3

Markov Random Fields in the Context of Stereo Vision

Lorenzo J. Tardón[1], Isabel Barbancho[1] and Carlos Alberola[2]
[1]*Dept. Ingeniería de Comunicaciones, ETSI Telecomunicación-University of Malaga*
[2]*Dept. Teoría de la Señal y Comunicaciones e Ingeniería Telemática, ETSI Telecomunicación-University of Valladolid*
Spain

1. Introduction

The term *stereo vision* refers to the ability of an observer (either a human or a machine) to recover the three-dimensional information of a scene by means of (at least) two images taken from different viewpoints. Under the scope of this problem—and provided that cameras are calibrated—two subproblems are typically considered, namely, the correspondence problem, and the reconstruction problem (Trucco & Verri, 1998). The former refers to the search for points in the two images that are projections of the same physical point in space. Since the images are taken from different viewpoints, every point in the scene will project onto different image points, i.e, onto points with different coordinates in every image coordinate system. It is precisely this *disparity* in the location of image points that gives the information needed to reconstruct the point position in space. The second problem, i.e., the reconstruction problem, deals with calculating the disparity between a set of corresponding points in the two images to create a disparity map, and to convert this into a three-dimensional map.

In this context, we will show how Markov Random Fields (MRFs) can be effectively used. It is well known that MRFs constitute a powerful tool to incorporate spatial local interactions in a global context (Geman & Geman, 1984). So, in this chapter, we will consider local interactions that define proper MRFs to develop a model that can be applied in the process of recovery of the 3D structure of the real world using stereo pairs of images.

To this end, we will briefly describe the whole stereo reconstruction process (Fig. 1), including the process of selection of features, some important aspects regarding the calibration of the camera system and related geometric transformations of the images and, finally, probabilistic analyses usable in the definition of MRFs to solve the correspondence problem.

In the model to describe, both a priori and a posteriori probabilities will be separately considered and derived making use of reasonable selections of the potentials (Winkler, 1995) that define the MRFs on the basis of specific analytic models.

In the next section, a general overview of a stereo system will be shown. In Sec. 3, a brief overview of some well known stereo correspondence algorithms is given. Sec. 4 describes the main stages of a stereo correspondence system in which MRFS can be applied. Sec. 5 describes the camera model that will be considered in this chapter together with some important related issues like: camera calibration, the epipolar constraint and image rectification. Sec. 6 describes the concept of Markov random fields, and related procedures, like simulated annealing. Sec. 7 introduces MRFs for the edge detection problem. Sec. 8 describes, in detail, how MRFs can

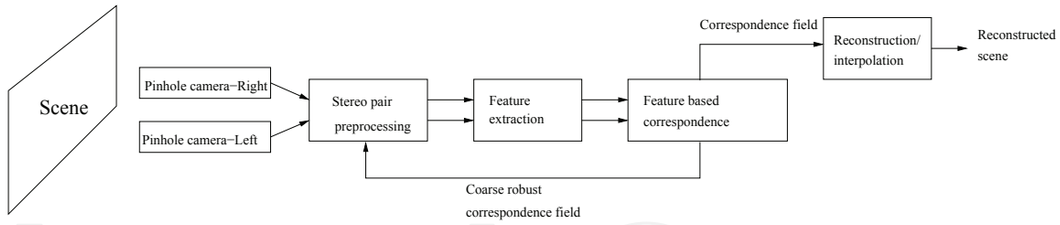


Fig. 1. Scheme of a stereo image reconstruction system.

be modeled using probabilistic analyses in the stereo correspondence context. Sec. 9 describes the implementation of the MRF based stereo system from the point of view of object models. Sec. 10 presents some illustrative experiments done with the MRF described. Finally, Sec. 11 draws some conclusions.

2. Processing stages in three-dimensional stereo

Now, we will briefly describe the main stages of a stereo system (Fig. 1).

Preprocessing: the images acquired by the camera system may require the application of some techniques to allow the reconstruction of three-dimensional scenes and/or to improve the performance of other stages. These techniques refer to many different aspects related to low level vision like: noise reduction, image enhancement, edge sharpening or geometrical transformations.

Feature extraction: this stage is required by feature based stereo systems, like the approach we will present. So, we will briefly introduce MRFs for the detection of edge pixels.

Matching: this stage refers to the process of resolution of the correspondence problem of the selected features. This stage will make use of a MRF defined upon specific probabilistic models.

Reconstruction/interpolation: after the correspondence problem is solved, the 3D scene can be reconstructed using information of the setup of the camera system and interpolating (if necessary) matched points or features.

Regarding these stages, stereo matching is often considered the most important and most difficult problem to solve. So, now, we are going to briefly overview some main ideas on solving the correspondence problem.

3. Solving the correspondence problem

The correspondence problem in stereo vision refers to the search for points in the two images that are projections of the same physical point in space (Trucco & Verri, 1998).
Correspondence methods can be broadly classified within two categories (Brown et al., 2003), namely, *local* and *global* methods. Local methods find the correspondence of a pixel using solely local information about that pixel. They can be very efficient, but also highly sensitive to local ambiguities. On the other hand, global methods provide global constraints on the image that may resolve these ambiguities, at the expense of a higher computational load. However, the following classification: area-based and feature-based methods is also widely used and accepted.
Area-based methods establish the correspondence mainly on the basis of the cross-correlation of image patches from each of the two images of a stereo pair. These techniques allow to obtain

dense disparity maps, but these are rather sensitive to noise and to perspective distortions although their efficiency matching images that contain natural elements solely.
Feature-based methods use specific similarity measures between pairs of selected features together with local and global restrictions regarding the disparity maps to obtain. These methods are often more robust but more difficult to implement and with higher computational burden. Regarding the features to match, it has been observed that edges are very important for the human visual system which makes these elements to be the most widely used features employed in stereo matching algorithms (observe the Fig. 9 b) which contains only detected edges. In this figure, the face of a woman is easily recognized).
A very short review of some main correspondence methods is given below.

3.1 Area-based methods

In (Cochran & Medioni, 1992), a deterministic and robust area-based correspondence method is proposed that used three levels of resolution to obtain dense disparity maps. The method defined is used in each of the three levels of resolution considered. The resolution levels are defined performing a Gaussian filtering and subsampling. The algorithm starts by cutting the image so that the disparity is zero at a fixed point and performing an epipolar alineation process. Then an area-based matching process starts which provides correspondence using a local measure of texture
Lane, Thacker and Seed (Lane et al., 1994) rely on the search of maxima of the correlation cross the pixel blocks previously deformed and the application of global constraints to eliminate the ambiguity due to the search of local maxima. The algorithm starts by aligning and correcting images according to the epipolar constraint.
Kanade (Kanade & Okutomi, 1994) proposed a model of the statistical distribution of the disparity at a point about the center. Such distribution is assumed to be Gaussian with variance proportional to the distance between the points.
Nishihara (Faugeras, 1993, sec. 6.4.2) proposed an improvement of area-based techniques introducing the use of sign of the Laplacian of Gaussian to reduce the sensitive to noise.

3.2 Feature-based methods

Pollard, Mayhew and Frisby (Pollard et al., 1985), (Pollard et al., 1986) proposed an algorithm to solve the problem of correspondence on the basis of the limits of the disparity gradient, derived from experiments performed on the human visual system's (HVS) ability to fuse stereograms.
According to their approach, the cyclopean separation is defined on the cyclopean image as (Fig: 2):

$$S = \sqrt{\left(\frac{x + x'}{2}\right)^2 + y^2} \tag{1}$$

and the disparity gradient is:

$$DG = \frac{|x' - x|}{S} \tag{2}$$

It is checked that a disparity gradient of 1 approximates the limit found for the human visual system (although it is also observed that when the matching dots are nearer the cameras, then it is more unlikely that this condition is maintained (Pollard et al., 1985)).

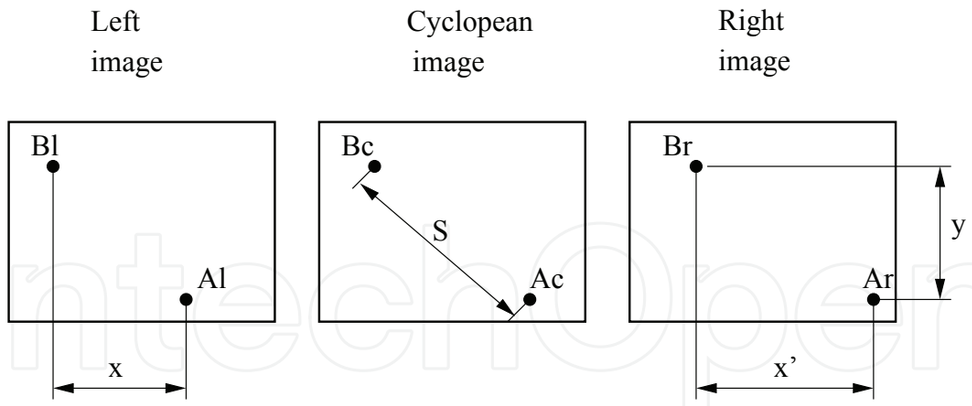


Fig. 2. Projections of the points A and B on the left and right image planes of a stereo system. Cyclopean image and cyclopean separation.

The disparity gradient is a main concept that will be used in the definition of the MRFs involved to solve the correspondence problem.
Barnard and Thompson (Barnard & Thompson, 1980) select the points to match using the Moravec operator (Moravec, 1977), which calculates the sum of squared differences of the intensity of adjacent pixels in the four directions at each position in windows of size 5×5 pixels; the minimum of these measures is stored. Then local maxima are found.
Ohta and Kanade (Ohta & Kanade, 1985) introduced a method based on dynamic programming to obtain optimal correlation paths between pairs of selected features.
On the basis of computational and psychophysical studies, Marr and Poggio (Grimson, 1985, sec. II), (Faugeras, 1993, sec. 6.5.1) develop a correspondence technique according to a hierarchical strategy to match zero crossings of the result of the application of the Lapacian of Gaussian filter to the images. Then, the continuity of the surfaces is imposed to solve the ambiguity the matchings. The matching process is repeated at different resolutions.
Marapane and Trivedi (Marapane & Trivedi, 1989), (Marapane & Trivedi, 1994) propose a hierarchical method in which at each stage of the correspondence process correspondence the most appropriate features should be used. Three main stages are considered to match: regions, line segments and edge pixels.

4. MRFs in a stereo correspondence system

Now, we describe the general stereo matching process. Note that MRFs can be used in two main stages: selection of features to match and resolution of the correspondence process. However, in this chapter, we pay special attention to the correspondence problem. The features that will be matched are edge pixels. Also, our process is supported by the calibration and the rectification processes. The complete scheme, with indication of the stages in which MRFs can be applied, is shown in Fig. 3.
Two stages are made to establish the correspondence in static scenes: the first one is used to rectify the images to apply the epipolar restriction (Faugeras, 1993) to help to simplify and to reduce the computational burden of the process of establishing true correspondences. The second one corresponds to the final stereo matching process.
The process of detection of edges will use the Nalwa-Binford (Nalwa & Binford, 1986) edge detector, but MRFs can also be defined to solve this stage (Tardón et al., 2006). Only edge pixels will be considered as features.

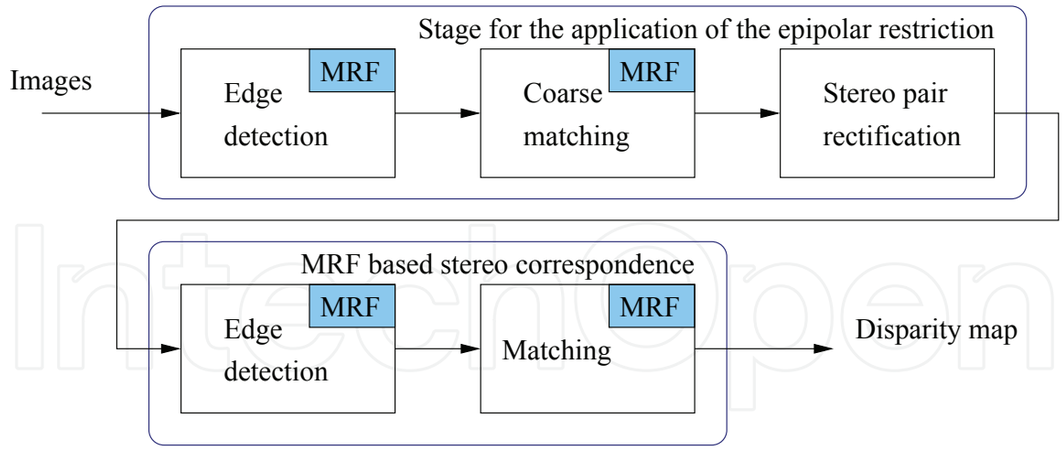


Fig. 3. An example of suitable application of MRFs in a stereo correspondence system.

After the edges are extracted, an initial matching stage is performed before image rectification:

– Area-based matching, using the normalized cross-covariance, is performed.
– Then, our iterative matching algorithm is employed to increase the reliability of the previous result by eliminating inconsistencies between correspondences and establishing new robust correspondences.

After obtaining a first map of correspondences, these are used to estimate the fundamental matrix (Mohr & Triggs, 1996). Once the fundamental matrix is estimated, the process of image rectification is done to easily apply the epipolar restriction.
Now, using the images properly rectified, the edges will be found again and then, the final matching process starts. Area-based matching using the normalized cross-covariance is employed and, then, the full MRF model is used to obtain the final disparity map for the selected features because of the modeling capability of MRFs (Li, 2001; Winkler, 1995) and their robust optimization capabilities (Boykov et al., 2001; Geman & Geman, 1984).
To begin with, we will make a description of MRFs and related concepts. Then, we must describe the camera model and the geometrical relations involved in the camera system considered because of their influence in the probabilities that help to define the MRFs involved in the formulation of the correspondence problem. Afterwards, we will describe the different stages in our stereo correspondence system.

5. Model of the binocular camera system

In this chapter, we consider a binocular system. One of the important factors involved is the main geometry of the system regarding the orientation of the optical axes. If the optical axes are parallel, then there exists a simple relation in the disparity (difference in the coordinates in the different images) between matching points (Barnard & Fischler, 1982) and depth (Bensrhair et al., 1992). This is a convenient case and it is usable in multitude of real cases.
The behavior of the cameras of the system must be described. We will consider the *pinhole* camera model (Faugeras, 1993, cap. 3), (Foley et al., 1992, cap. 6) (Fig. 4). Then a number of transformations expressed in homogeneous coordinates can be used to describe the relations between the real world coordinates and the image coordinate systems (Faugeras, 1993, cap. 3), (Foley et al., 1992, cap. 6), (Duda & Hart, 1973, cap. 10).

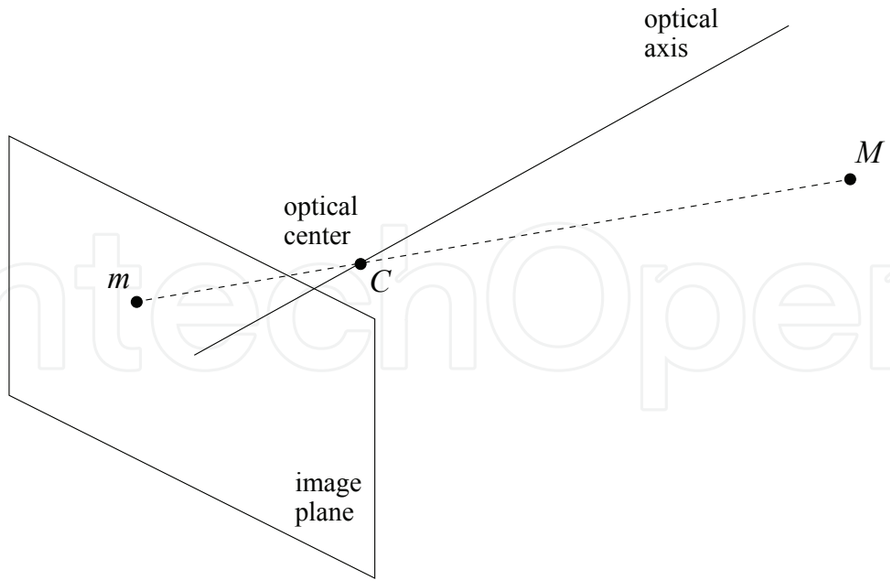


Fig. 4. The pinhole camera model.

According to the pinhole model, the camera is represented by a small point (hole), the optical center C, and an image plane at a distance F behind the hole (Duda & Hart, 1973) (Fig. 4). This model has a small drawback which is to reverse the images, so it is common to replace it by an equivalent one in which the optical center C is located behind the image plane. Then, the orthogonal projection that passes through the optical center is called the optical axis.
Homogeneous coordinates are suitable to describe the projection process in this model (Vince, 1995). First, consider the center of coordinates of the real world at the optical center and the following axes: Z orthogonal to the image plane and the axes X and Y orthogonal and, also orthogonal to Z. The origin of coordinates in the image plane will be the intersection of the Z axes with this plane and the axes u and v in the image plane will be orthogonal to each other and parallel to X and Y, respectively, then, the projected coordinates in the image plane $[U,V,S]^T$ of a point at $[x,y,z,1]^T$ will be given by (Faugeras, 1993, cap. 3):

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \qquad \vec{m} = P_0 \vec{M} \tag{3}$$

Now, we must also take into account all the possible transformations that can happen between the coordinates of a point in the space and a projection in the image plane. Consider a modification of the coordinates system in the image plane: a scaling of the axes and a translation. These operations, in the 2D space of projections, can be represented by:

$$H = \begin{bmatrix} k_u & 0 & t_u \\ 0 & k_v & t_v \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

so that we can obtain a new matrix $P_1 = H * P_0$ that takes into account these transformations. The parameters $\alpha_u = -fk_u$, $\alpha_v = -fk_v$, t_u y t_v are called the *intrinsic parameters* and they depend only on the camera itself.

Of course, we will probably desire to modify the usable coordinates system in the real world. Often, a rotation and a translation of the coordinates system is considered (Faugeras, 1993, sec. 3.3.2). These operations can be represented by the 4×4 matrix:

$$K = \begin{bmatrix} & R & & T \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

This matrix describes the position and the orientation of the camera with respect to the reference system and it defines the *extrinsic parameters*.
With all this, the projection matrix becomes:

$$P = P_1 * K = H * P_0 * K = \begin{bmatrix} \alpha_u \vec{r}_1 + t_u \vec{r}_3 & \alpha_u t_x + t_u t_z \\ \alpha_v \vec{r}_2 + t_v \vec{r}_3 & \alpha_v t_y + t_v t_z \\ \vec{r}_3 & t_z \end{bmatrix} = \begin{bmatrix} \vec{q}_1^T & q_{14} \\ \vec{q}_2^T & q_{24} \\ \vec{q}_3^T & q_{34} \end{bmatrix} \tag{6}$$

Note that only 10 parameters in the matrix are independent: scaling in the image plane (2 parameters), translation in the image plane (2), rotation in the real world (3) and translation in the real world (3). So, a valid projection matrix must satisfy certain conditions:

$$||\vec{q}_3|| = 1 \tag{7}$$

$$(\vec{q}_1 \wedge \vec{q}_3) \cdot (\vec{q}_2 \wedge \vec{q}_3) = 0 \tag{8}$$

The estimation of the projection matrix P can be done on the basis of the original equation that relates the coordinates of a point in the real world and the coordinates of its projection in the image plane:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = P \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{9}$$

with $u = \frac{U}{S}$ y $v = \frac{V}{S}$. Then, for each projected point two equation will be found (Faugeras, 1993, sec. 3.4.1.2):

$$\vec{q}_1^T \vec{C} - u \vec{q}_3^T \vec{C} + q_{14} - u q_{34} = 0 \tag{10}$$

$$\vec{q}_2^T \vec{C} - u \vec{q}_3^T \vec{C} + q_{24} - u q_{34} = 0 \tag{11}$$

where $\vec{C} = (x, y, z, 1)^T$. So, if N point are used in the calibration process, then $2N$ equation will be found. The set of equation can be compactly written $A\vec{q} = \vec{0}$ and restrictions, (7) and (8), in order to find a proper solution.
It is possible to fix one of the parameters (i.e. $q_{34} = 1$) and then, the modified system, $A'\vec{q'} = \vec{b}$, can be solved in terms of the minimum square error, for example. Afterward, the condition in (7) can be applied. With this idea, the result will be a valid projection matrix in our context, although its structure will not follow the one in (6), so, extrinsic and intrinsic parameters cannot be properly extracted.
A different option is to impose the condition $||\vec{q}_3|| = 1$. Then it will be possible to perform a minimization of $||A\vec{q}||$ as described in (Faugeras, 1993, Appendix. A).

5.1 The epipolar constraint

The epipolar constraint helps to convert the 2D search for correspondences in a 1D search since this constraint establishes the following: the images of a stereo pair are formed by pairs of lines, called *epipolar lines*, such that points in a given epipolar line in one of the images will find their matching point in the corresponding epipolar line in the other image of the pair.
First, we define the *epipolar planes* as the planes that pass through the optical centers of the two cameras and any point in the space. The intersections of these planes with the image planes define the pairs of epipolar lines (Fig. 5).
Pairs of epipolar lines can be found using the projection matrices of a stereo camera system (Faugeras, 1993, cap. 6). To describe the process, we write, now, the projection matrices as: $T = \begin{bmatrix} T_1^T \\ T_2^T \\ T_3^T \end{bmatrix}$, and let $\vec{M}$ denote a point. Then $T_3^T \vec{M} = 0$ represents a plane that is parallel to the image plane that contains the optical center ($T_3^T \vec{M} = 0 \rightarrow p_w = 0 \rightarrow \frac{p_x}{p_w} = \infty$, $\frac{p_y}{p_w} = \infty$). if, in addition to this, $T_2^T \vec{M} = 0$ ($\rightarrow p_y = 0$) and $T_1^T \vec{M} = 0$ ($\rightarrow p_x = 0$), we find the equation of two other planes that contain the optical center. The intersection of these three planes is the center of projection in global coordinates:

$$T\vec{C} = \begin{bmatrix} \vec{T}_1^T \\ \vec{T}_2^T \\ \vec{T}_3^T \end{bmatrix} \vec{C} = \begin{bmatrix} \vec{q}_1^T & q_{14} \\ \vec{q}_2^T & q_{24} \\ \vec{q}_3^T & q_{34} \end{bmatrix} \vec{C} = \vec{0} \tag{12}$$

The projection equation can be written as:

$$\begin{bmatrix} \vec{q}_1^T \\ \vec{q}_2^T \\ \vec{q}_3^T \end{bmatrix} \vec{O} = - \begin{bmatrix} q_{14} \\ q_{24} \\ q_{34} \end{bmatrix} \qquad \rightarrow \qquad \vec{O} = - \begin{bmatrix} \vec{q}_1^T \\ \vec{q}_2^T \\ \vec{q}_3^T \end{bmatrix}^{-1} \cdot \begin{bmatrix} q_{14} \\ q_{24} \\ q_{34} \end{bmatrix} \tag{13}$$

with $\vec{O} = (o_x, o_y, o_z)^T$.
Using the optical center, the *epipoles* E_1 y E_2 can be found. An epipole is the projection of and optical center in the opposite image plane. Then, the epipolar lines can be easily defined since

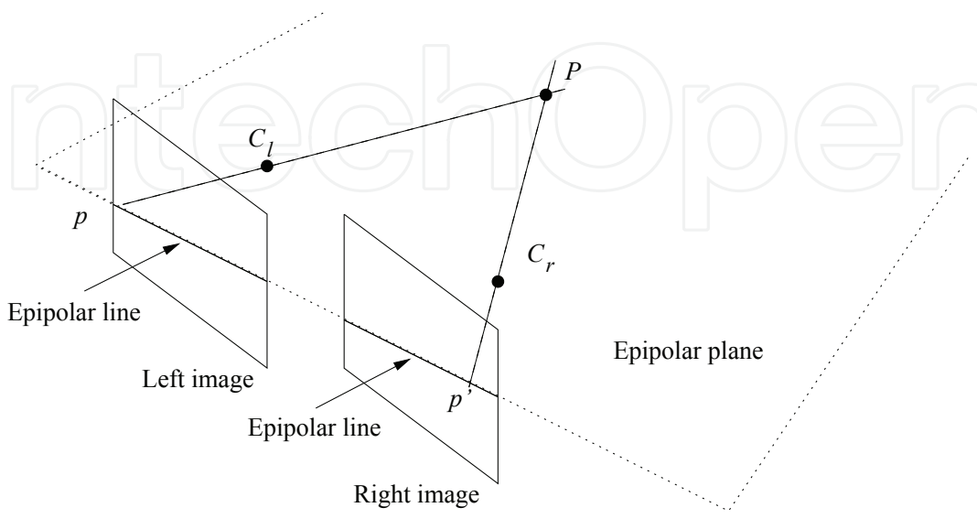
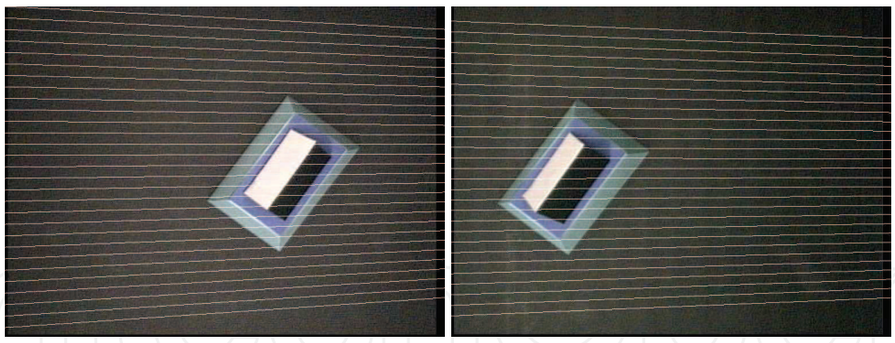


Fig. 5. Epipolar lines and planes.



(a) (b)

Fig. 6. Left a) and right b) images of a stereo pair with superimposed epipolar lines obtained with the calibration matrices using homogeneous coordinates.

they all contain the respective epipole. Fig. 6 shows an example of application of the epipolar constraint derived from the calibration matrices of a binocular stereo setup.
Note that it is also possible to find the the relation that defines the epipolar constraint without the projection matrices (Trivedi, 1986). To this end, we will pay attention to the fundamental matrix.

5.1.1 The fundamental matrix

Since the epipolar lines are the projection of a single plane in the image planes, then there exists a projective transformation that transforms an epipolar line in an image of a stereo pair into the corresponding epipolar line in the other image of the pair. This transformation is defined by the *fundamental matrix*.
Let $\vec{l}$ and $\vec{l}'$ denote two corresponding epipolar lines in the two images of a stereo pair. The transformation between these two lines is a collineation: a projective transformation of the projective space that $\mathcal{P}^n$ into the same projective space (Mohr & Triggs, 1996). Collineations in the projective space are represented by 3×3 non-singular matrices. So, let A represent a collineation, then $\vec{l}' = A\vec{l}$.
Let $\vec{m} = [x, y, t]^t$ represent a point in the first image of the stereo pair and let $\vec{e} = [u, v, w]^t$ represent the epipole in the first image. Then, the epipolar line through $\vec{m}$ y $\vec{e}$ is given by $\vec{l} = [a, b, c]^t = \vec{m} \times \vec{e}$ (Mohr & Triggs, 1996, sec. 2.2.1). This is a linear transform that can be represented as:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} yw - tv \\ tu - xw \\ xv - yu \end{bmatrix} = \begin{bmatrix} 0 & w & -v \\ -w & 0 & u \\ v & -u & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ t \end{bmatrix}; \ \vec{l} = C\vec{m} \tag{14}$$

where C is a matrix with rank 2.
Then, we can write $\vec{l}' = AC\vec{m} = F\vec{m}$. Since this expression is accomplished by all the points in the line l', we can write:

$$\vec{m}'^t F \vec{m} = 0 \tag{15}$$

where F is 3×3 matrix with rank 2, called the fundamental matrix:

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \tag{16}$$

Now, these relation must be estimated to simplify the correspondence problem. Linear and nonlinear techniques are available to this end (Luong & Faugeras, 1996). We will give a short discussion on the most frequently used procedures.

5.1.1.1 Estimation of the fundamental matrix

In the work by Xie and Yuan Li (Xie & Liu, 1995), it is considered that since the matrix F defines an application between projective spaces, than, any matrix $F' = kF$, where k is a scalar, defines the same transformation. Specifically, if an element F_{ij} of F is nonzero, say f_{33}, we can define $H = \frac{1}{f_{33}} F$, so that $\vec{m}' H \vec{m} = 0$, with

$$H = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \tag{17}$$

The transformation represented by this equation is called generalized epipolar geometry and, since no additional constraints are imposed on the rank of F, the coefficients of the matrix can be easily estimated using sets of known matching point using a conventional least squares technique.

Mohr and Triggs (Mohr & Triggs, 1996) propose a more elaborate solution since the rank of the matrix is considered. Since, for each pair of matching points, we can write $\vec{m}' F \vec{m} = 0$, then for each pair, we can write the following equation:

$$x x' f_{1,1} + x y' f_{1,2} + x f_{1,3} + y x' f_{2,1} + y y' f_{2,2} + y f_{2,3} + x' f_{3,1} + y' f_{3,2} + f_{3,3} = 0 \tag{18}$$

The set of all the available equation can be written $D\vec{f} = 0$, where $\vec{f}$ is a vector that contains the 9 coefficients in F. The first constraint that can be imposed is that the solution have unity norm and, if more than 8 pairs of matching points are available, then, we can find the solution in the sense of minimum squares:

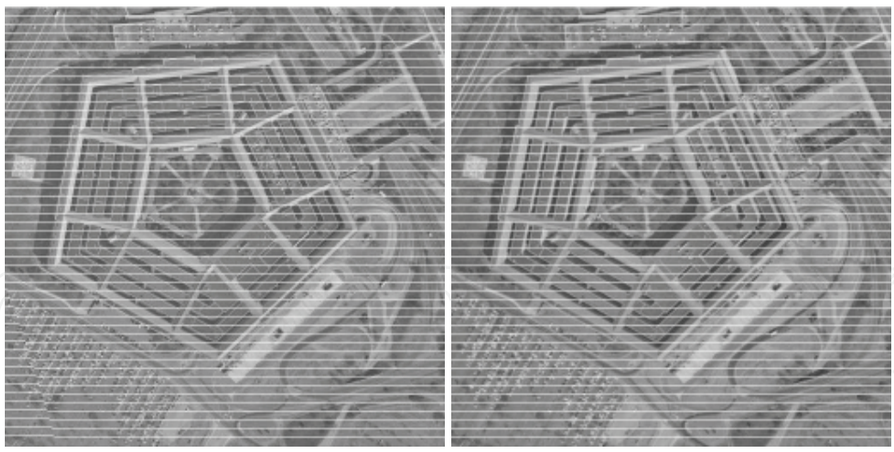
$$\min_{||\vec{f}||=1} ||D\vec{f}||^2 \tag{19}$$

which is equivalent to finding the eigenvector of the smallest eigenvalue in $D^t D$. The technique is similar to the one presented by Zhengyou Zhang in (Zhang, 1996, sec. 3.2). A different strategy is also shown in (Zhang, 1996, sec. 3.4), on the basis of the definition of proper error measures in the calculation of the fundamental matrix. Regardless of the technique employed, note that the process of estimation of the fundamental matrix is always very sensitive to noise

After the epipolar constraint is defined between the pairs of images, a geometrical transformation of the image is performed so that the corresponding epipolar lines will be horizontal and with the same vertical coordinate in both images.

Fig. 7 shows an example with selected epipolar lines, obtained using the fundamental matrix, superimposed on the images of a stereo pair.

Note that, in order to obtain reliable matching points to estimate the fundamental matrix, matching points should be well distributed over the entire image. In this example, we have



(a) (b)

Fig. 7. *Pentagon* stereo pair with superimposed epipolar lines. a) Left image. b) Right image.

used a set of the most probably correct matching points (about 200 points) obtained using the iterative Markovian algorithm that will be described.

5.2 Geometric correction of the images according to the epipolar constraint

Now, corrected pairs of images will be generated so that their corresponding epipolar lines will be horizontal and with the same vertical coordinate in both images to simplify the process of establishment of the correspondence. The process applied is the following:

- A list of vertical positions for the original images of the epipolar lines at the borders of the images will be generated.
- The epipolar lines will be redrawn in horizontal and the intensity values at the new pixel position of the rectified images will be obtained using a parametric bicubic model of the intensity surfaces (Foley et al., 1992), (Tardón, 1999).

6. Markov random fields

The formulation of MRFs in the context of stereo vision considers the existence of a set of irregularly distributed points or positions in an image, called (*nodes*) which are the image elements that will be matched. The set of possible correspondences of each node (*labels*) will be a discrete set selected from the image features extracted from the other image of the stereo pair, according to the disparity range allowed.
Our formulation of MRFs follows the one given by Besag (Besag, 1974). Note that the matching of a node will depend only on the matching of other nearby nodes called neighbors. The model will be supported by the Bayesian theory to incorporate levels of knowledge to the formulation:

- A priori knowledge: conditions that a set of related matchings must fulfill because of inherent restrictions that must be accomplished by the disparity maps.
- A posteriori knowledge: conditions imposed by the characterization of the matching of each node to each label.

Using this information in this context, restrictions are not imposed strictly, but in a probabilistic manner. So, correspondences will be characterized by a function that indicates

a probability that each matching is correct or not. Then, the solution of the problem requires the maximization of a complex function defined in a finite but large space of solutions. The problem is faced by dividing it into smaller problems that can be more easily handled, the solutions of which can be mixed to give rise to the global solution, according to the MRF model.

6.1 Random fields

We will introduce in this section the concept of random field and some related notation. Let S denote all positions where data can be observed (Winkler, 1995). These positions define a graph in $\mathcal{R}^2$, where each position can be denoted $s \in S$. Each position can be in state x_s in a finite space of possible states X_s. We will call *node* each of the objects or primitives that occupy a position: a selected pixel to be matched will be a node. In the space of possible configurations of X ($\Pi_{s \in S} X_s$), we can consider the probabilities $P(x)$ con $x \in X$. Then, a strictly positive probability measure in X defines a *random field*.

Let A a subset in S (A *subset*S) and X_A the set of possible configurations of the nodes that belong to A (x_A *in*X_A). Let $\bar{A}$ stand for the set of all nodes in S that do not belong to A. Then, it is possible to define the conditional probabilities $P(X_A = x_A / X_{\bar{A}} = x_{\bar{A}})$ that will be usually called *local characteristics*. These local characteristics can be handled with a reasonable computational burden, unlike the probability measures of the complete MRF.

The nodes that affect the definition of the local probabilities of another node s are called the neighborhood $\mathtt{V}(s)$. These are defined with the following condition: if node t is a neighbor of s, then s is a neighbor of t. *Clique* is another related and important concept: a set of nodes in S ($C \subset S$) is a clique in a MRF if all the possible pairs of nodes in a clique are neighbors.

With all this, we can define a Markov random field with respect to a neighborhood system $\mathtt{V}$ as a random field such that for each $A \subset S$:

$$P(X_A = x_A / X_{\bar{A}} = x_{\bar{A}}) = P(X_A = x_A / X_{\mathtt{V}(A)} = x_{\mathtt{V}(A)}) \tag{20}$$

Observe that any random field in which local characteristics can be defined in this way, is a random field and that positivity condition makes $P(X_A = x_A / X_{\bar{A}} = x_{\bar{A}})$ to be strictly positive.

6.2 Markov random fields and Markov chains

Now, more details on MRFs from a generic point of view will be given. Let $\Lambda = \{\lambda_{\mathtt{p}}, \lambda_{\mathtt{q}}, \ldots\}$ denote the set of nodes in which a MRF is defined. The set of locations in which the MRF is defined will be $\mathcal{P} = \{\mathtt{p}, \mathtt{q}, \mathtt{r}, \ldots\}$, which is very often related to rectangular structures, but this is not a requirement (Besag, 1974), (Kinderman & Snell, 1980). Let $\Delta = \{\delta_1, \delta_2, \ldots\}$ denote the set of possible labels, and $\Delta_{\mathtt{p}} = \{\delta_i, \delta_j, \ldots\}$, the set of possible labels for node $\lambda_{\mathtt{p}}$.

The matching of a node to a label will be $\lambda_{\mathtt{i}} = \delta_j$, and the probability of the assignation of a label to a node at position $\mathtt{p}$ will be $P(\lambda_{\mathtt{p}} = \delta_{\mathtt{p}})$. Since we are dealing with a MRF, then the following positivity condition is fulfilled:

$$P(\Lambda = \Xi) > 0 \tag{21}$$

where Ξ represents the set of all the possible assignments.

If the neighborhood $\mathtt{V}$ is the set of nodes with influence on the conditional probability of the assignation of a label to a node among the set of possible labels for that node:

$$P(\lambda_{\mathtt{p}} = \delta_{\mathtt{p}} | \lambda_{\mathtt{q}} = \delta_{\mathtt{q}}, \mathtt{q} \neq \mathtt{p}) = P(\lambda_{\mathtt{p}} = \delta_{\mathtt{p}} | \lambda_{\mathtt{q}} = \delta_{\mathtt{q}}, \mathtt{q} \in \mathtt{V}_{\mathtt{p}}) \tag{22}$$

where $\mathtt{V}_{\mathtt{p}}$ is the neighborhood of $\mathtt{p}$ in the random field, then:

– The process is completely defined upon the conditional probabilities: *local characteristics*.
– If $\mathtt{V_p}$ is the neighborhood of the node at $\mathtt{p}$, $\forall\, \mathtt{p} \in \mathcal{P}$, then Λ is a MRF with respect to $\mathtt{V}$ if and only if $P(\Lambda = \Xi)$ is a Gibbs distribution with respect to the defined neighborhood (Geman & Geman, 1984).

We can write the conditional probability as:

$$P(\lambda_A = \delta_A | \lambda_{\bar{A}} = \delta_{\bar{A}}) = \frac{e^{-\sum_{c \in \mathcal{C}_1} \mathtt{U}_c(\delta_{A_v})}}{\sum_{\gamma_A \in \Delta_A} e^{-\sum_{c \in \mathcal{C}_1} \mathtt{U}_c(\gamma_A, \delta_{V(A)})}} \tag{23}$$

This is a key result and some considerations must be done about it:

– Local and global Markovian properties are equivalent.
– Any MRF can be specified using the local characteristic. More specifically, these can be described using: $P(\lambda_{\mathtt{p}} = \delta_{\mathtt{p}} /\ \lambda_{\bar{p}} = \delta_{\bar{p}})$.
– $P(\lambda_{\mathtt{p}} = \delta_{\mathtt{p}} / \lambda_{\bar{p}} = \delta_{\bar{p}}) > 0$, $\forall\, \delta_{\mathtt{p}} \in \Delta_{\mathtt{p}}$, according to the positivity condition

Regarding neighborhoods, these are easily defined in regular lattices using the *order* of the field (Cohen & Cooper, 1987). In other structures, the concept of order can not be used, then the neighborhoods must be specially defined, for example, using a measure of the distance between the nodes.
The concept of clique is of main importance. According to its definition: if $C(\mathtt{t})$ is a clique in a certain neighborhood of $\lambda_{\mathtt{t}}$, $\mathtt{V_p}$, then if $\lambda_{\mathtt{o}}$, $\lambda_{\mathtt{p}}$,..., $\lambda_{\mathtt{r}} \in C(\mathtt{t})$, then $\lambda_{\mathtt{o}}$, $\lambda_{\mathtt{p}}$,..., $\lambda_{\mathtt{r}} \in \mathtt{V_s}$ $\forall\, \lambda_{\mathtt{s}} \in C(\mathtt{t})$. Note that a clique can contain zero nodes.
It is rather simple to define cliques in rectangular lattices (Cohen & Cooper, 1987), but is is a more complex task in arbitrary graphs and the condition of clique should be check for every clique defined. However, it can be easily observed that the cliques formed by up to two neighboring nodes are always correctly defined, so, since there is no reason that imposes us to define more complex cliques, we will use cliques with up to two nodes.
Regarding the local characteristic, it can be defined using information coming from two different sources: a priori knowledge about how the correspondence fields should be and a posteriori knowledge regarding the observations (characterization of the features to match). These two sources of information can be mixed up using the Bayes theorem which establishes the following relation:

$$P(x/\hat{y}) = \frac{P(x)P(\hat{y}/x)}{\sum_z P(z)P(\hat{y}/z)} \tag{24}$$

– $P(x)$: a priori probability of the correspondence fields.
– $P(\hat{y}/x)$ posterior probability of the observed data.
– $\sum_z P(z)P(\hat{y}/z) = P(\hat{y})$ represents the probability of the observed data. It is a constant.

6.2.1 A priori and posterior probabilities

The a priori probability density function (pdf) incorporates the knowledge of the field to estimate. This is a Gibbs function (Winkler, 1995) and, so, it is given by:

$$P(x) = \frac{e^{-H(x)}}{\sum_{x \in X} e^{-H(x)}} = \frac{1}{Z} e^{-H(x)} \tag{25}$$

where H is a real function:

$$\begin{array}{rccc} H: & X & \longrightarrow & \mathcal{R} \\ & x & \longrightarrow & H(x) \end{array} \tag{26}$$

Note that any strictly positive function in X can be written as a Gibbs function using:

$$H(x) = -\ln P(x) \tag{27}$$

The posterior probabilities must be strictly positive functions so that $P(\hat{y}/x)$ may follow the shape of a local characteristic of a MRF:

$$\exists\, G(\hat{y}/x)/G(\hat{y}/x) = -\ln P(\hat{y}/x) \tag{28}$$

6.3 Gibbs sampler and simulated annealing

Now, the problem that we must solve is that of generating Markov chains to update the configuration of the MRF in successive steps to estimate modes of the limit distributions (Winkler, 1995), (Tardón, 1999). This problem is addressed considering the *Gibbs sampler* with *simulated annealing* (Geman & Geman, 1984), (Winkler, 1995) to generate Markov chains defined by $P(y/x)$ using the local characteristic. The procedure is described in Table 1.

Note that there are no restrictions for the update strategy of the nodes, these can be chosen randomly. Also, the algorithm visits each node an infinite number of times. Note that the step Update Temperature T represents the modification of the original Gibbs sampler algorithm to give rise to the so-called *simulated annealing*. Recall that our objective is to estimate the modes of the limit distributions which are the *MAP* estimators of the MRF. Simulated annealing helps to find that state (Geman & Geman, 1984).

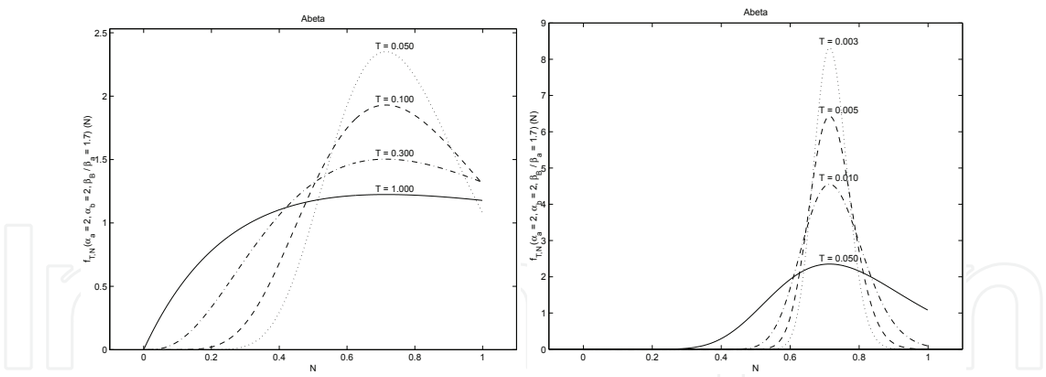
The main idea behind simulated annealing is now given. Consider a probability function $p(\psi) = \frac{1}{Z}e^{-H(\psi)}$ defined in $\psi \in \Psi$, where Ψ is a discrete and finite set of states. If the probability function is uniform, then any simulation of random variables that behaves according to that function will give any of the states, with the same probability as the other states. Instead, assume that $p(\psi)$ shows a maximum (mode). Then, the simulation will show that state with larger probability that the other states. Then, consider the following modification of the probability function in which the parameter *temperature* T is included:

$$p_T(\psi) = \frac{1}{Z_T}e^{-\frac{1}{T}H(\psi)} \tag{29}$$

This is the same function (a Gibbs function) as the original one when $T = 1$. If T is decreased towards zero, then $p_T(\psi)$ will have the same modes as the original one, but the difference in probability of the mode with respect to the other states will grow (see Fig. 8 as example).

A rigorous analysis of the behavior of the energy function H with T allows to determine the procedure to update the system temperature to guarantee the convergence, however, suboptimal simple temperature update procedures are often used (Winkler, 1995), (Tardón, 1999) (Sec. 9.2).

Now, simulated annealing can be applied to estimate the modes of the limit distributions of the Markov chains. According to our formulation, these modes will be to the MAP estimators of the correspondence map defined by the Markov random fields models we will describe.



(a) (b)

Fig. 8. Exaggeration of the modes of a probability function with decreasing temperature.

7. Using MRFs to find edges

Now, we are ready to consider the utilization of MRFs in a main stage of the stereo correspondence system. Since edges are known to constitute and important source of information for scene description, edges are used as feature to establish the correspondence.
As described in Tardón et al. (2006), MRFs can be used for edge detection. The likelihood can be based on the Holladay's principle (Boussaid et al., 1996) to relate the detection process to the ability of the human visual system (HVS) to detect edges. This information can be written in the form of suitable energy functions, $H(y/x)$ (here, x denotes the underlying edge field and y denotes the observation), that can be used to define MRFs.
Also, a priori knowledge about the expected behavior of the edges can be incorporated and expressed as an energy function, $H(x)$.
Then, using the Bayes rule, the posterior distribution of the MRF can be found:

$$p(x/y) \propto p(x)p(y/x) \tag{30}$$

and it will have the form of a Gibbs function. So, it will be possible to write the energy of the MRF as follows (Tardón et al., 2006):

START: *Iteration*
Update Temperature T
$\forall s_i \in S$
Select $s_i \in S_r$
START: *Comment*
s_i can be randomly selected from S_r.
$S_r \subset S$ is the subset of nodes in S that have not been yet updated in the present iteration.
END: *Comment*
Determine the local characteristic $P_{T,A_{s_i}}$
Randomly select the new state of s_i according to $P_{T,A_{s_i}}$
END: *Iteration*
GO TO: *Iteration*

Table 1. Gibbs sampler with simulated annealing.

(a) (b)

Fig. 9. a) Input image (*Lenna*). b) Edges detected using the MRF model in (Tardón et al., 2006).

$$H(x/y) = H(x) + H(y/x) \tag{31}$$

Fig. 9 shows an example of the performance of the algorithm. Simulated annealing is used (Sec. 6.3) with the following system temperature: $T = T_0 \cdot T_B^{k-1}$, where T_0 is the initial temperature, $T_B = 0.999$ and k stands for the iteration number. The number of iterations is 100. The parameter required by the algorithm is $C_w = 8$ (Tardón et al., 2006).
We have briefly introduced MRFS for the edge detection problem since MRFs are described in detail and they are used in the correspondence problem. However, the Nalwa-Binford edge detector Nalwa & Binford (1986) will be used in the stereo correspondence examples that will be shown in Sec. 10.

8. MRFs for stereo matching

In this section, we show how a Markovian model that makes use of an important psychovisual cue, the disparity gradient (DG) (Burt & Julesz, 1980), can be defined to help to solve the correspondence problem in stereo vision. We encode the behavior of the DG in a pdf to guide the definition of the energy function of the prior of a MRF for small baseline stereo. To complete the model based on a Bayesian approach, we also derive a likelihood function for the normalized cross-covariance (Kang et al., 1994) between any two matching points. Then, the correspondence problem is solved by finding the MAP solution using simulated annealing (Geman & Geman, 1984; Li et al., 1997) (Sec. 6.3).

8.1 Geometry of a stereo system for a MRF model of the correspondence problem

The setup of a stereo vision system is illustrated in Fig. 5. A point P in the space is projected onto the two image planes, giving rise to points p and p'. These two points are referred to as *matching* or *corresponding* points. Recall that these three points, together with the optical center of the two cameras, C_l and C_r, are constrained to lie on the same plane called the *epipolar* plane, and the line that joins p and p' is known as *epipolar* line.
As it has already been pointed, the DG is a main concept in stereo vision and for the correspondence problem (Burt & Julesz, 1980). Consider a pair of matching points $p \rightarrow p'$ and $q \rightarrow q'$. Their DG (δ) is defined by (Pollard et al., 1986):

$$\delta = \frac{\text{difference in disparity}}{\text{cyclopean separation}} = 2\frac{||(p' - q') - (p - q)||}{||(p' - q') + (p - q)||} \tag{32}$$

where the *cyclopean separation* represents the distance between the cyclopean image points ($\frac{p+p'}{2}$ and $\frac{q+q'}{2}$ as shown in figure 2) and the associated disparity vectors are $(p' - p)$ and $(q' - q)$.
Note that other constraints like surface continuity, figural continuity or uniqueness are subsumed by the DG (Faugeras, 1993), (Li & Hu, 1996).

8.2 Design of a MRF model for stereo matching

In this section, a methodology to design a MRF based on a Bayesian formulation on the basis of probabilistic analyses of the prior model of the expected correspondence maps and, also, on probabilistic analyses of the posterior information will be described (Tardón et al., 2006).

8.2.1 Neighborhood

The definition of the MRF requires the definition of the neighborhood system, so that each node, or feature for which a matching feature in the other image must be found, find some nearby nodes, neighbors, to define the local characteristic. In this case, a regular rectangular lattice can not be considered, and so, the concept of the order of the MRF can not be used to define neighbors or cliques.
We have decided to define a region around each node in which all the neighbors of the node can be found.
The neighborhood is defined upon the concept of superellipse (Fig. 10). This choice includes, in fact, different possibilities in the definition of the shape of the neighborhood. A superellipse with semi axes a and b and shape parameter p centered at the origin of the coordinate system is defined by:

$$\left(\frac{|x|}{a}\right)^p + \left(\frac{|y|}{b}\right)^p - 1 = 0 \tag{33}$$

with $a > 0$, $b > 0$ and $p > 0$.
Note that the structure of the neighborhood must be kept fixed along the image to guarantee the correct definition of the field in terms of neighbours and cliques.

8.2.2 Labels: sets of possible matchings

The region in which matching features for each node can be found is defined by superellipses, just like the neighborhoods. Labels are defined as the extracted features that can be found in the selected region of the other image of the stereo pair, plus the null-correspondence label (for the nodes that have no matching feature in the other image).
This search region is a superellipse (Fig. 10, eq. (33)) centered at the location point where we expect to find the correspondence of each node.
Note that if the images are correctly rectified, then the search region will become a segment in the corresponding epipolar line. This shape can, also, be easily described by the superellipse, with appropriate parameters.

8.3 A priori knowledge

Regarding a priori knowledge, the sources of information typically used in stereo matching are the maximum difference of disparity between two points (Barnard & Thompson, 1980),

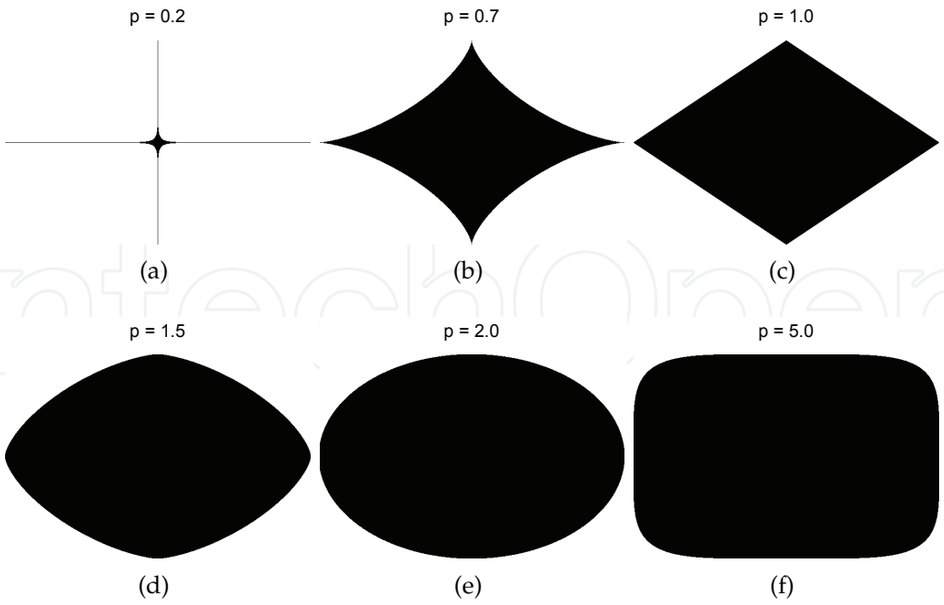


Fig. 10. Geometrical structures defined by the superellipse.

surface smoothness (Hoff & Ahuja, 1989), disparity continuity (Sherman & Peleg, 1990), ordering (Zhang & Gerbrands, 1995) and the disparity gradient *DG* (Olsen, 1990). However, the *DG* subsumes the rest of the constraints usually imposed for stereo matching (Li & Hu, 1996). Also, it is possible to obtain closed-from expressions of its probabilistic behavior under reasonable assumptions.

It has been demonstrated that the *DG* between two matching points should not be larger than 1 (Pollard et al., 1985), although this is a fuzzy limit, since it may vary slightly, depending on different factors (Wainman, 1997), (McKee & Verghese, 2002). Furthermore, in natural scenes, the *DG* between correct matches is usually small. We consider the limit of the *DG* as a soft threshold for the HVS, such that there should be a low probability that correct matches exceed this limit.

So, we define a MRF of matching points in which the information given by the *DG* is used to cope with the a priori knowledge (Tardón et al., 1999), (Tardón et al., 2004). To proceed with the design, notice that every match will be defined as the relationship between a selected feature in the left image (called *node*) and another feature in the right image (called *label*) (Fig. 11 and Sections 8.2.1 and 8.2.2).

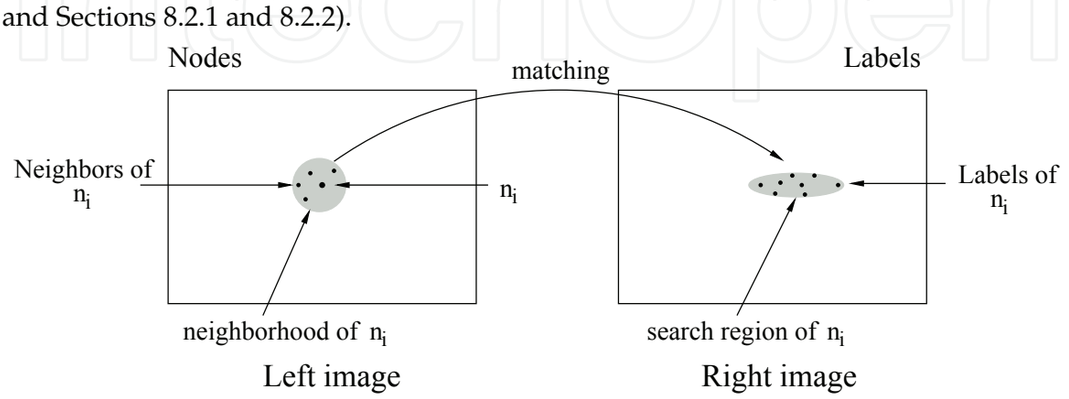


Fig. 11. Labels and nodes.

Consider a neighborhood system V for the set of sites S in the left image. Since the a priori knowledge will be based on the DG, which is defined for every pair of matching points, we will only use the set of binary cliques, $\mathcal{C}_b$, to build the a priori Gibbs function of the disparity map:

$$p_{S\Delta}(x) = \frac{1}{Z_x} e^{-H_{S\Delta}(x)} \tag{34}$$

where the energy function $H_{S\Delta}(x)$ consists of the potentials of the cliques in $\mathcal{C}_b$:

$$H_{S\Delta}(x) = \sum_{c \in \mathcal{C}_b} U_\Delta(\delta_c) \tag{35}$$

with δ_c the DG defined by the matches in the clique c. Note that when the DG is modeled as a random variable it will be denoted with the capital letter Δ, with δ a particular value of it. The same criterion will be used for other random variables in this section.
To derive the potential functions, consider, as an illustration, a node n_i that has a single neighbor n_j. Then a single clique $c_{i,j}$ contains the node n_i and the corresponding local characteristic will be (Winkler, 1995):

$$p(n_i = x_i / X_R = x_R, R = S - \{n_i\}) \propto p(n_i = x_i / X_{n_j} = x_{n_j}) \propto e^{-U_\Delta(\delta_{c_{i,j}})} \tag{36}$$

This function must be consistent with the behavior of the DG, so a natural choice for the potential functions is:

$$U_\Delta(\delta_c) \propto -\ln p(n_i = x_i / X_{n_j} = x_{n_j}) \propto -\ln f_\Delta(\delta) \tag{37}$$

In this way, the probabilistic behavior of the DG is easily accounted for in the prior. Recall that this is not an attempt to use the pdf of the DG to define the marginals of the MRF but to derive suitable potential functions using psycho-visual information. Now, we must derive the pdf the of the disparity gradient.

8.3.1 Pdf of the disparity gradient

Consider a simple geometry of parallel cameras of small aperture. Figure 12 shows a top view of the system with the Y axis protruding from the paper plane upwards; the terminology and the relationship between the parameters involved are described in figures 12 and 13.
The DG is defined upon the relationship between the projection of two points in 3D space, P and Q, the coordinates of which in the world reference system are given by the following relations:

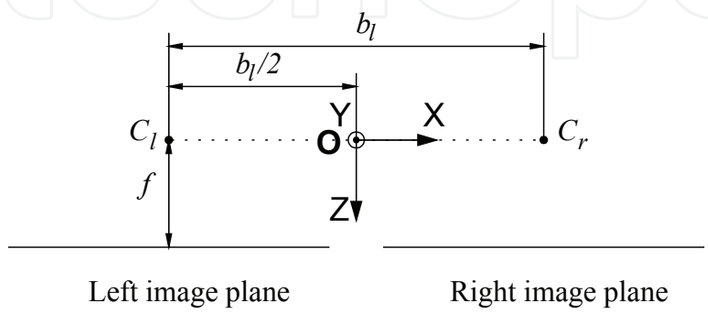


Fig. 12. Stereo system with parallel cameras of small aperture.

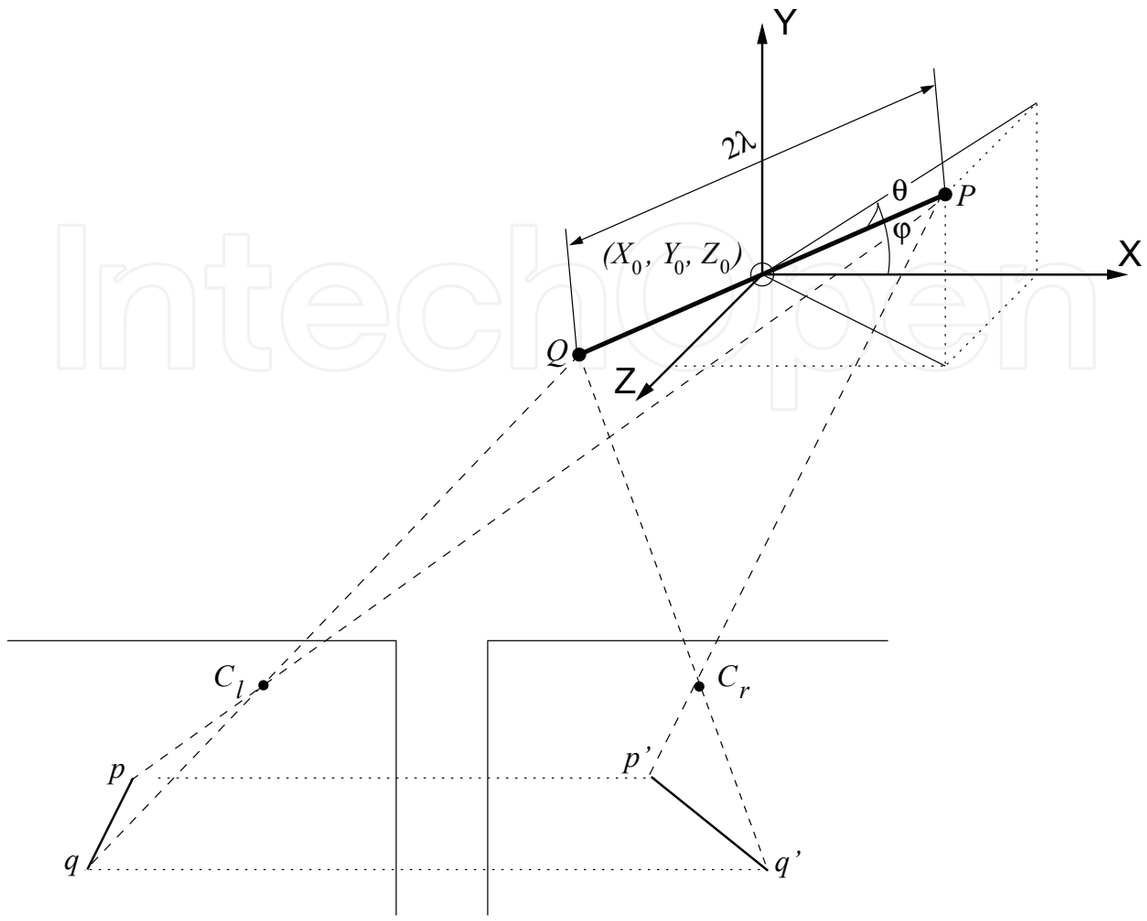


Fig. 13. Stereo system with parallel cameras of small aperture: projections and disparity gradient scenario.

$$P = (P_x, P_y, P_z) = (X_0 + \lambda\cos\theta\cos\psi, Y_0 + \lambda\cos\theta\sin\psi, Z_0 - \lambda\sin\theta) \tag{38}$$

$$Q = (Q_x, Q_y, Q_z) = (X_0 - \lambda\cos\theta\cos\psi, Y_0 - \lambda\cos\theta\sin\psi, Z_0 + \lambda\sin\theta) \tag{39}$$

where

– 2λ is an arbitrary distance that separates P and Q.

– (X_0, Y_0, Z_0) is a point which is equidistant between P and Q and belongs to the segment $\overline{PQ}$.

– ψ and θ are the angles that describe the orientation of $\overline{PQ}$.

Note that it is reasonable to model these variables, in the absence of any other type of knowledge, as independent uniform random variables, Ψ and Θ, in the intervals $(-\pi, \pi)$ and $(0, \pi)$, respectively(Law & Kelton, 1991).

The projections of P and Q on the left and right image planes are given by:

$$p = \left(-\frac{f}{P_z}(P_x + \frac{b_l}{2}), -\frac{f}{P_z}P_y \right) \tag{40}$$

$$q = \left(-\frac{f}{Q_z}(Q_x + \frac{b_l}{2}), -\frac{f}{Q_z}Q_y \right) \tag{41}$$

$$p' = \left(-\frac{f}{P_z}(P_x - \frac{b_l}{2}), -\frac{f}{P_z}P_y \right) \tag{42}$$

$$q' = \left(-\frac{f}{Q_z}(Q_x - \frac{b_l}{2}), -\frac{f}{Q_z}Q_y \right) \tag{43}$$

where b_l and f represent the baseline and the focal distance, respectively (Fig. 12).
Substituting equations (38)—(43) in equation (32) we obtain

$$\delta = \frac{||b_l \sin\theta||}{||\left(-X_0 \sin\theta - Z_0 \cos\theta \cos\psi, -Y_0 \sin\theta - Z_0 \cos\theta \sin\psi\right)||} \tag{44}$$

An approximated expression can be determined for the pdf of the DG for this general case (Tardón, 1999); however, a much more tractable and useful expression can be obtained if we assume that the primitives P and Q are approximately centered between the two cameras or if we use small aperture cameras. In this case, the conditions $Z_0 \gg X_0$, $Z_0 \gg Y_0$ and $\theta \neq \frac{\pi}{2}$ (not that in this case occlusions can not occur) are satisfied. Then, the DG can be expressed in the following simplified way:

$$\delta = \frac{b_l \sin\theta}{Z_0 |\cos\theta|} \tag{45}$$

and the pdf will be (Tardón, 1999), (Tardón et al., 2004):

$$f_\Delta(\delta) = \frac{\frac{2}{\pi}\frac{b_l}{Z_0}}{\delta^2 + \left(\frac{b_l}{Z_0}\right)^2} \tag{46}$$

This is a unilateral Cauchy pdf with parameters 0 and $\frac{b_l}{Z_0}$ ($UCau(0, \frac{b_l}{Z_0})$) (see figure 14). This pdf favors the label assignments with low DG values as required. This tendency to favor low DG matches increases when the ratio $\frac{b_l}{Z_0}$ decreases, as expected.

8.4 The likelihood function

Now, we consider the information that can be extracted form the observations that will be used for matching. In other words, we deal now with a measure of the probability of a certain observation y given an outcome of the MRF x. Observe that the intensity values of the pixels in the two images of the stereo pair located in a window centered at the matching primitives should be similar. So, a similarity measure defined taking into account this idea should be higher in windows centered about correct matching primitives than in windows centered at unrelated projections.
We will use a function, $\mathcal{V}$ (t.b.d.), of the normalized-cross-covariance $\mathcal{N}$ (Kang et al., 1994) to measure the similarity between every pair of corresponding primitives and to model $p(y/x)$ accordingly. Using the selected measure, the role played by the observation y will be played, here, by $\mathcal{V} = \mathcal{N}^2$, given the underlying disparity map x. Then, the likelihood function will be denoted by

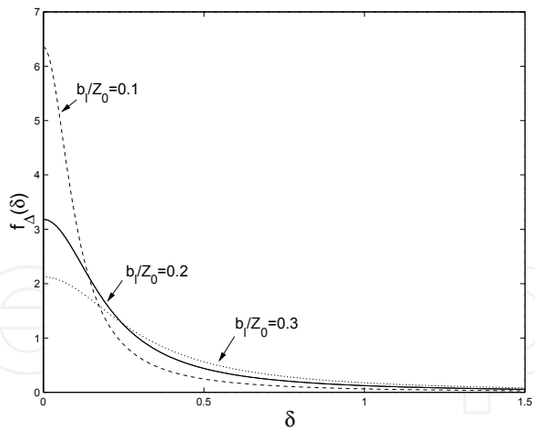


Fig. 14. Unilateral Cauchy pdf.

$$p_{S\mathcal{N}}(y/x) = \frac{1}{Z_{y/x}} e^{-H_{S\mathcal{N}}(y/x)} \tag{47}$$

And the energy of the system due to the similarity measures will be:

$$H_{S\mathcal{N}}(y/x) = \sum_{i=1}^{N} U_{\mathcal{N}}(n_i, l_{n_i}) \tag{48}$$

where the node n_i is matched to the label l_{n_i}. A natural choice for the potential functions is

$$U_{\mathcal{N}}(n_i, l_{n_i}) \propto -\ln f_{\mathcal{V}}(v) \tag{49}$$

where $f_{\mathcal{V}}(v)$ stands for the probability density function of the square of the normalized cross-covariance $\mathcal{V} = \mathcal{N}^2$. We use $f_{\mathcal{V}}(v)$ to derive a suitable form of the potential function as stated in (49) (Tardón, 1999),(Tardón et al., 2006).

8.4.1 Probabilistic analysis of the normalized cross-covariance

First of all, we recall the correlation coefficient (also called normalized cross-covariance (Kang et al., 1994)):

$$\mathcal{N}(N_i, L_j) = \frac{E\left[\{N_i - E[N_i]\} \cdot \left\{L_j - E[L_j]\right\}\right]}{\left(E\left[\{N_i - E[N_i]\}^2\right] \cdot E\left[\left\{L_j - E[L_j]\right\}^2\right]\right)^{\frac{1}{2}}} \tag{50}$$

where E represents the mathematical expectation operator and N_i and L_j, are the gray levels of the image windows considered, which will be treated as random variables, of node n_i, in the left image, and label l_j, in the right image, respectively. Needless to say, this coefficient must be replaced in practice by its estimation from the available data.

We assume that the image intensity can be considered Gaussian in each estimation window (Lim, 1990), with additive Gaussian noise. We will assume that only one of the image will corrupted by noise (Kanade & Okutomi, 1994). Specifically, let η denote a vector of independent and identically distributed Gaussian random variables, then $N_i = G + \eta$ and and $L_j = G$, where $G \sim N(\eta_l, \sigma_l)$ stands for the gray level in the absence of noise and $\eta \sim N(0, \sigma_\eta)$

represents the noise that corrupts the image with the labels. Using these conditions and operating in (50) we can find the following expression for the square of $\mathcal{N}$:

$$\mathcal{N}^2 = \frac{\sigma_l^2}{\sigma_l^2 + \sigma_\eta^2} \tag{51}$$

We will use the natural estimators of σ_n^2 and σ_l^2 and, so, we obtain the sample unbiased variances $\hat{\sigma}_n^2$ and $\hat{\sigma}_l^2$ using windows placed on both sides of each edge detected.
The noise η is obtained from the difference between the matched windows. The estimated unbiased variances $\hat{\sigma}_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{m}_x)^2$ will behave as gamma r.v.'s (Bain & Engelhardt, 1989) with parameters

$$\alpha = \frac{N-1}{2} \text{ and } \phi = \frac{2\sigma_x^2}{N-1} \tag{52}$$

For simplicity, for each window, let $\mathcal{V} = \hat{\mathcal{N}}^2$ and denote $a = \hat{\sigma}_l^2$ and $b = \hat{\sigma}_\eta^2$ which are two independent gamma r.v.'s: $A \sim \gamma(\alpha_a, \phi_a)$ and $B \sim \gamma(\alpha_b, \phi_b)$. Their joint pdf will be the product of the two gamma pdfs, and then, the pdf of

$$\mathcal{V} = \frac{A}{A+B} \tag{53}$$

is readily obtained (Tardón & Portillo, 1998). Using those results, one arrives at

$$f_{\mathcal{V}}(\nu) = \begin{cases} \frac{(1-\nu)^{\alpha_b - 1}\nu^{\alpha_a - 1}}{B(\alpha_a, \alpha_b)} \cdot \frac{\phi_a^{\alpha_b}\phi_b^{\alpha_a}}{(\phi_b\nu - \phi_a\nu + \phi_a)^{\alpha_a + \alpha_b}} & , \nu \in [0,1] \\ 0 & , \text{otherwise} \end{cases} \tag{54}$$

with $B(\cdot,\cdot)$ the beta function. We call this pdf *generalized beta* and denote it by $Gbeta_{\mathcal{V},\left(\alpha_a,\alpha_b,\frac{\phi_b}{\phi_a}\right)}(\nu)$ (Tardón, 1999), (Tardón & Portillo, 1998) (Fig. 15). Observe that the *Gbeta* pdf is far more versatile that the *beta* pdf and the former naturally subsumes the behavior of the latter.
However, we have not finished with our model yet since a good estimate of the noise power will not be available at the early stages of the algorithm. In fact, the difference between the matched windows incorporates both actual noise and noise due to the incorrect matches. Then, the main idea, now is to consider the estimated noise power as an upper bound of the actual noise power.
Consider the same variables A and B, but assume, now, that ϕ_b is a uniform r.v. (Φ_b) (Law & Kelton, 1991) within the interval $[0, \phi_B]$, with ϕ_B the upper bound. Then, the conditional pdf of B given $\Phi_b = \phi_b$ is *gamma*, and the joint pdf of B and Φ_b is $f_{B,\Phi_b}(b,\phi_b) = f_{B/\Phi_b}(b/\phi_b) f_{\Phi_b}(\phi_b)$.
Then, it is possible to obtain the pdf of $\mathcal{V}$ defined by (53) ((Tardón & Portillo, 1998)):

$$f_{\mathcal{V}}(\nu) = \begin{cases} \frac{1}{\nu^2}\frac{\alpha_a}{\alpha_b - 1}\frac{\phi_a}{\phi_B} I_{\frac{\phi_B\nu}{\phi_a - \phi_a\nu + \phi_B\nu}}(\alpha_a + 1, \alpha_b - 1) & , \nu \in [0,1] \\ 0 & , \text{otherwise} \end{cases} \tag{55}$$

where $I_*(\cdot)$ stands for the incomplete beta function (Abramowitz & Stegun, 1970) and α_* and ϕ_* are defined in (52).
We call this function *asymmetric beta* pdf and we will denote it by $Abeta_{\mathcal{V},\left(\alpha_a,\alpha_b,\frac{\phi_B}{\phi_a}\right)}(\nu)$. Figure 16 illustrates the behavior of this function.

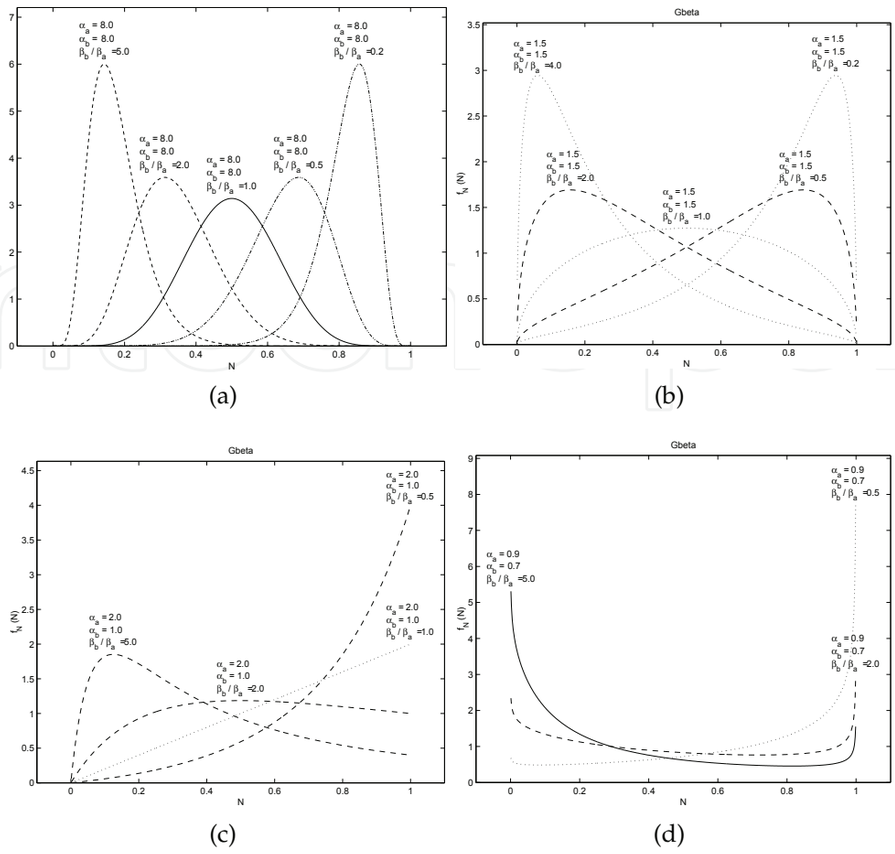


Fig. 15. $Gbeta_{\mathcal{V},\left(\alpha_a,\alpha_b,\frac{\phi_b}{\phi a}\right)}(v)$.

Since $Abeta(\cdot) > 0$ for $\mathcal{N}^2 \in (0,1)$, reasoning as in section 8.3, we can use the derived *asymmetric beta* pdf for the normalized-cross-covariance to define the energy $H_{S\mathcal{N}}(y/x)$ (equation (48)), as stated in equation (49).

8.5 The posterior pdf

After all the pdfs are available, the posterior distribution will be found suing the Bayes rule:

$$p_S(x/y) \propto p_{S\Delta}(x) p_{S\mathcal{N}}(y/x) \tag{56}$$

Its energy can be written as follows:

$$H_S(x/y) = H_{S\Delta}(x) + H_{S\mathcal{N}}(y/x) \tag{57}$$

Since $p_{S\Delta}(x)$ and $p_{S\mathcal{N}}(y/x)$ are Gibbs functions, then $p_S(x/y)$ is also a Gibbs function and, consequently, it describes a MRF.

Once the posterior pdf has been defined, the MAP estimator of the disparity map can be obtained by well-known procedures (Winkler, 1995; Boykov et al., 2001; Geman & Geman, 1984) (Sec. 6.3).

Note that, after equation (57), it is clear that classical area correlation techniques only make use of the information that would be included in $H_{S\mathcal{N}}$.

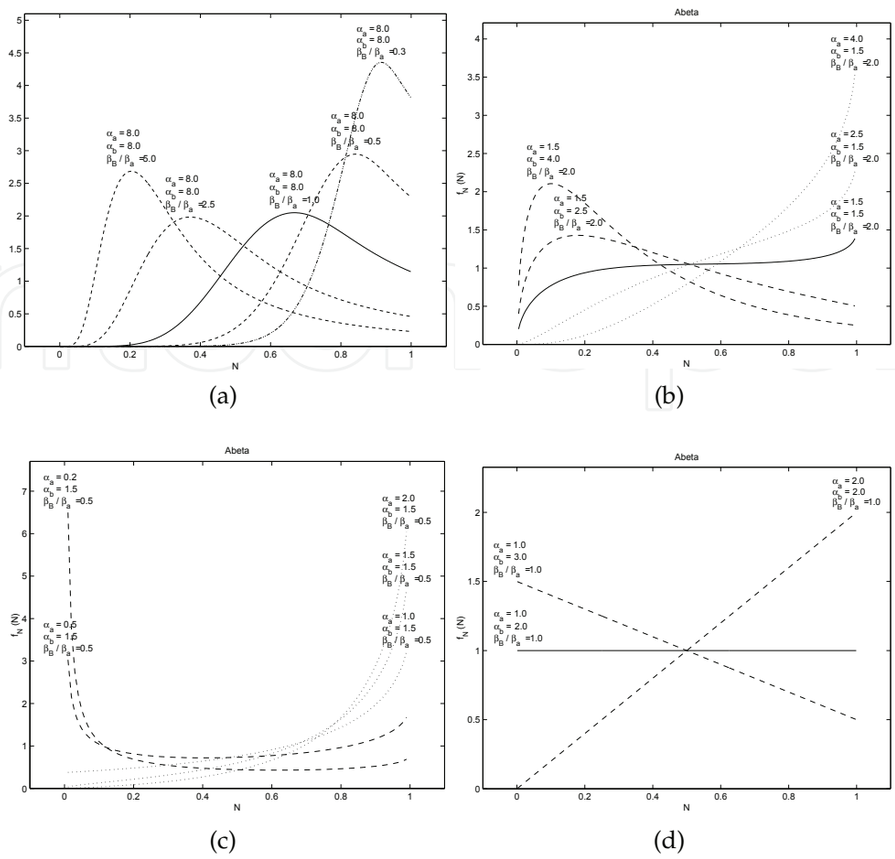


Fig. 16. $Abeta_{\mathcal{V},\left(\alpha_a,\alpha_b,\frac{\phi_B}{\phi a}\right)}(v)$.

9. Implementation of a stereo correspondence system with a MRF model

In this section we include a number of notes about the model presented, the implementation and the technique used to solve the problem. Afterwards, we show some examples of the application of the algorithm to solve different stereo pairs.

9.1 Implementation details. Object model

The use of Markov fields allows not only to specify how the correspondence of each node with respect to each neighborhood and a similarity measure of the nodes must be established but also to define of a stereo correlation system intrinsically parallelizable (Geman & Geman, 1984). In fact, the system that implements the MRF based stereo matching algorithm is implemented according to an object-oriented paradigm. So, we briefly describe the implementation of the system using the Object Modeling Technique *OMT* (Rumbaugh, 1991). In accordance with the description of simulation algorithm of Markov chains (Gibbs sampler with simulated annealing, Table 1), the decision on the correspondence of each node is done at each node, according to a certain set of neighbors which are used to build the functions involved in the model. A set of labels (including the null-correspondence label) will be available to establish its correspondence according to the local characteristic. Specifically, each node at each iteration computes the prior and the likelihood pdfs, according to the neighborhood system defined, to solve its own correspondence.

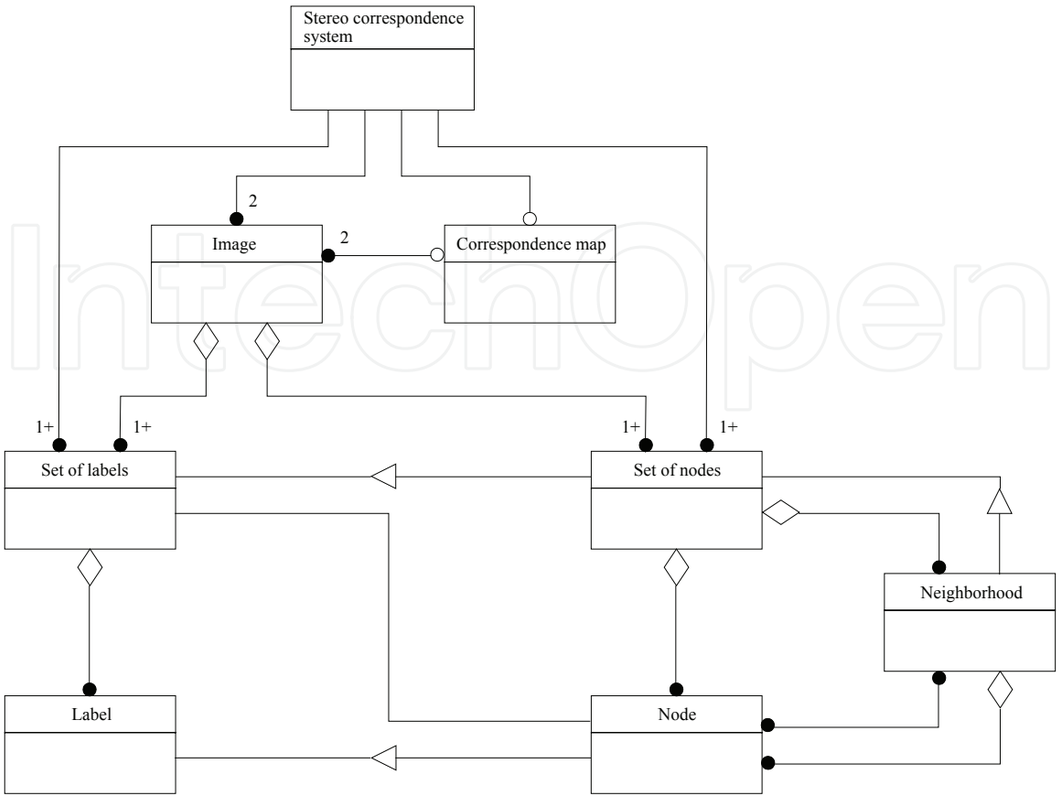


Fig. 17. MRF based stereo correspondence system. Object model (Rumbaugh, 1991).

Fig. 17 shows an object model that describes the relations between the main entities in the system.
The object that establishes the correspondence will be connected with at least other two that represent the real-world images and an initial correspondence map (if available). This object will contain a set of nodes and a set of labels (their roles are interchangeable: correspondence can be established from an image to another and vice versa), which are the features to match in both images of a stereo pair. Each of these sets is made up of many nodes or labels, respectively. Each node will be related to a neighborhood (a subset of the nodes in that image) and to a set of labels in the other image (plus the null correspondence label).
The sets of nodes and labels are identical, except for addition of some extra features in the set of nodes. So, the set of nodes and each particular node are derived from the set of labels and each particular label objects, respectively.
The main functionalities of the nodes, which constitute the main processing unit, are the following:

1. Define the set of possible labels to establish its own correspondence.
2. Define its neighborhood.
3. Determine an initial correspondence selecting a label from the available set for that node.
4. Establish its own correspondence using the local characteristic according to the information given by the neighborhood and using its particular set of possible labels.

The operation of the system is based on the activity of each node, which performs a relatively simple task at each stage: select randomly a label in accordance with the local characteristic to iteratively evolve toward the MAP estimator of the correspondence field.

9.2 Implementation details. Parameters and procedure

Correspondences are sought only within preselected features, instead of searching in the whole image, this helps to reduce computational burden. The features selected are edge pixels obtained by the Nalwa-Binford edge detector (Nalwa & Binford, 1986); specifically, the central pixel in the fitted surface is used. Other edge detectors could be used, including the MRF based edge detector described, however, the Nalwa-Binford edge detector has been selected because of its availability and because it extracts edge pixels with subpixel accuracy.

The matching procedure is performed solely from the left to the right image; uniqueness is imposed by the DG constraint itself (Li & Hu, 1996). The neighborhood system is defined by the nodes that lie inside a region defined around every node and a similar criterion is used to define the set of possible matches or labels for a node. The set of possible matches for a node will be defined by all the labels that lie inside the corresponding search window: a region centered at a likely match position. This selection is not critical since large windows will almost-surely contain the right label. Superellipses are used to define these regions in a compact widely usable form.

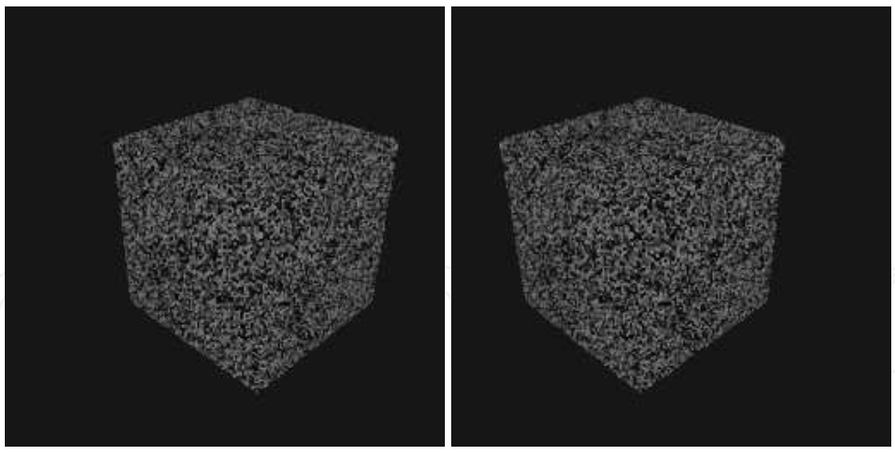
The null match, i.e., the label that leaves a node unmatched, must always belong to the set of possible matches, so its energy must be adequately defined. To this end, consider, separately, the a priori and likelihood information.

- Regarding the a priori information, and recalling how the HVS works, we define the energy of the null match as the energy of a *virtual match* in which all the neighbors have a DG equal to 0.8; this means that a null match is (probabilistically) preferred to other matches with a larger DG.
- With respect to the likelihood term, since the null match has obviously no data, we need to define it. We have implemented this choice as follows: recalling equation (54), for every node, obtain the energy of the current assignment (the one from the previous iteration) and pick the maximum of the *Gbeta* pdf under these working conditions. Let ν_{max} denote the mode of the *Gbeta* function. Then, find the argument ν_n of *Gbeta* (leftwards from the mode since no assignment tends to uncorrelation) such that $Gbeta_{\nu,\left(\alpha_a,\alpha_b,\frac{\phi_b}{\phi_a}\right)}(\nu_n)$ is half $Gbeta_{\nu,\left(\alpha_a,\alpha_b,\frac{\phi_b}{\phi_a}\right)}(\nu_{max})$. Finally, use this value, ν_n, as the argument to define the energy of the null match according to the likelihood information.

To evolve towards the MAP estimator we have resorted to a practical suboptimal cooling scheme (Winkler, 1995), defined by the following system temperature: $T = T_0 \cdot T_B^k$, where $T_0 = 1$, $T_B = 0.9998$ and k is the sweep number.

Different techniques to establish the initial matchings can be selected, however the initial state is significant only during the first stages of the algorithm, and after a number of iterations, the algorithm evolves to a solution independently of the initial state (Winkler, 1995).

The ratio $\frac{b_l}{Z_0}$ ($\frac{\text{baseline}}{\text{subject distance}}$) modifies the sharpness of the a priori pdf (46) and so, the selection of this parameter has an influence on the system performance; if this parameter is too small, the algorithm could be easily trapped in local maxima. The ratio $\frac{b_l}{Z_0}$ has been manually tuned. However, note that it could be accurately estimated for every tentative matching using the calibration parameters.



(a) (b)

Fig. 18. *Cube* stereo pair. a) Left image. b) Right image.

10. Experiments

We can observe the performance of the MRF based stereo matching system presented in this chapter in a number of experiments done with synthetic and real world stereo pairs (see Acknowledgments).

10.1 Synthetic images

Consider the synthesized random dot stereogram (RDS) (*cube*) shown in Fig. 18. The epipolar lines are horizontal so the search window becomes a segment of the corresponding epipolar line in the right image. We have used a horizontal disparity search within the interval $[-50,-20]$ pixels. The image size is 256×256 with 256 gray levels. Nodes and labels have been defined as those points that exceed an intensity threshold of 80, giving rise to the number of features shown in Table 2. The ratio $\frac{b_l}{Z_0}$ (recall figures 12 and 13) is approximately 0.3 and the cooling schedule is as described in section 9.2.

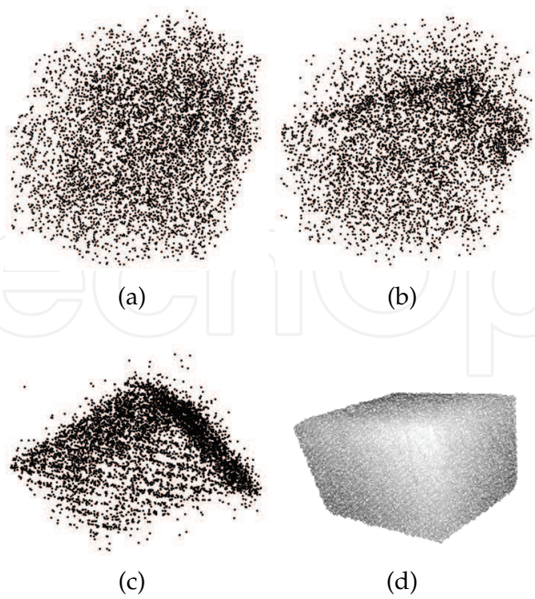
Note that, since there is no other information available, the neighborhood includes all the nodes in a circular region centered at each node. Regarding the size of the neighborhood, it should be large enough so that a sufficiently large set of nearby nodes can be employed to define the local interactions (Besag, 1974).

In this experiment, we have consciously ignored the brightness information of nodes and labels; this is equivalent to assuming that the likelihood pdf is non informative, i.e., the disparity map will only be a function of the *DG*.

Figure 19 shows a perspective view of the evolution of the disparity map with the number of iterations of the simulated annealing algorithm. The initial disparity map, shown in figure 19 a), is obtained randomly; it is just a random cloud of points. Also the final disparity

	Size		Selected features		
	Rows	Columns	# of nodes	# of labels	$\frac{b_l}{Z_0}$
Cube	256	256	6284	6332	0.3
rd1	250	250	8269	12834	0.3

Table 2. Synthetic images

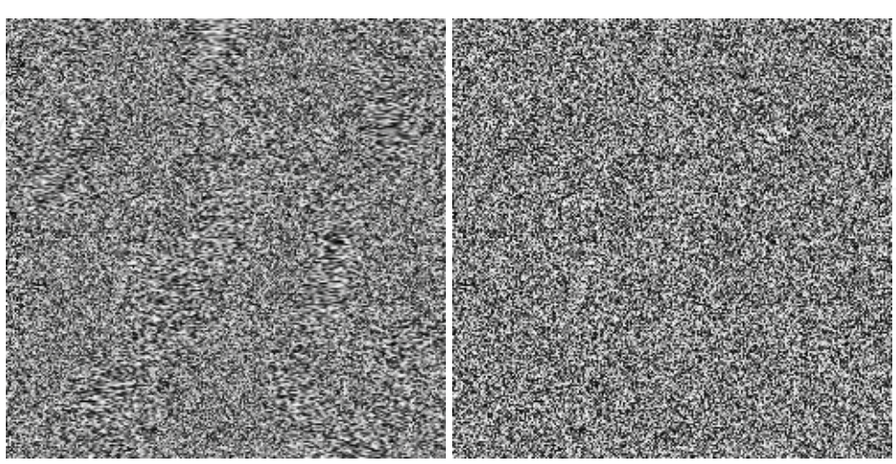


(a) (b)

(c) (d)

Fig. 19. *Cube* disparity map. a) Initial (random) configuration. b) After 500 iterations. c) After 5000 iterations. Three faces of the cube are clearly visible. d) After 10000 iterations. Interpolated disparity map of the stereo pair *cube*. Modified Hardy interpolation used with $b = 3$, radius$= 15$ and number of base functions$= 15$ (Vázquez, 1998).

map obtained after 10000 iterations, interpolated using a Hardy-like interpolation technique (Franke, 1982), (Bradley & Vickers, 1993), (Vázquez, 1998), is shown in Fig. 19 d).
We have also applied our stereo algorithm to the synthesized random dot stereogram *rd1* shown in Fig. 20.
Again, the epipolar lines are horizontal. The search region is a segment of the corresponding epipolar line in the right image defined by the following interval: $[-20,20]$ pixels. The image size is 250×250 with 256 gray levels. Nodes and labels have been defined as those points that



(a) (b)

Fig. 20. *Rd1* stereo pair. a) Left image. b) Right image.

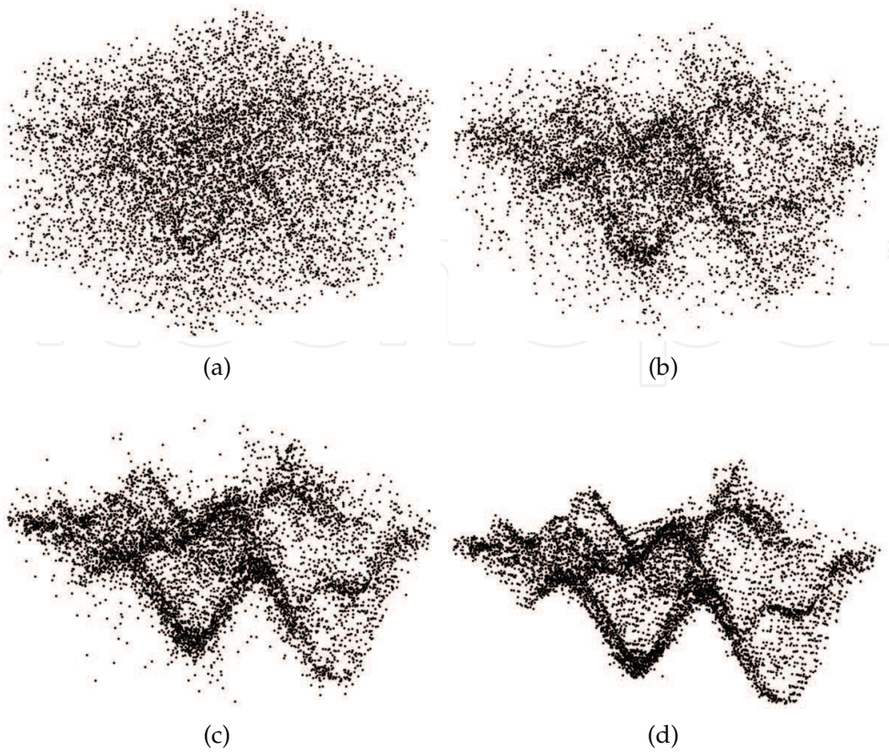


Fig. 21. Evolution of the correspondence map for the *rd1* stereo pair. a) Initial (random) configuration. b) After 1000 iterations. c) After 3000 iterations. d) After 7000 iterations.

exceed an intensity threshold of 80, giving rise to the number of features shown in Table 2. The ratio $\frac{b_I}{Z_0}$ (recall figures 12 and 13) is approximately 0.3. The cooling scenario is unchanged. Brightness information is ignored.
Fig. 21 shows the evolution of the correspondence map as the iteration number increases.

10.2 Real world images

In this section, we show the results found on some real world images. The images have been geometrically corrected to make the epipolar lines horizontal before performing the matching procedure. The rectification is done using the fundamental matrix (Sec. 5.1.1), which is estimated using a number of matches obtained carrying out a preliminary matching stage using the MRF based matching technique described. In this case, a smaller number of iterations are performed and the neighborhood and the search regions are defined by large superellipses with parameter $p = 2$. The search windows are, in this stage, circles (Sec. 8.2), centered at the expected matching label. The diameter of the window is large enough to capture the real matching, if any, even for high disparity values. Afterwards, only the matchings with highest probability are selected to estimate the fundamental matrix (usually between 100 and 200 matching points) (Tardón, 1999).
Nodes and labels (edges) are detected (see Sec. 9.2). Note that only the pixel that lays at the center of the edge detector window with a contrast larger than 70 is selected as node or label in the left and right images, respectively. The information of the node position and the edge orientation will be used to place the windows from which the normalized cross-covariance

(a) (b)

Fig. 22. *Pentagon* stereo pair. a) Left image. b) Right image.

will be calculated. A window at each side of the edge is considered to calculate the normalized cross-covariance. The outcomes of this measure, at each side of the edge, are considered to be independent.
Recall that we consider that the image intensity levels are of Gaussian nature and that these variables are affected by Gaussian noise in one of the images. Then, the asymmetric beta function can be used to model the behavior of the normalized-cross-covariance.
For the rectified stereo pair *pentagon*, shown in Fig. 7 c) and d), table 3 shows the size of the images, the number of features (nodes and labels) selected to establish correspondence (Fig. 23) and the approximate ratio $\frac{b_l}{Z_0}$.

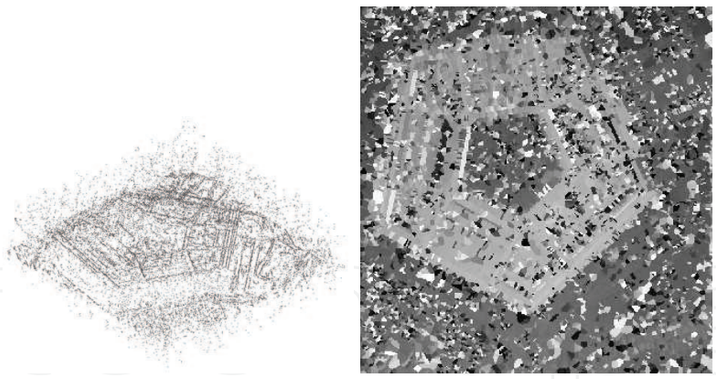
	Size		Selected features		
	Rows	Columns	Left image	Right image	$\frac{b_l}{Z_0}$
Pentagon	512	512	26491	28551	0.01
Baseball	512	512	23762	24809	0.15

Table 3. Real world images

In order to establish the correspondence in the *pentagon* stereo pair, the horizontal search range is ± 15 pixels and the neighborhood of a node n_i is composed of the nodes ranging less than 25 pixels from n_i (the neighborhood area is a superellipse with $a = b = 25$ and $p = 2$).

(a) (b)

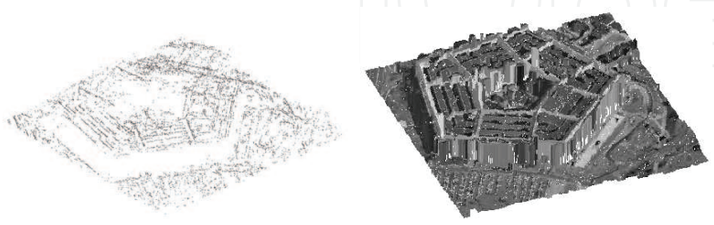
Fig. 23. Nodes and labels selected in the *pentagon* stereo pair to establish the correspondence. a) Left image (nodes). b) Right image (labels).



(a) (b)

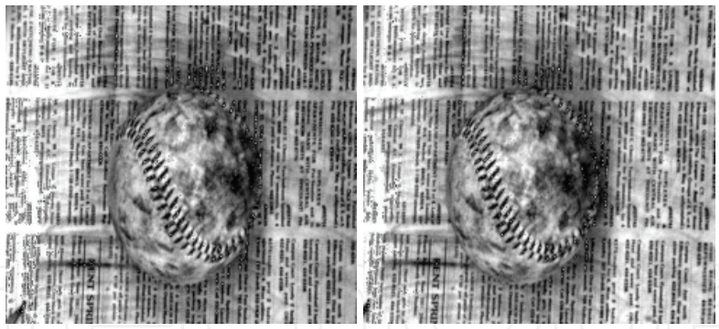
Fig. 24. Disparity map for the *pentagon* stereo pair obtained using the normalized-cross-covariance. a) Matched points. b) Top view, with coded disparity, of the disparity map interpolated using planar patches (Bradley & Vickers, 1993).

Fig. 24 a) shows the disparity map obtained using only the likelihood information: the normalized cross-covariance. Fig. 24 b) shows a top view of the interpolated disparity map (Bradley & Vickers, 1993) (planar patches are grown around each matched node) with coded disparity (brighter color for larger disparity). Observe the noisy disparity map obtained.
Fig. 25 a) shows the disparity map obtained after 5000 iterations of the algorithm with simulated annealing using both a priori and likelihood. Fig. 25 b) shows the final disparity map interpolated using the Sheppard technique (Bradley & Vickers, 1993), the original gray levels where applied to the 3D representation.
The second example in this section is the baseball pair shown in Fig. 26. Table 3 shows the size of the *baseball* images, the number of nodes selected to establish correspondence and the approximate ratio $\frac{b_l}{Z_0}$. The search region ranges from -50 to -5 pixels and the neighborhood area is a circle of radius 15 pixels. Results are shown in figure 27 with an isometric plot of the matched nodes, a disparity coded view and the interpolated data with the same technique as before. Note that in this case, the lack of 3D information is evident in the reconstructed image. An objective of the evaluation of the performance of a stereo correspondence system can be found in (Tardón et al., 2006).



(a) (b)

Fig. 25. Disparity map for the *pentagon* stereo pair after 5000 iterations of the MRF based stereo correspondence algorithm. a) Matched points. b) 3D reconstruction. Surface interpolated using planar patches (Bradley & Vickers, 1993)



(a) (b)

Fig. 26. *Baseball* stereo pair. a) Left image. b) Right image.

11. Concluding remarks

In this chapter, we have shown how MRFs can be effectively used to solve the stereo correspondence problem and how the fields can be designed making use of the main concepts of cliques, energy and potentials that contribute to define the local characteristic of the MRF.
Local interactions between edge pixels and between matching points have been incorporated to a specific MRF model to solve the correspondence problem using a Markovian formulation. It has been shown how both a priori and a posteriori probabilities can be derived and incorporated in the MRF model. Probabilistic analyses have been described that lead to the definition of the functions that gave rise to the MRF model to solve the correspondence problem.
A Bayesian approach to edge detection based on MRFs has been briefly introduced because of its connection to the correspondence problem through MRF models.
Regarding the specific MRF model for stereo correspondence. We have described a complete Bayesian approach in which the a priori information is derived upon the probabilistic characterization of the disparity gradient obtained after a detailed analysis of its behavior under a specific camera model (the pinhole camera model). The likelihood term is derived upon the probabilistic characterization of the normalized-cross-covariance.
It is important to observe how MRFs can take into account psychovisual cues. Another main aspect of MRFs in the stereo vision context is that MRFs are able to cope, simultaneously, with both prior information extracted from the HVS (in our case related to the disparity gradient) and likelihood information (related to the normalized-cross-covariance in our model).
Note that in a stereo correspondence system, the null-correspondence must be taken into account since occlusions may happen and, then, some points in an image will not be able

(a) (b)

Fig. 27. *Baseball*.a) Disparity map after 5000 iterations. b) 3D reconstruction of the *baseball* scene.

to find their correspondence in the other image. This must be taken into account in any probabilistic correspondence method.

12. Acknowledgments

The image *Lenna* was obtained from the Electrical Engineering Department at the Signal & Image Processing Institute from the University of Southern California (USC).
The stereo pairs *cube*, *rd*, *pentagon* and *baseball* were obtained from the Vision and Autonomous Systems Center Database from the Carnegie Mellon University (CMU) (they were provided by Bill Hoff, University of Illinois).
This work has been partly funded by Junta de Andalucía under Project Number P07-TIC-02783, by the Spanish Ministerio de Ciencia e Innovación under Project Number TIN2010-21089-C03-02 and by the Spanish Ministerio de Industria Turismo y Comercio under Project Number TSI-020201-2008-0117.

13. References

Abramowitz, M. & Stegun, I. A. (1970). *Hanbook of Mathematical Functions*, Dover Publications Inc., New York.

Bain, L. J. & Engelhardt, M. (1989). *Introduction to Probability and Mathematical Statistics*, PWS-Kent Publishing Company.

Barnard, S. T. & Fischler, M. A. (1982). Computational stereo, *Computing Surveys* 14(4): 553 – 572.

Barnard, S. T. & Thompson, W. B. (1980). Disparity analysis of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-2(4): 333 – 340.

Bensrhair, A., Miché, P. & Debrie, R. (1992). Binocular stereo matching algorithm using prediction and verification of hypotheses, *Proc. ISSPA 92, Signal Processing and Its Applications*, pp. 167 – 170.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *J. Royal Statistical Society* 34: 192 – 236. Series B.

Boussaid, K. B., Beghdadi, A. & Dupoisot, H. (1996). Edge detection using Holladay's principle, *Proc. ICIP'96, IEEE Int. Conference on Image Processing*, Vol. I, pp. 833 – 836.

Boykov, Y., Veksler, O. & Zabih, R. (2001). Fast approximate energy minimization via graphs cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(11): 1222 – 1239.

Bradley, C. & Vickers, G. W. (1993). Free-form surface reconstruction for machine vision rapid prototyping, *Optical Engineering* 32(9): 2191 – 2200.

Brown, M. Z., Burschka, D. & Hager, G. D. (2003). Advances in computational stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-25(8): 993 – 1008.

Burt, P. & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area, *Perception* 9: 671 – 682.

Cochran, S. D. & Medioni, G. (1992). 3-d surface description from binocular stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(10): 981 – 994.

Cohen, F. S. & Cooper, D. B. (1987). Simple parallel hierarchical and relaxation algorithms for segmenting noncausal markovian random fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9(2): 195 – 219.

Duda, R. O. & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York.

Faugeras, O. (1993). *Three-Dimensional Computer Vision. A Geometric Viewpoint*, The MIT Press, Cambridge.

Foley, vanDam, Feiner & Hughes (1992). *Computer Graphics. Principles and Practice*, second edn, Addison- Wesley, Reading, Massachusetts.

Franke, R. (1982). Scatterd data interpolation: Test of some methods, *Mathematics of Computation* 38(157): 181 – 200.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6(6): 721 – 741.

Grimson, W. E. L. (1985). Computational experiments with a feature based stereo algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-7(1): 17 – 34.

Hoff, W. & Ahuja, N. (1989). Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(2): 121 – 136.

Kanade, T. & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(9): 920 – 932.

Kang, M. S., Park, R.-H. & Lee, K.-H. (1994). Recovering an elevation map by using stereo modeling of the aerial image sequence, *Optical Engineering* 33(11): 3793 – 3802.

Kinderman, R. & Snell, J. L. (1980). *Markov Random Fields and Their Applications*, Providence RI, American Mathematical Society.

Lane, R. A., Thacker, N. A. & Seed, N. L. (1994). Stretch-correlation as a real-time alternative to feature-based stereo matching algorithms, *Image and Vision Computing* 12(4): 203 – 212.

Law, A. M. & Kelton, W. D. (1991). *Simulation Modeling & Analysis*, second edn, McGraw-Hill International Editions.

Li, S. Z. (2001). *Markov Random Field Modeling in Image Analysis*, Springer-Verlag.

Li, S. Z., Wang, H., Chan, K. L. & Petrou, M. (1997). Minimization of MRF energy with relaxation labeling, *Journal of Mathematical Imaging and Vision* 7: 149 – 161.

Li, Z.-N. & Hu, G. (1996). Analysis of disparity gradient based cooperative stereo, *IEEE Transactions on Image Processing* 5(11): 1493 – 1506.

Lim, J. S. (1990). *Two-Dimensional Signal and Image Processing*, Prentice Hall Inc., Englewood Cliffs, New Jersey.

Luong, Q.-T. & Faugeras, O. D. (1996). The fundamental matrix: Theory, algorithms and stability analysis, *Int. Journal of Computer Vision* 17: 43 – 75.

Marapane, S. B. & Trivedi, M. M. (1989). Region-based stereo analysis for robotic applications, *IEEE Transactions on Systems, Man and Cybernetics* 19: 1447–1464. Special issue on computer vision.

Marapane, S. B. & Trivedi, M. M. (1994). Multi-Primitive Hierarchical (MPH) stereo analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(3): 227 – 240.

McKee, S. P. & Verghese, P. (2002). Stereo transparency and the disparity gradient limit, *Vision Research* 42: 1963 – 1977.

Mohr, R. & Triggs, B. (1996). Projective geometry for image analysis. Tutorial given at ISPRS, Vienna.

Moravec, H. P. (1977). Towards automatic visual obstacle avoidance, *Proc. 5th Int. Joint Conf. Artificial Intell*, Cambridge, MA, p. 584.

Nalwa, V. S. & Binford, T. O. (1986). On detecting edges, *IEEE Transactions on Pattern Analysis*

and Machine Intelligence PAMI-8(6): 699 – 714.

Ohta, Y. & Kanade, T. (1985). Stereo by intra- and inter-scanline search using dynamic progamming, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-7(2): 139 – 154.

Olsen, S. I. (1990). Stereo correspondence by surface reconstruction, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(3): 309 – 315.

Pollard, S. B., Mayhew, J. E. W. & Frisby, J. P. (1985). PMF: A stereo correspondence algorithm using a disparity gradient limit, *Perception* 14: 449 – 470.

Pollard, S. B., Porrill, J., Mayhew, J. E. W. & Frisby, J. P. (1986). Disparity gradient, Lipschitz continuity and computing binocular correspondences, *Robotics Research: The Third International Symposium* pp. 19 – 26.

Rumbaugh, J. (1991). *Object-Oriented Modelling and Design*, Prentice-Hall International Editions, London.

Sherman, D. & Peleg, S. (1990). Stereo by incremental matching of contours, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(11): 1102 – 1106.

Tardón, L. J. (1999). *A robust method of 3D scene reconstruction using binocular information*, PhD thesis, E.T.S.I. Telecomunicación, Univ. Politécnica de Madrid. In spanish.

Tardón, L. J., Barbancho, I. & Marquez, F. (2006). A markov random field approach to edge detection, *Proceedings of the IEEE Mediterranean Electrotechnical Conference MELECON 2006*, pp. 482 – 485.

Tardón, L. J. & Portillo, J. (1998). Two new beta-related probability density functions, *IEE Electronics Letters* 34(24): 2347 – 2348.

Tardón, L. J., Portillo, J. & Alberola, C. (1999). Markov Random Fields and the disparity gradient applied to stereo correspondence, *Proc. of the IEEE International Conference on Image Processing, ICIP-99*, Vol. III, pp. 901 – 905.

Tardón, L. J., Portillo, J. & Alberola, C. (2004). A novel markovian formulation of the correspondence problem in stereo vision, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 34(6): 779 – 788.

Trivedi, H. P. (1986). On the reconstruction of a scene from two unregistered images, *Proceedings of the AAAI*, pp. 652 – 656.

Trucco, E. & Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*, Prentice-Hall.

Vázquez, J. I. (1998). *Surface reconstruction from sparse data*, Master's thesis, E.T.S.I. Telecomunicación, Univ. Politécnica de Madrid, Madrid. In spanish.

Vince, J. A. (1995). *Virtual Reality Systems*, ACM Press Books, Siggraph Series, Addison-Wesley.

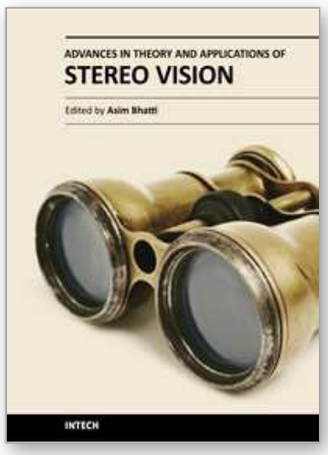
Wainman, G. (1997). *The effect of stimulus properties on the disparity gradient threshold for diplopia*, Master's thesis, York University.

Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Vol. 27 of *Applications of Mathematics*, Springer-Verlag.

Xie, M. & Liu, L. Y. (1995). Color stereo vision: Use of appearance constraint and epipolar geometry for feature matching, *in* S. Z. Li, D. P. Mital, E. K. Teoh & H. Wang (eds), *Recent Developments in Computer Vision*, Lecture Notes in Computer Science, Springer, pp. 255 – 264. Second Asian Conf. on Computer Vision, ACCV'95, Singapore, Invited Session Papers.

Zhang, Y. & Gerbrands, J. J. (1995). Method for matching general stereo planar curves, *Image and Vision Computing* 13(8): 645 – 655.

Zhang, Z. (1996). Determining the epipolar geometry and its uncertainty, *Technical Report 2927*, Institut National de Recherche en Informatique et en Automatique, INRIA. Rev. ver.



Advances in Theory and Applications of Stereo Vision
Edited by Dr Asim Bhatti

ISBN 978-953-307-516-7
Hard cover, 352 pages
Publisher InTech
Published online 08, January, 2011
Published in print edition January, 2011

The book presents a wide range of innovative research ideas and current trends in stereo vision. The topics covered in this book encapsulate research trends from fundamental theoretical aspects of robust stereo correspondence estimation to the establishment of novel and robust algorithms as well as applications in a wide range of disciplines. Particularly interesting theoretical trends presented in this book involve the exploitation of the evolutionary approach, wavelets and multiwavelet theories, Markov random fields and fuzzy sets in addressing the correspondence estimation problem. Novel algorithms utilizing inspiration from biological systems (such as the silicon retina imager and fish eye) and nature (through the exploitation of the refractive index of liquids) make this book an interesting compilation of current research ideas.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Lorenzo J. Tardón, Isabel Barbancho and Carlos Alberola (2011). Markov Random Fields in the Context of Stereo Vision, Advances in Theory and Applications of Stereo Vision, Dr Asim Bhatti (Ed.), ISBN: 978-953-307-516-7, InTech, Available from: http://www.intechopen.com/books/advances-in-theory-and-applications-of-stereo-vision/markov-random-fields-in-the-context-of-stereo-vision

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen