynamical system whose updating formula is

$$p_r^{t+1} = \frac{\hat{p}_r + \lambda \sum_{s \in \mathcal{N}_r} p_s^t}{1 + \lambda |\mathcal{N}_r|} \quad (15)$$

where $\mathcal{N}_r$ is the set of neighbors of r.

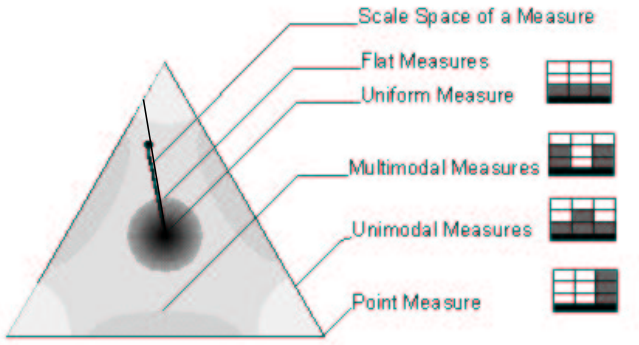

Fig. 3.
The simplex W^{m-1}, the space where discrete measures live.

Since the sum of the coefficients of the measures in the right is 1, the updated vector will lie within the convex hull of those measures and therefore will lie on the same plane. This means that if the initial vectors are measures (their components are non-negative and add up to unity), then the updated vectors in this process will keep being measures.

Since $U(p_r; \hat{p}_r)$ is convex its unique minimum will be the fixed point of this dynamical system. It has thus been shown that estimators based on posterior marginal probability distributions exist and can be computed by solving a simple optimization problem.

Parameter λ in the GMMF energy functional (see eq. 12) is generally not known, but it may be noted that when the observations term is eliminated from this formula, then the regularization parameter becomes equivalent to time and proportional to the number of iterations (see [60], [61]). The corresponding updating formula turns out to be an ordinary average

$$p_r^{t+1} = \frac{\sum_{s \in \mathcal{N}_r} p_s^t}{|\mathcal{N}_r|}. \tag{16}$$

This dynamical system is the foundation of the cellular automata used to implement regularization in this proposal although of course, there are many efficient methods that can find the regularized estimator, such as conjugate gradient or using the cosine transform. In any case, GMMF's are a much more efficient methodology for obtaining optimal estimators than the use of Monte Carlo techniques and provide a simple way to incorporate any *a priori* information. For example, Dynamic Programming matching algorithms naturally facilitate enforcement of the ordering constraint, which is important a priori knowledge, difficult to implement in other contexts. To illustrate this point, a solution from a Dynamic Programming matching algorithm (run on the Matiushka

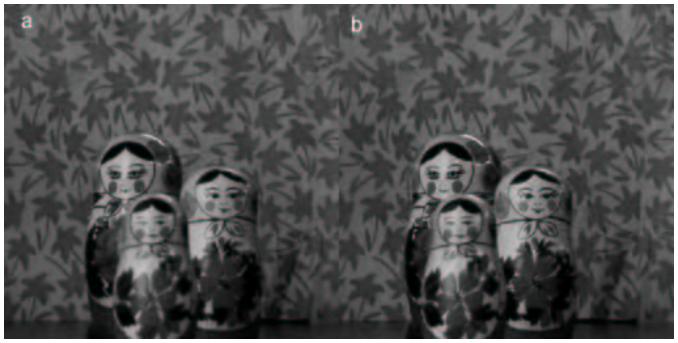

Fig. 4.
Matiushka stereoscopic pair. Left image (a), right image (b).

Fig. 5.
Incorporating information from different sources using GMMF's.

stereoscopic pair, see Fig. 4) was used to reinforce the entries in each measure of a GMMF that resulted from a matching algorithm that ignored the ordering constraint. In each measure the entry corresponding to a (order constraint compliant) match was multiplied by a constant factor $1 + \varepsilon$, and measures were subsequently renormalized. In Fig. 5(left) the MLE after reinforcement looks exactly like the output of the DP algorithm, showing the typical artifacts of those algorithms (flat surfaces and horizontal lines). After a few regularization iterations most of them have disappeared while the ordering constraint prevails on the disparity surfaces, see Fig. 5 (center). The frame on the right of Fig. 5 shows the disparity results superposed to one of the original images of the stereo-pair.

V. Disparity Consistency

The epipolar constraint (see [5]), states that the two possible projections of a point of the 3D scene (one on each camera), lie on corresponding linear segments of the cameras, so matching corresponding points is an inherently 1-dimensional problem.

The basic relation for disparity in stereo using coplanar cameras has a simple expression when stated for corre-

sponding segments:

$$d = x_l - x_r = \frac{fb}{Z}, \tag{17}$$

where x_l and x_r are the 1D-coordinates corresponding to each segment, f is the focal length, b is the baseline and Z is the depth coordinate.

After any matching process for binocular stereo, two disparity functions can be obtained, one referred to the left camera and another to the right one. Discrete disparity functions $d_i(x), \quad i \in \{l, r\}$ assign a real disparity value to each discrete position of corresponding segments $\mathcal{S}_i, \quad i \in \{l, r\}$ in either camera, i.e.;

$$d_i(x_i) \quad \in \quad [d_{\min}, d_{\max}] \subseteq \mathbb{R}, \quad i \in \{l, r\}. \tag{18}$$

Then, a basic model for matching can be written as follows

$$\begin{aligned} g_r(x_r) &= a_1 g_l(x_r + d_r(x_r)) + n(x_r, \sigma, \varepsilon) + a_2, \\ g_l(x_l) &= a_1 g_r(x_l + d_l(x_l)) + n(x_l, \sigma, \varepsilon) + a_2, \end{aligned} \tag{19}$$

where g_r and g_l are the intensity functions for each corresponding segment, a_1 models difference in contrast between cameras, a_2 represents a difference in bias and $n(x_i, \sigma, \varepsilon), i \in \{l, r\}$ stands for noise, modelled with a random variable having a normal distribution contaminated with a constant, (see eqn. 4).

Assumption of the following ideal conditions:

- perfect matching,
- absence of noise,
- no photometric variation and
- no projective distortion,

simplifies the statement of the relations between corresponding camera points. If $x_r \in \mathcal{S}_r$ and $x_l \in \mathcal{S}_l$ are corresponding points in segments $\mathcal{S}_r$ and $\mathcal{S}_l$, respectively, then under these ideal conditions their intensity values should be the same, i.e.,

$$g_r(x_r) = g_l(x_l),$$

where

$$\begin{aligned} x_l &= x_r + d_r(x_r), \\ x_r &= x_l + d_l(x_l), \end{aligned} \tag{20}$$

so we may state that

$$d_r(x_r) = -d_l(x_l). \tag{21}$$

Relations (20) show that if d_l and d_r are known, then it is possible to map points from one camera of the stereo-pair onto points of the opposite camera. Let corresponding points be called "conjugates". Notice that the conjugation mapping $\overline{x_r} : x_r \mapsto x_r + d_r(x_r)$ is defined only for those points that are visible in the right camera of the stereo-pair, and similarly $\overline{x_l} : x_l \mapsto x_l + d_l(x_l)$ is only defined for those points visible in the left camera. The *image* of the left camera on the right camera under this mapping is $\overline{\mathcal{S}_l}$. The complement of its intersection with the points of the right camera are the *left occlusions*, and the *right occlusions* are defined in a similar way, i.e.:

$$\begin{aligned} O_l &= \left(\overline{\mathcal{S}_l} \cap \mathcal{S}_r\right)', \\ O_r &= \left(\overline{\mathcal{S}_r} \cap \mathcal{S}_l\right)'. \end{aligned} \tag{22}$$

For all points x in either camera but not in the occluded areas, the conjugate mapping is involutive, i.e., if $O = O_l \cup O_r$, we have

$$\overline{\overline{x}} = x, \quad \forall x \notin O. \tag{23}$$

In practice, the assumptions mentioned above will be violated to some degree, resulting in inconsistent disparity assignements, i.e., equality will not hold in eqn. (23). So it is possible to use the **SSD** metric to assess the degree of inconsistency between left and right disparity functions:

$$\mathcal{I}_O(d_i) = \sum_{x \in \mathcal{S}_i} \left(x - \overline{\overline{x}}\right)^2, \quad i \in \{l, r\}, \quad x \notin O. \tag{24}$$

When a process, such as that proposed in the next section, assigns disparity values to the occluded regions O, the inconsistency of the assignement can be evaluated with essentially the same metric, though applied to all points or to specific regions in either image:

$$\mathcal{I}(d_i) = \sum_{x \in \mathcal{S}_i} \left(x - \overline{\overline{x}}\right)^2, \quad i \in \{l, r\}. \tag{25}$$

VI. Proposed Approach

The proposed general strategy consists in finding those regions of each map whose values cannot be reliably determined with an ordinary matching process, and to fill them by propagating the best disparity hypotheses based on known, sound principles of stereoscopic vision, for example, the coherence principle (see section III). These areas are occlusions, regions with very homogeneous texture and points where matching is suspected to be unreliable because some matching quality criterion is not met, such as those having a very flat marginal probability distribution of disparity or having its maximum in either end of the searching interval considered.

After this process, left and right disparity consistency can be checked and inconsistent regions filled with the same hypotheses propagation processes.

Large unreliable regions are not desirable because they take longer to fill and because reliability of the substitute values decreases as those areas increase (coherence is valid only in small neighborhoods, unless available a priori information determines large coherent regions). The

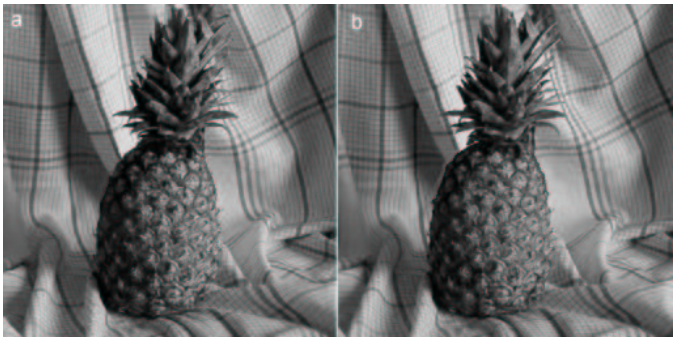

Fig. 6.
Pineapple stereoscopic pair. Left image (a), right image (b).

Fig. 7.
Maximum likelihood disparity, pixelwise matching.

straightforward approach would be to find all *pathological* regions first (occlusions, homogeneous, unreliable and inconsistent areas) , and then to propagate hypotheses in a second stage, tends to produce large unreliable areas. So, it is wiser not to deal with all these regions at the same time. The following general steps lead to good results (they are later discussed in detail):

1. **Compute** initial MLE of disparity referred to the target view.
2. **Detect** homogeneous areas on the target view, where matching is known to be deficient.
3. **Regularize** non-homogeneous regions.
4. **Define** unreliable regions from homogeneous areas.
5. **Propagate** hypotheses in unreliable regions
6. **Detect** occlusions determined by the obtained disparity maps.
7. **Define** unreliable regions from occlusions
8. **Propagate** hypotheses in unreliable regions.
9. **Repeat**
10. **Define** unreliable regions from areas with inconsistent left-right disparity.
11. **Propagate** hypotheses in unreliable regions.
12. **Until** convergence

The strategy was applied to the Pineapple stereoscopic pair (see Fig. 6) to illustrate the results.

A. Step 1: Computing Initial MLE of Disparity

The MLE of disparity is found by computing the likelihood field $\hat{p}$ and choosing the value of disparity $q_{\max}$ that maximizes the likelihood defined on that position; i.e., $q_{\max} = \arg\max_q \hat{p}_r(q)$. An example of such an estimator can be seen in Figure 7.

B. Step 2: Detecting Homogeneous Areas

Homogeneous areas are detected with a threshold on gradient magnitude. The gradient is computed using convolution with gaussian derivatives. The parameter values used are $\sigma = 0.33$ for the gaussians and the threshold $\theta = 0.015$. Then, a point (i, j) is considered to be in a homogeneous area if the gradient magnitude $\|\nabla_\sigma(i,j)\| \leq \theta$. An example of these areas can be seen in Figure 8.

Fig. 8.
Homogeneous regions.

Homogeneous regions are *filled* propagating the best disparity hypotheses with a three step diffusion process described below in section VI-E.

C. Step 3: Regularizing Non-homogeneous Regions

Those areas complementing the homogeneous regions found in the previous step are regularized with a diffusion process implemented with a cellular automaton updating layers of the disparity GMMF only on those positions outside the marked homogeneous regions.

Algorithm 1: Cellular Automaton for Diffusion in a Selected Area

Input:

- A GMMF F consisting of the marginal probability distributions of disparity $p_r(q_k)$, $q_k \in \mathcal{Q}$, $k \in \{1, \ldots, m\}$, over a lattice $\mathcal{S}$ corresponding to either the left or right image of a stereoscopic pair, $r \in \mathcal{S}$.
- A mask (2D array) H_r having a value of 1 on those positions $r \in \mathcal{S}$ lying on a homogeneous region and 0

elsewhere.

Description: This algorithm processes each layer of the GMMF by substituting the value in each position marked as non-homogeneous with the average of its neighbors (if they exist). Values in homogeneous areas are left without change. It implements the updating formula shown in eqn. (16).

After all layers of the disparity GMMF have been processed, a new estimate of disparity can be obtained by choosing for each position r the disparity value that maximizes the measure P_r defined on that position (see section IV). This process should be applied for as many iterations as necessary to clean the treated areas from noise without excessive blurring of contours. The results of this step are shown in Figure 9 (a).

D. Step 4: Defining Unreliable Regions from Homogeneous Areas

The homogeneous regions found in Step 2 must be initialized before diffusion with boundary conditions is performed. These conditions are determined by the following Algorithm 2:

Algorithm 2: Determination of Boundary Conditions

Input: A disparity map D_{ij} where $i = 1, \ldots, n$, $j = 1, \ldots, m$, taking values from a known range $[d_{\inf}, d_{\sup}]$. A mask array Z_{ij} where $Z_{ij} = 1$ iff (i, j) is included in an unreliable region, and $Z_{ij} = 0$ elsewhere.

Description: The algorithm scans each line of the mask Z_{ij}. When it enters an interval marked with $1's$, it records the corresponding disparity value from D_{ij} at the beginning of the interval, and records the disparity value at the other end. It then compares both disparity values and marks the end corresponding to the smaller value as initial condition for diffusion by setting $Z_{ij} = -1$. This leaves mask Z_{ij} with value 1 on unreliable regions, 0 outside unreliable regions and -1 on positions whose disparity value should be taken into account, but not updated, by the diffusion process.

E. Step 5: Propagating Hypotheses in Unreliable Regions

Unreliable regions coming from occlusion or homogeneous regions should be assigned disparity values on principles other than matching, such as coherence [34], continuity [10], and adjacency).

Propagation of disparity hypotheses on unreliable regions is carried out with non-homogeneous diffusion with boundary conditions. This step can be implemented with the following cellular automata that updates the disparity map, using information from map Z_{ij} as left by the just described Algorithm 2.

Algorithm 3: Cellular Automaton for Diffusion with Boundary Conditions

Input:

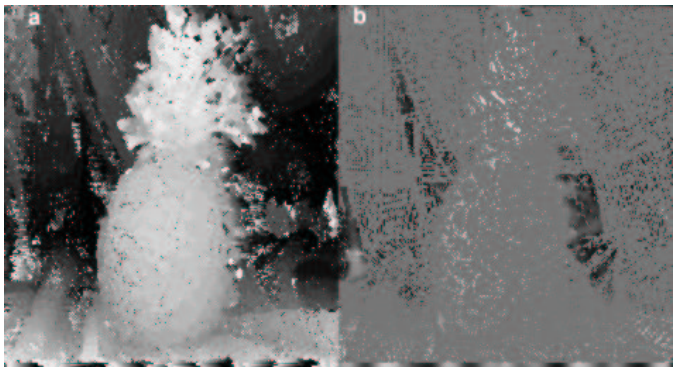

Fig. 9.
Left: Regularizing the complement of homogeneous areas. Right: Regularization of homogeneous regions.

- A disparity map D_r, where $r \in \mathcal{S}$, the pixel lattice of either view (left or right).
- An array Z_r, where $r = (i, j)^\top \in \mathcal{S}$, and Z_r is the output of Algorithm 2, indicating boundary conditions for diffusion.

Description: This algorithm performs diffusion with boundary conditions along an unreliable region. It does so by iteratively updating the value of the disparity map with the average of those neighbors that are either boundary conditions or points inside the unreliable region. It is based on formula (16), modified to diffuse with boundary conditions.

Updating is done until convergence.

The results of this step are illustrated in Figure 9 (b).

F. Step 6: Detecting Occlusions Determined by the Obtained Disparity Maps

Right and left occlusions determined by the so far estimated disparities are detected as described in section V.

The results of this process are illustrated in Figure 10 (b).

G. Step 7: Defining Unreliable Regions from Occlusions

Occlusions obtained in Step 6 are usually too irregular since they contain lattice induced occlusions, which are produced by discrete changes in disparity. A regularized unreliable region is obtained by the following method:

1. Construct a single layer GMMF with $p(r) = 1$ if r is in the occluded region O_r as determined by Step 7 (see VI-G), and $p(r) = 0$ otherwise. Apply Algorithm 1 (see VI-C) to GMMF $p(r)$ for as many iterations as necessary to achieve a regular region.
2. Apply a threshold θ on the result such that $\hat{O}_r := 1$ if $O_r > \theta$, $\hat{O}_r := 0$ otherwise.
3. Apply Algorithm 2 (see VI-D) to $\hat{O}_r$, to determine boundary conditions.

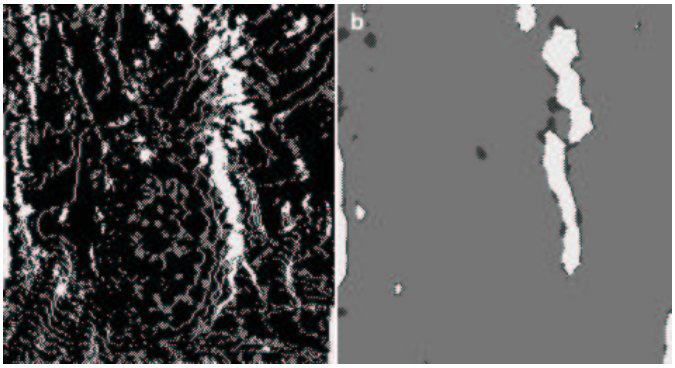

Fig. 10.
Left: Right occlusions. Right: Significant occlusions are masked and their boundaries are regularized.

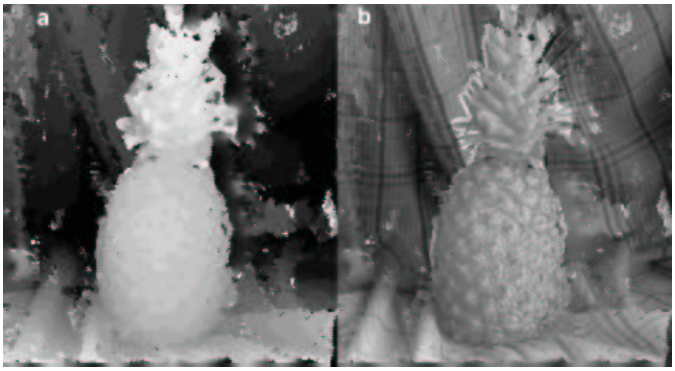

Fig. 11.
Left: Disparity after propagation of the most likely hypothesis in occluded regions. Right: Disparity function superposed to original view.

The results of this process are illustrated in Figure 10 (b) and Figure 11.

H. Step 8: Propagating Hypotheses in Unreliable Regions

By definition occluded points can not be matched and no disparity value can be assigned to occluded regions *solely on the basis of matching*. Disparity hypotheses for these regions must be derived from application of other principles, like Marr and Poggio's continuity constraint [10], Prazdny's coherence constraint [34], the occlusion constraint and the adjacency principle (see [14]). This is just what this step achieves and is executed exactly as Step 5 , only over the output of Step 7 . Results after hypotheses propagation in occluded areas are illustrated in Figure 11.

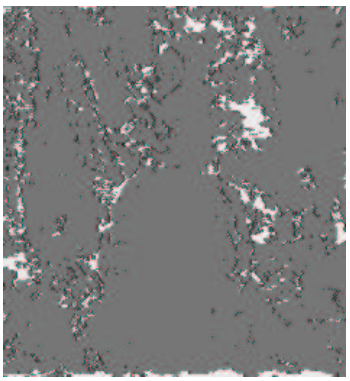

Fig. 12.
Regions with inconsistent disparities.

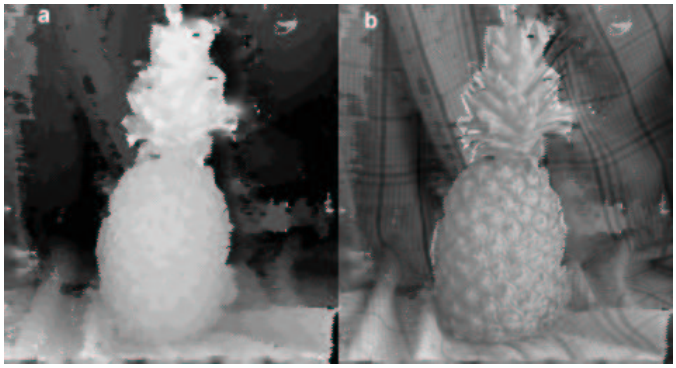

Fig. 13.
Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

I. Step 9: Defining Unreliable Regions from Areas with Inconsistent Left-Right Disparity

Disparity functions map points from one image to another. Repeated application of this kind of mapping should bring us map to the starting point. When this is not true, disparity functions are inconsistent. A threshold on the degree of inconsistency can be used to define regions on the domain of disparity functions, that need to be recalculated. See Figure 11. Consistency between disparity functions is found as described in section V, see Fig. 12.

J. Step 10: Propagating hypotheses in Unreliable Regions

This step propagates hypotheses exactly as in Step 5, only over the results of Step 9. The results of this step are shown in Figure 13.

VII. Results

A. Assumptions and Limitations

The proposed approach is useful under certain circumstances. The stereoscopic pair images are assumed to be rectified; i.e., their epipolar lines are horizontal and parallel. The effective size of the images of the stereoscopic pair depends on sensor characteristics, available memory and processor speed, but the proposed method keeps being competitive with image sizes from 256×256 to 512×512. A non-optimized program running on a PC with a 450 MHz Pentium III typically takes 1 minute and 15 seconds to process an ordinary stereoscopic pair of 256×256 images.

The effective range of disparities in this method (as in all other considered methods) is limited by factors such as baseline length, apparent size of objects of interest, effective size of images, and scene complexity. Nevertheless, with present day resources it is possible to obtain good results resolving up to 32 disparity levels requiring time in the order of minutes. The GMMF's memory requirements increase linearly with respect to disparity range. Nevertheless, a fundamental limitation of this method (as of any other method based in matching) is baseline length.

Big homogeneous areas represent a problem for diffusion based methods because they may take a long time to converge. Also, the larger the occluded areas, the longer they will take to be filled.

Untextured backgrounds favor wrong hypotheses diffusion and it is all too easy in this case to go past the occluded areas found by mapping from one image to the other. This inconvenience can be alleviated using an erosion morphological operator to enlarge the occluded region so that it can reach the true edges of the foreground object.

B. Visual and Quantitative Comparison

To aid in comparing the performance of the proposed approach with that of well known state of the art algorithms, a set of algorithms were applied to the same stereoscopic pair. First, a visual comparison of the results on a natural scene is made and later a quantitative comparison is made using a synthetic stereoscopic pair.

B.1 Visual Comparison

Two versions of point matching algorithms were tested on the pineapple stereoscopic pair. The first algorithm matches intensities and the sign of the first derivative of the intensity function at each pixel along epipolar lines. The results are shown in Fig. 14, where the granular nature of this kind of matching is apparent. The algorithm tested in second place matches just the values of the intensity function and obtains an initial MLE which is then regularized (see Fig. 15). It is apparent that using a token as simple as plain intensity matching can be very satisfactory if it is followed by a regularization stage. However, a certain degree of "spill" is unavoidable.

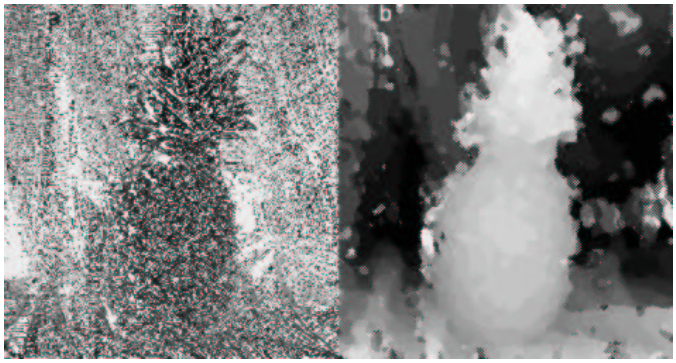

Fig. 14.
Derivative matching tokens. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

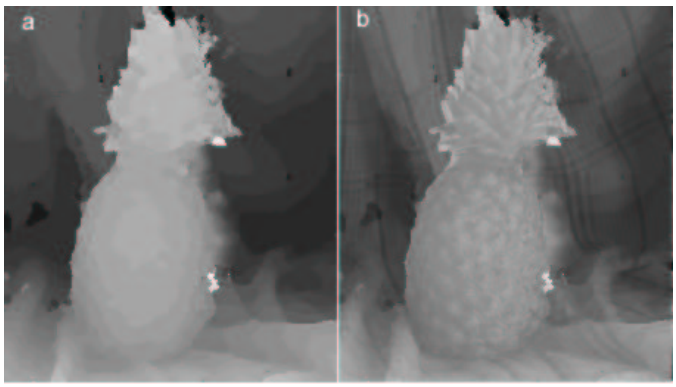

Fig. 15.
Point tokens. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

Window based methods tend to produce more conspicuous "spill" on occluded regions; for example, results of a 9×9 correlation window [5] can be seen in Fig. 16. Bhat and Nayar's robust window method using ranking and sophisticated metrics [15] can be seen in Fig. 17. Kanade's variable window approach [20] is shown in Fig. 19. Robert Maas' variable window with model based window size selection [21] can be seen in Fig. 20. All of them lack definition in areas to the right of the foreground object, where occlusions are. Also, a big homogeneous zone is still filled to some degree with a wrong disparity hypothesis in all these methods. Scharstein and Szeliski's stereo matching with non-linear diffusion [24] can be seen to have very good definition (at the cost of a noisy disparity map) except on right occlusions (see Fig. 18), a "spill" is also appreciable over the very homogeneous area to the middle-

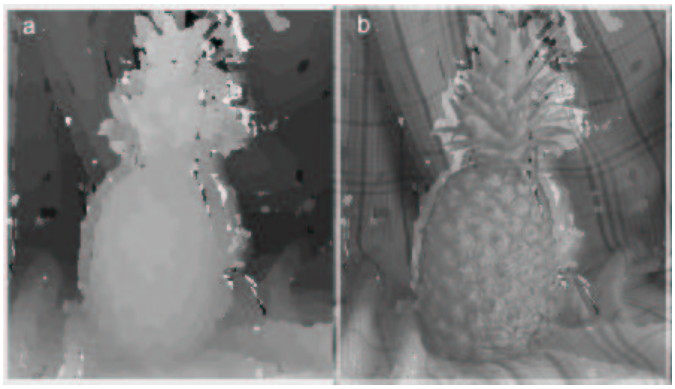

Fig. 16.
Correlation $9x9$ window. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

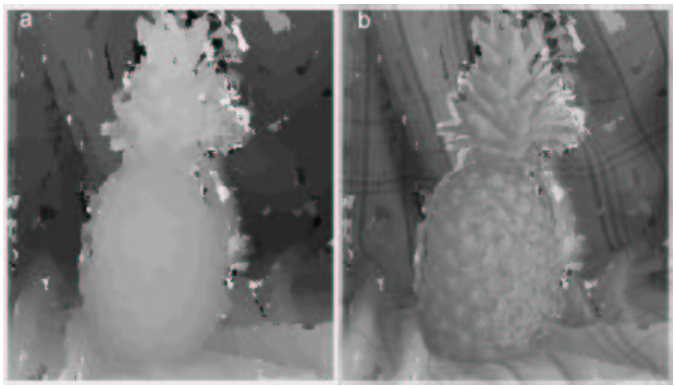

Fig. 17.
Using ranking as in ref. [15]. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

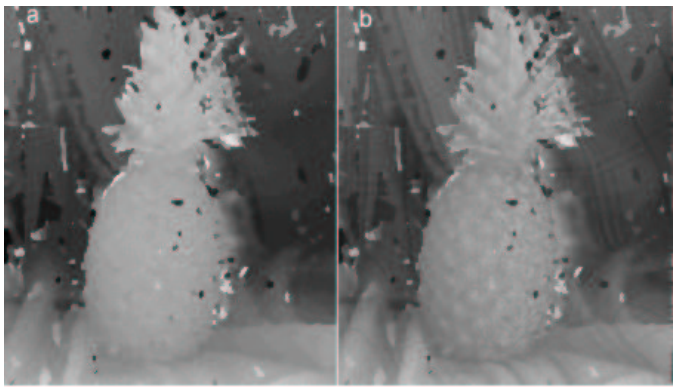

Fig. 18.
Implicit support as in ref. [24]. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

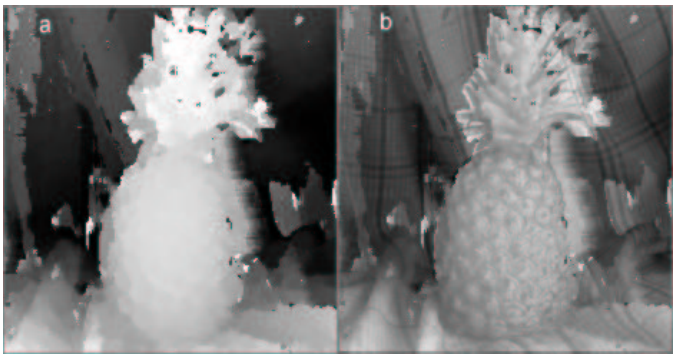

Fig. 19.
Kanade variable window, as in ref. [20]. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

right part of the pineapple. More regularization steps for adequate denoising necessarily blur detail.

The proposed method can give a more detailed disparity map without compromising with noise because the last stages take care of disparity inconsistencies (see Fig. 13), so very few initial regularization iterations are needed. Unreliable areas that cannot be matched are filled with the most likely hypotheses given the principles of coherence [34], continuity [10], and adjacency.

Another example of application of the proposed method can be seen in the Castle stereoscopic pair (Figs. 21 and 22).

B.2 Quantitative Comparison

In order to quantify the performance of different methods, a synthetic stereoscopic pair and a performance measure were devised. The synthetic stereo pair shown in

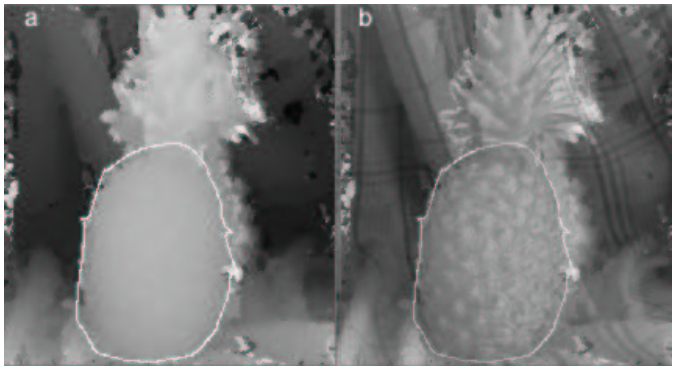

Fig. 20.
Maas variable window, as in ref. [21]. Left: Reconstructed disparity function. Right: Disparity values superposed to original image.

Fig. 21.
Castle stereoscopic pair. Cropped from the original CMU test pair. Left hand (a), right hand (b).

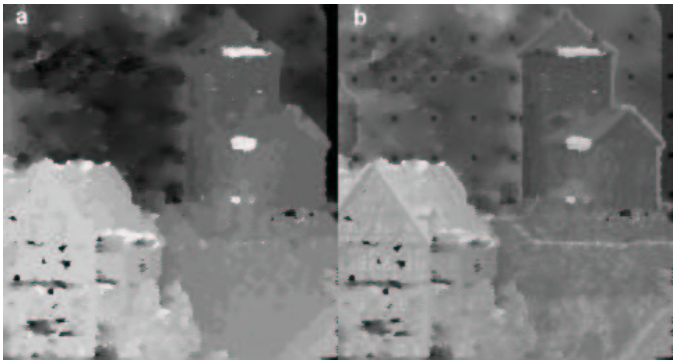

Fig. 22.
Results for the Castle stereoscopic pair. The right hand picture depicts disparity overlayed with original image.

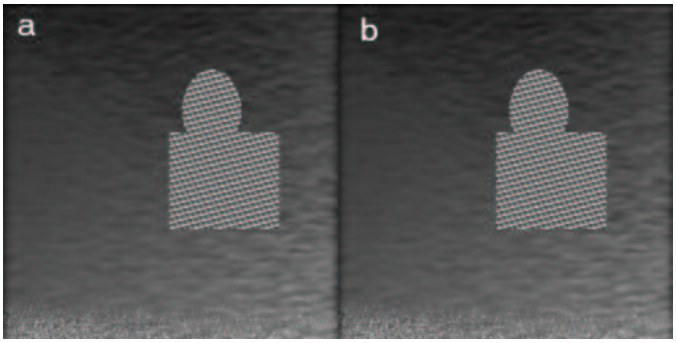

Fig. 23.
Synthesized stereoscopic pair. a. left, b. right

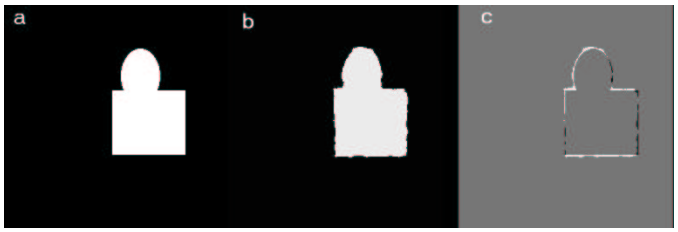

Fig. 24.
a.- Ground truth, b.- Segmented output, c.- Classification error

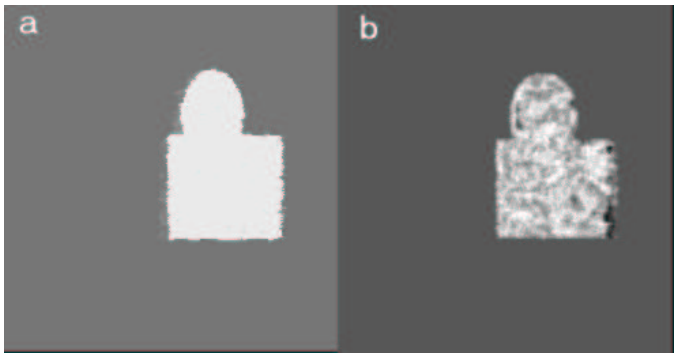

Fig. 25.
Results for $\sigma = 0.7$ noise, a.- Proposed method, b.- Scharstein-Szeliski

Fig. 23 was built. It consists of a single, simple geometric object in the foreground filled with a periodic pattern with a period bigger than the range of searched disparity values, to eliminate aliasing, and an arbitrary texture in the background.

The considered algorithms are compared in their ability to resolve true disparity edges, so a special metric was devised with this purpose in mind. The performance measure takes a disparity map produced by the studied algorithm as input. A binary image is obtained by thresholding the disparity map (the threshold was exactly one half the disparity range and was the same for all methods), segmenting foreground and background. The resulting image is compared to the mask put in the corresponding position and classification errors are counted and expressed as a percentage of the total number of pixels in the image (see Fig. 24). Disparity borders overriding the true ones appear as background pixels labeled as foreground (white in Fig. 24, c); this will be known as type I error. Disparity borders not reaching the corresponding true one appear as foreground pixels labeled as background (black in Fig. 24, c); this will be type II error.

To test the algorithms' sensitivity to noise, the test stereoscopic pair was altered with normally distributed noise with $\mu = 0$ and three levels of $\sigma = \{0.2, 0.7, 1.0\}$.

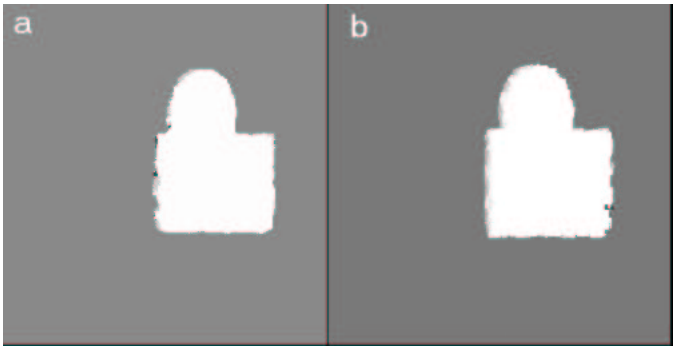

Fig. 26.
Results for $\sigma = 0.7$ noise. a.- Bhat & Nayar, b.- Okutomi-Kanade

TABLE I
PROPOSED METHOD:

Left Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	0.43	0.43	0.58	0.38
II	0.21	0.30	0.10	0.48
Total	0.64	0.73	0.68	0.86
Right Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	0.49	0.44	0.85	0.43
II	0.19	0.32	0.15	0.51
Total	0.68	0.76	1.00	0.94

TABLE II
SCHARSTEIN-SZELISKI:

Left Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	1.49	1.15	1.03	1.15
II	0.15	0.54	1.39	0.67
Total	1.64	1.69	2.42	1.82
Right Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	1.41	1.09	0.99	0.81
II	0.14	0.52	0.96	1.69
Total	1.55	1.61	1.95	2.50

TABLE III
BHAT & NAYAR:

Left Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	1.64	1.17	1.08	0.96
II	0.15	0.14	0.06	0.11
Total	1.79	1.31	1.14	1.07
Right Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	1.97	1.31	1.13	0.97
II	0.21	0.13	0.04	0.07
Total	2.18	1.44	1.17	1.04

TABLE IV
OKUTOMI-KANADE:

Left Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	3.09	3.44	2.99	2.51
II	0.00	0.00	0.00	0.01
Total	3.09	3.44	2.99	2.52
Right Image				
Error Type	$\sigma = 0$	$\sigma = 0.2$	$\sigma = 0.7$	$\sigma = 1$
I	3.12	3.42	2.95	2.43
II	0.00	0.00	0.00	0.00
Total	3.12	3.42	2.95	2.43

Outputs from the algorithms were analyzed with the performance measure; some results can be seen in Fig. 25 and Fig. 26, and the computed error is reported in tables I to IV.

The studied algorithms tend to show higher values for error type I than for error type II, which essentially means that foreground objects tend to appear fatter than they are, although some dents into the true edges can always be found, too. Nevertheless, the proposed method shows significantly less error levels for both types.

VIII. CONCLUSION

It has been shown that the matching approach to solve the so called correspondence problem in stereo vision has intrinsic limitations. Specifically, matching is of no use in occluded areas because there is nothing to match in those regions. Other kinds of problems, like large regions of the image with a very homogeneous texture will result in erroneous matching in almost every case. A method was proposed to compute disparity in such regions using a different approach, based on well known facts and principles of stereo vision, and its performance was compared to state of the art stereo algorithms. The proposed methodology is based upon diffusion of the most likely dis-

parity hypotheses for pathological regions. This diffusion approach is founded upon well known principles of stereo vision, such as Marr and Poggio's continuity constraint [10], Prazdny's coherence constraint [34], the occlusion constraint and the adjacency principle (see [14]). The principle of consistency of left and right disparity functions is precisely stated and used to measure the goodness of disparity assignements on regions where matching should not be used.

References

[1] B. Julesz, *Foundations of Cyclopean Perception*. Chicago and London: The University of Chicago Press, 1971.

[2] W. E. L. Grimson, *From Images to Surfaces*. Cambridge, Massachusetts: MIT Press, 1981.

[3] H. L. F. V. Helmholtz, *Treatise on Physiological Optics*. New York: Dover, 1925.

[4] B. K. P. Horn, *Robot Vision*. Cambridge, Massachusetts: MIT Press, 1986.

[5] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[6] R. N. Klaus Voss and M. Schubert, *Monokulare Rekonstruktion Für Robotvision*. Verlag Shaker, 1996.

[7] B. Julesz, "Binocular depth perception of computer-generated patterns," *Bell System Tech.*, vol. 39, pp. 1125–1161, September 1960.

[8] J. I. Nelson *Journal of Theoretical Biology*, vol. 49, pp. 1–xx, 1975.

[9] P. Dev *International Journal of Man-Machine Studies*, vol. 7, pp. 420–xxx, 1975.

[10] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *SCIENCE*, pp. 283–287, 1976.

[11] S. T. Barnard and M. A. Fischler, "Computational stereo," *Computing Surveys*, vol. 14, no. 4, pp. 553–572, 1982.

[12] U. R. Dhond and J. K. Aggarwal, "Structure from stereo - a review," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489–1510, 1989.

[13] I. J. Cox, S. L. Hingorani, and S. B. Rao, "A maximum likelihood stereo algorithm," *Computer Vision and Image Understanding*, vol. 63, pp. 542–567, May 1996.

[14] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 269–293, 1999.

[15] D. N. Bhat and S. K. Nayar, "Ordinal measures for visual correspondence," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*, pp. 351–357, 1996.

[16] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and binocular stereo," *IJCV*, vol. 14, pp. 211–226, April 1995.

[17] D. G. Jones and J. Malik, "A computational framework for determining stereo correspondences from a set of linear spatial filters," in *Second European Conference on Computer Vision (ECCV'92)*, (Santa Margherita Liguere, Italy), pp. 397–410, Springer-Verlag, 1992.

[18] H. H. Baker, *Edge Based Stereo Correlation*, pp. 168–175. L.S. Baumann (Ed.), 1980.

[19] R. D. Arnold, "Automated stereo perception," Tech. Rep. AIM-351, Artificial Intelligence Laboratory, Stanford University, 1983.

[20] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 16, pp. 920–932, September 1994.

[21] M. A. V. Robert Maas, Bart M. Ter Haar Romeny, "Area-based computation of stereo disparity with model-based window size selection," in *Proceedings of the IEEE Computer Society Conferenceon Computer Vision and Pattern Recognition (CVPR'99)*, pp. 106–112, IEEE, 1999.

[22] S. Gutiérrez, *Robust Methods for Disparity Estimation in Stereo Vision*. Ph.D. thesis, Centro de Investigación en Matemáticas (CIMAT), Apdo. Postal 402, Guanajuato, Guanajuato, México, C.P. 36000, Mar 2001.

[23] R. Szeliski and G. Hinton, "Solving random-dot stereograms using the heat equation," (San Francisco, California), pp. 284–288, IEEE Computer Society Press, 1985.

[24] D. Scharstein and R. Szeliski, "Stereo matching with non-linear diffusion," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPR'96), San Francisco, California*, pp. 343–350, 1996.

[25] R. Szeliski and P. Golland, "Stereo matching with transparency and matting," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 45–61, 1999.

[26] S. Roy, "Stereo without epipolar lines: A maximum-flow formulation," *International Journal of Computer Vision*, vol. 34, no. 2/3, pp. 147–161, 1999.

[27] Y. Ohta and T. Kanade, "Stereo by intra- and inter- scanline search using dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, pp. 139–154, March 1985.

[28] P. N. Belhumeur and D. Mumford, "A bayesian treatment of the stereo correspondence problem using half-occluded regions," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 1992 (CVPR92)*, pp. 506–512, 1992.

[29] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 181–200, 1999.

[30] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Santa Barbara, California), 1998.

[31] J. L. Marroquín, "Random measure fields and the integration of visual information," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 4, pp. 705–716, 1992.

[32] J. Marroquín, F. Velasco, S. Gutiérrez, and M. Rivera, "Gauss-markov measure fields models for image processing," Tech. Rep. I-97-16 (CC/CIMAT), Centro de Investigación en Matemáticas (CIMAT), 1997.

[33] S. Pollard, J. Mayhew, and J. Frisby, "A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449–470, 1985.

[34] K. Prazdny, "Detection of binocular disparities," *BioCyber*, vol. 52, pp. 93–99, 1985.

[35] H. V. R. Figueroa, *A Filtering Approach to the Integration of Stereo and Motion*. PhD thesis, The University of Sussex, 1993.

[36] D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin, "Phase-based disparity measurement," *CVGIP: Image Understanding*, vol. 53, pp. 198–210, March 1991.

[37] O. Faugeras and R. Keriven, "Variational principles, surface evolution, PDE's, level set methods and the stereo problem," Tech. Rep. 3021, Institut National de Recherche en Informatique et en Automatique (INRIA), 1996.

[38] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, pp. 675–684, July 2000.

[39] R. Bolles, H. Baker, and D. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, pp. 7–55, 1987.

[40] L. Matthies, R. Szeliski, and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," *International Journal of Computer Vision*, vol. 3, pp. 209–236, 1989.

[41] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 353–363, April 1993.

[42] S. B. Kang, J. Webb, L. Zitnick, and T. Kanade, "A multi-baseline stereo system with active illumination and real-time image acquisition," in *Proceedings of the Fifth International Conferenceon Computer Vision (ICCV'95)*, (Cambridge, Massachusetts), pp. 88–93, 1995.

[43] R. T. Collins, "A space-sweep approach to true multi-image

matching," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*, pp. 358–363, 1996.

[44] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proceedings of the Fifth International Joint Conf. Artificial Intelligence*, (Cambridge, Massachusetts), p. 584, 1977.

[45] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proceedings of the Royal Society of London, Series B*, vol. 204, pp. 301–328, 1979.

[46] W. Hoff and N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 11, no. 2, pp. 121–136, 1989.

[47] E. Maeda, A. Shio, and M. Okudaira, "Layered neural network for stereo disparity detection," *Neural Networks*, vol. 2, pp. 141–153, 1992.

[48] J. Cruz, G. Pajares, and J. Aranda, "A neural-network model in stereovision matching," *NeurNet*, vol. 8, no. 5, pp. 805–813, 1995.

[49] C. Tomasi and R. Manduchi, "Stereo matching as a nearest-neighbor problem," *PAMI*, vol. 20, pp. 333–340, March 1998.

[50] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Nad Tanaka, "A stereo-machine for video-rate dense depth mapping and its new applications," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*, (San Francisco, California), pp. 196–202, 1996.

[51] T. Ryan, R. Gray, and B. Hunt, "Prediction of correlation errors in stereo-pair images," *Optical Engineering*, vol. 19, no. 3, pp. 312–322, 1980.

[52] J. L. Marroquín, "Local harmonic analysis and stereo tokens," tech. rep., Centro de Investigación en Matemáticas (CIMAT), 1987.

[53] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," *Computer Graphics (SIGGRAPH'95)*, pp. 39–46, 1995.

[54] R. Szeliski and S. Kang, "Direct methods for visual scene reconstruction," in *IEEE Workshop on*, (Cambridge, Massachusetts), pp. 26–33, 1995.

[55] e. A. Blonde, L., "A virtual studio for live broadcasting: The mona lisa project," *IEEE Multimedia*, vol. 3, no. 2, pp. 18–29, 1996.

[56] J. L. Marroquín. PhD thesis, MIT, 1985.

[57] K. Abend, *Pattern Recognition*, pp. 207–249. Thompson Book Co., 1968.

[58] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions and the bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 6, pp. 721–741, 1984.

[59] J. Marroquín, S. Botello, F. Calderón, and B. Vemuri, "The MPM-MAP algorithm for image segmentation," in *Proc. 15th Int. Conf. In Pattern Recognition ICPR-2000* (I. C. Soc., ed.), (Barcelona, Spain), pp. 303–308, IEEE Comp. Soc., 2000.

[60] M. Nielsen, L. Florack, and P. Deriche, "Regularization and scale space," Tech. Rep. 2532, INRIA, Sep. 1994.

[61] B. M. T. H. Romeny, ed., *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.