



Relaxation labeling in stereo image matching

Gonzalo Pajares*, Jesús Manuel de la Cruz, José Antonio López-Orozco

Dpto. Arquitectura de Computadores y Automática, Facultad de CC Físicas, Universidad Complutense de Madrid 28040 Madrid, Spain

Received 8 July 1998; accepted 14 December 1998

Abstract

This paper outlines a method for solving the global stereovision matching problem using edge segments as the primitives. A relaxation scheme is the technique commonly used by existing methods to solve this problem. These techniques generally impose the following competing constraints: *similarity, smoothness, ordering and uniqueness*, and assume a bound on the disparity range. The smoothness constraint is basic in the relaxation process. We have verified that the smoothness and ordering constraints can be violated by objects close to the cameras and that the setting of the disparity limit is a serious problem. This problem also arises when repetitive structures appear in the scene (i.e. complex images), where the existing methods produce a high number of failures. We develop our approach from a relaxation labeling method ([1] W.J. Christmas, J. Kittler, M. Petrou, structural matching in computer vision using probabilistic relaxation, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 749–764), which allows us to map the above constraints. The main contribution is made, (1) by applying a learning strategy in the similarity constraint and (2) by introducing specific conditions to overcome the violation of the smoothness constraint and to avoid the serious problem produced by the required fixation of a disparity limit. Consequently, we improve the stereovision matching process. A better performance of the proposed method is illustrated by comparative analysis against some recent global matching methods. © 1999 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Stereo correspondence; Probabilistic relaxation; Similarity; Smoothness; Uniqueness

1. Introduction

A major portion of the research efforts of the computer vision community has been directed towards the study of the three-dimensional (3-D) structure of objects using machine analysis of images [2–4]. Analysis of video images in stereo has emerged as an important passive method for extracting the 3-D structure of a scene. Following the Barnard and Fischler [5] terminology, we can view the problem of stereo analysis as consisting of the following steps: image acquisition, camera modeling, feature acquisition, image matching, depth determination and interpolation. The key step is that of image matching, that is, the process of identifying the corresponding points in two images that are cast by the same physical point in 3-D space. This paper is devoted solely to this

problem. The basic principle involved in the recovery of depth using passive imaging is triangulation. In stereopsis the triangulation needs to be achieved with the help of only the existing environmental illumination. Hence, a correspondence needs to be established between features from two images that correspond to some physical feature in space. Then, provided that the position of centers of projection, the effective local length, the orientation of the optical axis, and the sampling interval of each camera are known, the depth can be established using triangulation [6].

1.1. Techniques in stereovision matching

A review of the state of art in stereovision matching allows us to distinguish two sorts of techniques broadly used in this discipline: *area based* and *feature based* [3,4,7,8]. Area-based stereo techniques use correlation between brightness (intensity) patterns in the local neighborhood of a pixel in one image with brightness patterns

* Corresponding author. Tel.: + 0034-91-3944477; fax: + 0034-91-3944687.

E-mail address: pajares@eucmax.sim.ucm.es (G. Pajares)

in the local neighborhood of the other image [9–14], where the number of possible matches is intrinsically high, while feature-based methods use sets of pixels with similar attributes, normally, either pixels belonging to edges [15–27], or the corresponding edges themselves [4,8,28–38]. These latter methods lead to a sparse depth map only, leaving the rest of the surface to be reconstructed by interpolation; but they are faster than area-based methods as there are many fewer points (features) to be considered [3]. We select a feature-based method as justified below.

1.2. Factors affecting the physical stereovision system and choice of features

There are extrinsic and intrinsic factors affecting the stereovision matching system: (a) *extrinsic*, in a practical stereovision system, the left and right images are obtained with two cameras placed at different positions/angles, although they capture both the same scene, each camera may receive a different illumination and also, incidentally, different reflections; (b) *intrinsic*, the stereovision system is equipped with two physical cameras, always placed at the same relative position (left and right), although they are the same commercial model, their optical devices and electronic components are different, hence each camera may convert the same illumination level to a different gray level independently.

As a result of the above-mentioned factors the corresponding features in both images may display different values. This may lead to incorrect matches. Thus, it is very important to find features in both images which are unique or independent of possible variation in the images [39]. This research paper use a feature-based method where edge segments are to be matched, since our experiment has been carried out in an artificial environment where edge segments are abundant. This environment is where a mobile robot equipped with the stereovision system navigates.

Such features have been studied in terms of reliability [7] and robustness [39] and, as mentioned before, have also been used in previous stereovision matching works. This justifies our choice of features, although they may be too local. Four average attribute values (module and direction gradient, variance and Laplace) are computed for each edge-segment as we will see later. The ideas described here can be used for other simple geometric primitives with other attributes, even though in the details of the current implementation the line segments play an important role.

1.3. Constraints applied in stereovision matching

Our stereo correspondence problem can be defined in terms of finding pairs of true matches, namely, pairs of edge segments in two images that are generated by the

same physical edge segment in space. These true matches generally satisfy some competing constraints [22,23,27]: (1) *similarity*, matched edge segments have similar local properties or attributes; (2) *smoothness*, disparity values in a given neighborhood change smoothly, except at a few depth discontinuities; (3) *ordering*, the relative position among two segments in the left image is preserved in the right one for the corresponding matches; (4) *uniqueness*, each edge segment in one image should be matched to a unique edge segment in the other image.

Immediately after, we introduce the very important disparity concept. Assume a stereo pair of edge segments, where one segment belongs to the left image and the other to the right one. If we superpose the right image over the left one, for example, the two edge segments of the stereo pair appear horizontally displaced. Following horizontal (epipolar) lines we determine the two intersection points for each epipolar line with the pair of edge segments. The relative displacement of the x-coordinates of the two intersection points is the disparity for such points. We compute the disparity for the pair of edge segments as the average displacement between points belonging to both edge segments along the common length.

The similarity constraint is associated to a *local* matching process and the smoothness and ordering constraints to a *global* matching one. The major difficulty of stereo processing arises due to the need to make *global correspondences*. A local edge segment in one image may match equally well with a number of edge segments in the other image. This problem is compounded by the fact that the local matches are not perfect due to the above-mentioned extrinsic and intrinsic factors. These ambiguities in local matches can only be resolved by considering sets of local matches globally. Hence, to make global correspondences given a pair of edge segments, we consider a set of neighboring edge segments, where a bound on the disparity range defines the neighborhood. Relaxation is a technique commonly used to find the best matches globally and it refers to any computational mechanism that employs a set of locally interacting parallel processes, one associated with each image unit, that in an iterative fashion update each unit's current labeling in order to achieve a globally consistent interpretation of image data [40,41].

1.4. Relaxation in stereovision

Relaxation labeling is a technique proposed by Rosenfeld et al. [42] and developed to deal with uncertainty in sensory data interpretation systems and to find the best matches. It uses contextual information as an aid in classifying a set of interdependent objects by allowing interactions among the possible classifications of related objects. In the stereo paradigm the problem involves assigning unique labels (or matches) to a set of features in

an image from a given list of possible matches. So, the goal is to assign each feature (edge segment) a value corresponding to disparity in a manner consistent with certain predefined constraints.

Two main relaxation processes can be distinguished in stereo correspondence: optimization-based and probabilistic/merit-based. In the optimization-based processes stereo correspondence is carried out by minimizing an energy function which is formulated from the applicable constraints. It represents a mechanism for the propagation of constraints among neighboring match features for the removal of ambiguity of multiple stereo matches in an iterative manner. The optimal solution is ground state, that is, the state (or states) of the lowest energy. In the probabilistic/merit-based processes, the initial probabilities/merits, established from a local stereo correspondence process and computed from similarity in the feature values, are updated iteratively depending on the matching probabilities/merits of neighboring features and also from the applicable constraints. The following papers use a relaxation technique: (a) probabilistic/merit [1,4,17,18,20–22,25,26,41–54] and (b) optimization through a Hopfield neural network [12,14,24,27,36,55,56]; we use (a).

1.5. Motivational research and contribution

For a given pair of features the smoothness constraint makes global correspondences based on how well its neighbors match. The neighborhood region is defined by a bound on the disparity range. In our experiments we have found complex images with highly repetitive structures or with objects very close to the cameras or with occlusions. In complex images setting a disparity limit is a difficult task because according to such a limit the smoothness constraint can be violated. Additionally, in complex images the ordering constraint is also easily violated.

We propose a probabilistic relaxation approach for solving the edge-segment stereovision matching problem and base our method on the work of Christmas et al. [1] (similar to these proposed by Rosenfeld et al. [42] or Peleg [48]) for the following reasons: (a) it incorporates the recent advances of the theory of probabilistic relaxation, (b) the similarity, smoothness and ordering constraints can be mapped directly, and (c) it has been satisfactorily applied in stereovision matching tasks [1].

The main contribution of the paper is the use in the relaxation process of a compatibility measure between pairs of matched segments. This enforces consistency imposing smoothness and ordering constraints when no suspicion exists about the violation of such constraints, otherwise a different compatibility measure is used and the lost global consistency is replaced by local ones through the matching probabilities. With this approach, our method is valid for common stereo images and it is

extended to deal robustly with complex stereo images where the smoothness and ordering constraints are commonly violated. The extension of our approach is based upon the idea of Dhond and Aggarwal [57] that deals with occluding objects.

Additional contributions are made in the following ways: (a) by using the learning strategy of Cruz et al. [31] and Pajares [4] to compute the matching probability for each pair of edge segments. The matching probabilities are used to start the relaxation process and then to obtain a compatibility measure when the smoothness and ordering constraints are violated, and (b) by using the minimum differential disparity criterion of Medioni and Nevatia [8] and the approach of Ruichek and Postaire [27] to map the smoothness and ordering constraints, respectively.

Therefore, our method integrates several criteria following the idea of Pavlidis [58] and we get that it works properly even if the images are complex.

1.6. Paper organization

The paper is organized as follows: in Section 2 the stereovision probabilistic relaxation matching scheme is described. In Section 3 we summarize the proposed matching procedure. The performance of the method is illustrated in Section 4, where a comparative study against existing global relaxation matching methods is carried out, and the necessity of using a relaxation matching strategy rather than a local matching technique is also made clear. Finally in Section 5 there is a discussion of some related topics.

2. Probabilistic relaxation scheme in stereovision matching

As mentioned earlier, this paper proposes a probabilistic relaxation technique for solving the stereovision matching problem. This technique is extended to work properly with complex images and integrates several criteria. In this section we deal with these issues.

2.1. Feature and attribute extraction

We believe, as in Medioni and Nevatia [8], that feature-based stereo systems have strong advantages over area-based correlation systems. However, detection of the boundaries is a complex and time-consuming scene-analysis task. The contour edges in both images are extracted using the Laplacian of Gaussian filter in accordance with the zero-crossing criterion [59,60]. For each zero crossing in a given image, its gradient vector, Laplacian and variance values are computed from the gray levels of a central pixel and its eight immediate neighbors. The gradient vector (magnitude and direction)

is computed as in Leu and Yau [61], the Laplacian value as in Lew et al. [62] and the variance value as in Krotkov [63]. To find the gradient magnitude of the central pixel we compare the gray-level differences of the four pairs of opposite pixels in the 8-neighborhood, and the largest difference is taken as the gradient magnitude. The gradient direction of the central pixel is the direction out of the eight principal directions whose opposite pixels yield the largest gray-level difference and also points in the direction in which the pixel gray level is increasing. We use a chain-code and we assign eight digits to represent the eight principal directions; these digits are integer numbers from 1 to 8. This approach allows the normalization of the gradient direction, so its values fall in the same range as the remainder of the attribute values. In order to avoid noise problems in edge detection that can lead to later mismatches in realistic images, the following two global consistent methods are used: (1) the edges are obtained by joining adjacent zero crossings following the algorithm of Tanaka and Kak [64]; a margin of deviation of $\pm 20\%$ in gradient magnitude and of $\pm 45^\circ$ in gradient direction is allowed; (2) each detected contour is approximated by a series of piecewise linear line segments [65]. Finally, for every segment, an average value of the four attributes is obtained from all computed values of its zero crossings. All average attribute values are normalized in the same range. Each segment is identified with its initial and final pixel coordinates, its length and its label.

Therefore, each pair of features has two associated four-dimensional vectors \mathbf{x}_l and \mathbf{x}_r , where the components are the attribute values, and the sub-indices l and r denote features belonging to the left and right images, respectively. A four-dimensional difference measurement vector \mathbf{x} is then also obtained from the above \mathbf{x}_l and \mathbf{x}_r vectors, $\mathbf{x} = \mathbf{x}_l - \mathbf{x}_r = \{x_m, x_d, x_b, x_v\}$. The components of \mathbf{x} are the corresponding differences for module and direction gradient, Laplacian and variance values,

respectively, associated to the same four attributes used in our approach.

Of all the possible combinations of pairs of matches formed by segments of the left and right images, only those pairs with a difference value in the direction of the gradient less than $\pm 45^\circ$, taking into account that they overlap will be processed. This is called the initial condition. The remaining pairs that do not meet such conditions are directly classified as *False* correspondences. The overlap is a concept introduced by Medioni and Nevatia [8], two segments u and v overlap if by sliding one of them in a direction parallel to the epipolar line, they would intersect. We note that a segment in one image is allowed to match more than one segment in the other image. Therefore, we make a provision for broken segments resulting in possible multiple correct matches. The following pedagogical example in Fig. 1 clarifies the above.

The edge segment u in the left image matches with the broken segment represented by v and w in the right image, but under the condition that v and w do not overlap and that their orientations, measured by their corresponding \mathbf{x}_l , do not differ by more than $\pm 10^\circ$.

2.2. Computation of the local matching probability

The local probabilities for all pairs of features that meet the above initial conditions are computed following a learning strategy [4,31]. This is briefly described below.

Due to the above-mentioned extrinsic and intrinsic factors, we have verified that the differences between the attributes for the true matches cluster in a cloud around a center and we have assumed an underlying probability density function (PDFt). The form of the PDFt is known. Our experience lead us to model this PDFt as a Gaussian one with two unknown parameters, the *mean* difference vector and the *covariance* matrix. The first is the center of

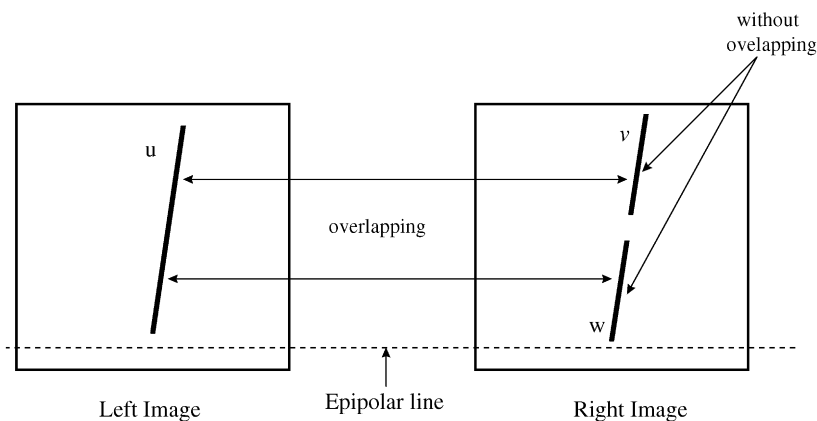


Fig. 1. Broken edge segment v and w may match with u .

the cluster and the second gives us information about the dispersion of the differences of the attributes in the cluster. Both parameters are estimated through a learning process based on the parametric Bayes classifier by a maximum likelihood method [4,66]. The learning process is carried out with a number of samples through a training process.

Once the PDFt is fully described, it allows us to compute the local matching probability associated with any new and current pair of edge segments using the differences of its attributes according to the following expression [4,31,66–68].

$$p(\mathbf{x}) = \frac{1}{|\Sigma|^{0.5}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (1)$$

where $\boldsymbol{\mu}$ and Σ are the corresponding learned cluster center, and the covariance matrix, respectively, \mathbf{x} is the four-dimensional difference measurement vector (see Section 2.1) and t stands for the transpose.

During the computation of the local probabilities, we explicitly apply the similarity constraint between features from the left and right images. These local probabilities are the starting point for the next relaxation matching process, i.e. initial probabilities for the global relaxation process.

2.3. A review of the relaxation labeling processes

Recent advances in the theory of probabilistic relaxation can be found in Refs. [1,69], although it has been studied in previous works [42,44,48,51,54]. The probabilistic relaxation process is made applicable to general matching problems. The matching problem is formulated in the Bayesian framework for contextual label assignment. The formulation leads to an evidence-combining formula which prescribes in a consistent and unified manner how unary attribute measurements relating to single entities, binary relation measurements relating to pairs of objects, and prior knowledge of the environment (initial probabilities) should be jointly brought to bear on the object labeling problem. The theoretical framework also suggests how the probabilistic relaxation process should be initialized based on unary measurements (according to Section 2.2). This contrasts with the approach adopted by Li [70] who started the iterative process from a random assignment of label probabilities.

A set A of N objects of the scene is to be matched, $A = \{a_1, a_2, \dots, a_N\}$. The goal is to match the scene to a model. Each object a_i has assigned a label θ_i , which may take as its value any of the M model labels that form the set $\Omega : \Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$. The notation ω_{θ_i} indicates that a model label is associated with a particular scene label θ_i . At the end of the labeling process, it is expected that each object will have one unambiguous label value. For convenience two sets of indices are defined:

$N_0 \equiv \{1, 2, \dots, N\}$ and $N_i \equiv \{1, 2, \dots, i-1, i+1, \dots, N\}$. For each object a_i a set of m_1 measurements \mathbf{x}_i is available, corresponding to the unary attributes of the object: $\mathbf{x}_i = \{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m_1)}\}$. The abbreviation $x_{i,i \in N_0}$ denotes the set of all unary measurement vectors x_i made on the set A of objects, i.e., $x_{i,i \in N_0} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$.

For each pair of objects a_i and a_j a set of m_2 binary measurements is available $A_{ij} : A_{ij} = \{A_{ij}^{(1)}, A_{ij}^{(2)}, \dots, A_{ij}^{(m_2)}\}$. The same classes of unary and binary measurements are also made on the model. After introduction of the above notations and following Christmas et al. [1] the theoretical framework of Bayesian probability for object labeling using probabilistic relaxation is as follows: the label θ_i of an object a_i will be given the value ω_{θ_i} , provided that it is the most probable label given all the information we have for the system, i.e. all unary measurements and the values of all binary relations between the various objects. By mathematical induction it can be shown that there is no need for the inclusion of all binary relations of all objects in order to identify an object, and the most appropriate label of object a_i is ω_{θ_i} , given by

$$\begin{aligned} P(\theta_i = \omega_{\theta_i} | x_{j, j \in N_0}, A_{ij, j \in N_i}) \\ = \max_{\omega_{\lambda} \in \Omega} P(\theta_i = \omega_{\lambda} | x_{j, j \in N_0}, A_{ij, j \in N_i}), \end{aligned} \quad (2)$$

where the upper-case P represents the probability of an event; thus $P(\theta_i = \omega_{\lambda})$ and $P(\theta_i = \omega_{\theta_i})$ denote the probabilities that scene label θ_i is matched to model labels ω_{λ} and ω_{θ_i} , respectively. Under certain often-adopted assumptions, the conditional probabilities in this equation can be expressed in a form that indicates that a relaxation updating rule would be an appropriate method of finding the required maximum. Using the Bayes's rule and after an exhaustive treatment (see Christmas et al. [1]) the desired solution to the problem of labeling, as defined by Eq. (2) can be obtained using an iterative scheme as follows:

$$P^{(n+1)}(\theta_i = \omega_{\theta_i}) = \frac{P^{(n)}(\theta_i = \omega_{\theta_i}) Q^{(n)}(\theta_i = \omega_{\theta_i})}{\sum_{\omega_{\lambda} \in \Omega} P^{(n)}(\theta_i = \omega_{\lambda}) Q^{(n)}(\theta_i = \omega_{\lambda})}, \quad (3)$$

where

$$\begin{aligned} Q^{(n)}(\theta_i = \omega_{\theta_i}) \\ = \prod_{j \in N_i} \sum_{\omega_{\beta} \in \Omega} P^{(n)}(\theta_j = \omega_{\beta}) p(A_{ij} | \theta_i = \omega_{\theta_i}, \theta_j = \omega_{\beta}), \end{aligned} \quad (4)$$

where $P^{(n)}(\theta_i = \omega_{\theta_i})$ is the matching probability at level (iteration) n that scene label θ_i is matched to model label ω_{θ_i} and $P^{(n+1)}(\theta_i = \omega_{\theta_i})$ is the updated probability of the corresponding match at level $n+1$. The quantity $Q^{(n)}(\theta_i = \omega_{\theta_i})$ expresses the support the match $\theta_i = \omega_{\theta_i}$ receives at the n th iteration step from the other objects in the scene, taking into consideration the binary relations

that exist between them and object a_i . The density function $p(A_{ij} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta)$ corresponds to the compatibility coefficients of other methods [42,44]; i.e. it quantifies the compatibility between the match $\theta_i = \omega_\alpha$ and a neighboring match $\theta_j = \omega_\beta$. The density function ranges from 0 to 1 and will be expressed as a compatibility coefficient in Eq. (6).

2.4. Stereo correspondence using relaxation labeling

At this stage there are two goals in mind: (a) to show how the relaxation labeling theory may be applied to stereovision matching problems and (b) to map the similarity, smoothness and ordering stereovision matching constraints in Eqs. (3) and (4). The uniqueness constraint will be considered later during the decision process in Section 4.3.

2.4.1. Stereovision matching by relaxation labeling

As explained in the introduction, the stereovision matching is the process of identifying the corresponding features in two images (left and right) that are cast by the same physical feature in 3-D space. In order to apply the relaxation labeling theory to the stereovision matching and without loss of generality, we associate the left stereo-image to the scene and the right one to the model. With this approach we have to match the scene to a model and the stereovision matching problem is now a relaxation matching one.

In our approach, the objects in the scene and the model are edge segments. Following the notation introduced in Section 2.3, the set A , of N objects, is identified with the set of N edge segments in the left image from the stereo pair, each left edge segment (object) is identified by its corresponding label θ_i . The edge segments from the right image are associated to the model labels that form the set Ω . Then the notation ω_θ , in relaxation labeling indicates now, that we wish to match a right edge segment with a particular left edge segment.

According to Section 2.3 the unary attributes m_1 , to map the similarity constraint are module and direction gradient, Laplacian and variance values, as explained in Section 2.1. Therefore m_1 takes a value of 4. The binary relations are: (a) the differential disparity of Ref. [8] where we apply the smoothness constraint, and (b) the ordering relation to apply the ordering constraint. Hence m_2 takes the value of 2. According to Section 1.3, we can infer that unary and binary measurements are associated to local and global correspondences, respectively.

2.4.2. Mapping the similarity, smoothness and ordering stereovision matching constraints

The next step consists in the mapping of the similarity, smoothness and ordering stereovision matching

constraints [20–23,27], into the relaxation labeling Eqs. (3) and (4). To achieve this we have made two important contributions to solve the stereovision matching problem: (a) by applying a learning strategy in the similarity constraint, and (b) by introducing specific conditions where we take into account the possible violation of the smoothness and ordering constraints (objects in the 3-D scene very close to the cameras and occlusions) and the problem of fixing a disparity limit (repetitive structures). We can also point out that unlike Christmas et al. [1], and due to the possible violation of the smoothness and ordering constraints, we have not considered structural or geometrical binary relations for mapping the smoothness constraint. Some of these binary relations can be found in Ref. [1] (angle between line segments, minimum distance between the endpoints of line segments, distance between the midpoints of line segments).

(a) *Mapping the similarity constraint*: The similarity constraint is mapped as follows: Given a pair of edge-segments, from left and right images, to be matched and labeled as θ_i and ω_λ , respectively, we have the four-dimensional difference measurement vector \mathbf{x}_i , which is obviously a measurement of similarity between the two edge segments involved. This vector is obtained from the corresponding unary measurement vectors associated to the left edge segment \mathbf{x}_l , and the right edge segment \mathbf{x}_r , as explained in Section 2.1, so $\mathbf{x}_i = \{x_{im}, x_{id}, x_{il}, x_{iv}\}$. The matching probability at level n , left-hand side of Eq. (1), that the left edge-segment θ_i is matched to the right edge-segment ω_λ is now $P^{(n)}(\theta_i = \omega_\lambda)$ and \mathbf{x} is now \mathbf{x}_i

(b) *Mapping the smoothness constraint*: Assuming no violation of the smoothness constraint and unlike in Christmas et al. [1] the density function $p(A_{ij} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta)$ in the support of the match, given in Eq. (4), is replaced by a compatibility coefficient where the smoothness and ordering constraints are mapped. This is the idea proposed by Hummel and Zucker [44] and Rosenfeld et al. [42]. Traditional interpretations of compatibility coefficients have been in terms of statistical measures such as correlation [42] or structural information [1]. In this paper we define a new compatibility coefficient.

However, beforehand we change the notation for simplicity. So, given two pairs of matched edge segments (θ_i, ω_j) and (θ_h, ω_k) . Throughout the remainder of the paper the pairs of edge segments (θ_i, ω_j) and (θ_h, ω_k) are denoted as (i, j) and (h, k) respectively, and the matching probabilities $P^{(n)}(\theta_i = \omega_j)$ and $P^{(n)}(\theta_h = \omega_k)$ as $P_{ij}^{(n)}$ and $P_{hk}^{(n)}$, respectively. We measure the consistency between (i, j) and (h, k) by means of the compatibility coefficient $c(i, j; h, k)$ which ranges in the interval $[0, 1]$.

According with Medioni and Nevatia [8], the smoothness constraint assumes that neighboring edge segments have similar disparities, except at a few depth discontinuities [12]. Generally, to define the neighboring region a

bound is assumed on the disparity range allowed for any given segment. This limit is denoted as $maxd$, fixed at eight pixels in this paper. For each edge segment “ i ” in the left image we define a window $w(i)$ in the right image in which corresponding segments from the right image must lie and, similarly, for each edge segment “ j ” in the right image, we define a window $w(j)$ in the left image in which corresponding segments from the left image must lie. It is said that an edge segment “ h ” must lie if at least the 30% of the length of the segment “ h ” is contained in the corresponding window. The shape of these windows is a parallelogram, two sides are fixed by “ i ” or “ j ” and the other two are lines of length $2 \cdot maxd$ (see Fig. 3). The smoothness constraint implies that “ i ” in $w(j)$ implies “ j ” in $w(i)$.

Now, given “ i ” and “ h ” in $w(j)$ and “ j ” and “ k ” in $w(i)$ where “ i ” matches with “ j ” and “ h ” with “ k ” the differential disparity $|d_{ij} - d_{hk}|$, to be included later in $c(i, j; h, k)$, measures how close the disparity between edge segments “ i ” and “ j ” denoted as d_{ij} is to the disparity d_{hk} between edge segments “ h ” and “ k ”. The disparity between edge segments is the average of the disparity between the two edge segments along the length they overlap. This differential disparity criterion is used in Refs. [8,27,36,55,56] among others.

(c) *Mapping the ordering constraint:* Assuming, as in the smoothness constraint, no violation of the ordering constraint, we extend the concept defined in Ruicheck and Postaire [27] for edge points to edge segments. As described in Ref. [27], if an edge point “ i ” in the left image is matched with an edge point “ j ” in the right image, then it is not possible for an edge point “ h ” in the left image, such that $x_h < x_i$, to be matched with an edge point “ k ” in the right image for which $x_j > x_k$, where x denotes the x -coordinate in a Cartesian system with its origin in the bottom left corner of each image.

We define the ordering constraint coefficient O_{ijhk} for the edge segments as follows:

$$O_{ijhk} = \frac{1}{N} \sum_N o_{ijhk} \quad \text{where} \quad o_{ijhk} = |S(x_i - x_h) - S(x_j - x_k)|$$

$$\text{and} \quad S(r) = \begin{cases} 1 & \text{if } r > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

this upper-case O_{ijhk} coefficient measures the relative position of edge segments “ i ” and “ h ” in the left image with respect to “ j ” and “ k ” in the right image. The lower-case o_{ijhk} coefficient is defined in Ref. [27] for edge points taking the discrete values of 0 and 1. O_{ijhk} is the ordering average measure for edge points along the common overlapping length for the four edge segments, it ranges from 0 to 1 and is included later in the compatibility coefficient $c(i, j; h, k)$. We trace N scanlines along the common overlapping length, each scanline produces a set of four intersection points with the four edge-segments. With this intersection points lower-case o_{ijhk} is computed. Fig. 2 clarifies the above.

(d) *Problems to apply the smoothness and ordering constraints:* During our experiments we have found the following problems related with these constraints, namely:

- (1) Violation of such constraints in systems with parallel geometry due to the presence of objects close to the cameras and occlusions. This type of objects is particularly dangerous and must be considered immediately in order to avoid, for example, undesired collisions in navigation systems.
- (2) The fixation of a bound on the disparity is a hard task, because one can never be sure which is the best limit. In images with repetitive structures this problem is more evident, because an edge segment in a window does not imply that its couple is in the other window.

As mentioned in Section 1.5, we call complex images the images in which these problems appear and we try to solve them by introducing a specific condition in the

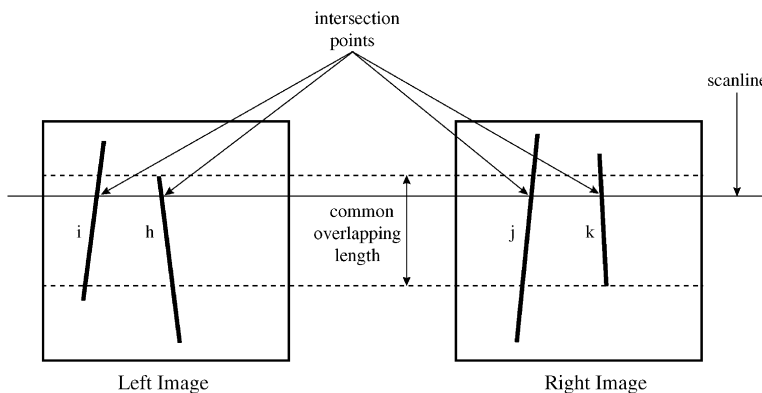


Fig. 2. Common overlapping length and intersection points produced by a scanline.

compatibility coefficient, as we will see later. So, our system works properly with complex and no complex images.

(e) *Definition of the compatibility coefficient:* In our stereovision relaxation labeling model, the probabilities computed according to Eq. (3), are used to determine the various possible correspondences between edge segments. The goal of the relaxation process is to increase the consistency of a given pair of edge segments among the constraints so that the probability of a correct match can be increased and the probability of any incorrect match can be decreased during each iteration. Therefore, we look for a compatibility coefficient, which must be able to represent the consistency between the current pair of edge segments under correspondence and the pairs of edge segments in a given neighborhood.

When no violation of the smoothness or ordering constraints is considered, the compatibility coefficient makes global consistency between neighbor pairs of edge segments based on such constraints. This is usual in stereovision matching systems and is made by condition C_1 .

When violation of the smoothness and ordering constraints is assumed, the global consistency is lost and such constraints cannot be applied, we must fall back on to local consistency, as this is the unique resource available. The local consistency is represented by the similarity constraint, that is by the matching probability, but now considering the corresponding neighbors edge segments. This is made through condition C_2 .

The compatibility coefficient $c(i, j; h, k)$ is finally defined as follows:

$$c(i, j; h, k) = \begin{cases} \frac{O_{ijk} \delta_{ijk}}{l + |d_{ij} - d_{hk}|} & \text{if } C_1, \\ \delta_{ijk} \frac{P_{ij}^{(n)} + P_{hk}^{(n)}}{2} & \text{if } C_2, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where $\delta_{ijk} = (\text{overlap}(i, j) + \text{overlap}(h, k))/2$, which is an overlap function in the interval $[0, 1]$ and measures the common average overlap rates of the edge segments under matching (see Section 2.1); the term d_{ij} is the average disparity, which is the average of the disparity between the two segments “ i ” and “ j ” along the common overlap length and $P_{ij}^{(n)}$ is the matching probability of edge segments “ i ” and “ j ” at level n . O_{ijk} is the order coefficient.

Condition C_1 : Assumes the mapping of the ordering and smoothness constraints. The assumption that the ordering constraint is not violated is simple, and only requires that O_{ijk} is greater than a fixed threshold, set to 0.85 in this paper. The assumption that the smoothness constraint is not violated is based on the following criterion. If the edge segment “ h ” is a neighbor of the edge

segment “ i ” (h in $w(j)$) its preferred match “ k ” is a neighbor of the edge segment “ j ” (k in $w(i)$). This is equivalent to say that both pairs of edge segments have similar disparities, this is the minimum differential disparity criterion in Ref. [8]. Given a segment “ h ” in the left image (LI) and a segment “ k ” in the right image (RI), we define a preferred match “ k ” of “ h ”, denoted as $pm(h, k)$ if $P_{hk}^{(n)} \geq A * P_{\max_{hk'}^{(n)}} \forall k' \in RI$. Where $P_{hk}^{(n)} > T$ (with the threshold T set to 0.5 in this experiment) and $P_{\max_{hk'}^{(n)}} = \max_{k' \in RI} P_{hk'}^{(n)}$ is the maximum matching probability of the edge segment “ h ” with any edge segment “ k ” in RI . The constant A is a coefficient to fix the degree of exigency to define the preferred match (set to 0.85 in this paper). Hence, the compatibility coefficient is maximum if the differential disparity is minimum and O_{ijk} and the overlap function are maximum and vice versa.

Condition C_2 : Unlike C_1 this condition assumes the violation of the smoothness or ordering constraints (due to objects close to the cameras or occlusions) or the possibility that an edge segment “ h ” in $w(j)$ has a preferred match “ k ” out of $w(i)$ (repetitive structures). C_2 is only considered when C_1 is false. The loss of global consistency is replaced by local consistency in the compatibility coefficient as the average of the matching probabilities where the overlap function is also taken into account. We must point out that this condition C_2 overrides the mapping of the ordering and smoothness constraints, to take up again the similarity constraint through the matching probabilities, so that the pair of edge segments under correspondence is not penalized during the relaxation process due to violation of the smoothness and ordering constraints. The compatibility coefficient is maximum if the matching probabilities and overlap function are maximum and vice versa.

(f) *A pedagogical example:* The following is a simple example to clarify the above concepts related with conditions C_1 and C_2 , where we consider the evaluation of a pair of edge segments (i, j) , (Fig. 3):

- (1) The preferred match of “ h ” is “ k ”, where $h \in w(j)$ and $k \in w(i)$; ordering and smoothness are not violated.
- (2) The preferred match of “ h ” is “ l ”, hence the ordering constraint is preserved but the smoothness constraint is violated because $l \notin w(i)$.
- (3) The preferred match of “ h ” is “ m ”, hence the ordering constraint is violated but the smoothness constraint is preserved because $m \in w(i)$.
- (4) The preferred match of “ h ” is “ n ”, in which case both ordering and smoothness constraints are violated

In the first case condition C_1 reinforces the match (i, j) as expected, but in the remainder of the cases the reinforcement is assumed by condition C_2 . This is the appropriate process. Without condition C_2 the match (i, j) should be penalized and this is unwanted as the edge

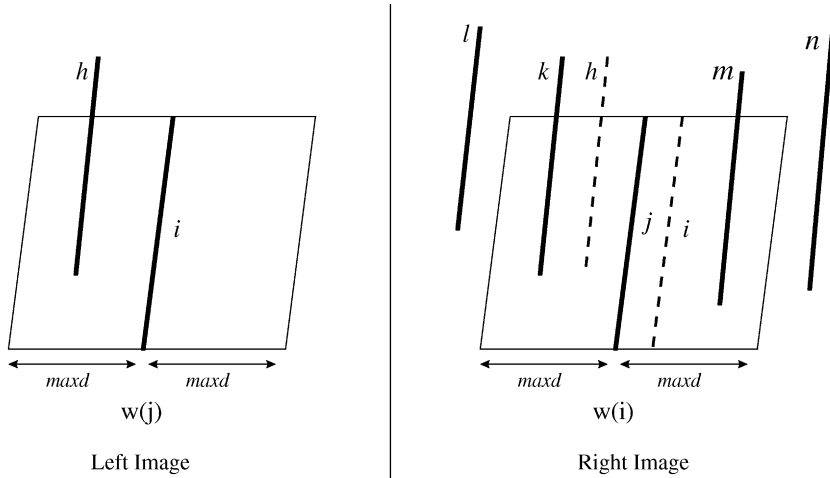


Fig. 3. Evaluation of the pair (i, j) considering its neighborhood.

segments in the right image are all preferred matches of “ h ”. The dotted lines in the right image represent the positions of “ i ” and “ h ” when the left image is superimposed to the right one.

3. Summary of the matching procedure

After mapping the similarity, ordering and smoothness stereovision matching constraints, the correspondence process is achieved by allowing the system to evolve so that it reaches a stable state, i.e. when no changes occur in the matching probabilities during the updating procedure.

The whole matching procedure can be summarized as follows:

- (1) The system is organized with a set of pairs of edge segments (from the left and right images) with a difference in the direction of the gradient of less than $\pm 45^\circ$ and considering that they overlap (initial conditions in Section 2.1). Each pair of edge segments is characterized by its initial matching probability.
- (2) The assignment of the initial matching probabilities are computed according to Eq. (1).
- (3) The variable $npair$, which represents the number of pairs of edge-segments for which the matching probabilities are modified by the updating procedure, is set to 0.
- (4) The pairs of edge segments are sequentially selected and the relaxation procedure applied to each one.
- (5) At the n th iteration, the matching probability is computed according to Eq. (3).
- (6) If $|P^{(n+1)}(\theta_i = \omega_{\theta_i}) - P^{(n)}(\theta_i = \omega_{\theta_i})| > \varepsilon$ then $npair$ is increased. This is the criterion used in Ref. [1], so that for each object a_i there is one label ω_{θ_i} for which

the matching probability at a given level is within some small distance ε from unity. The constant value ε is introduced to accelerate the convergence and is set to 0.01 in this experiment.

- (7) If $npair$ is not equal to 0 then go to step 3.
- (8) *Matching decisions*: A left edge segment can be assigned to a unique right edge segment (unambiguous pair) or several right edge segments (ambiguous pairs). The decision about whether a match is correct or not correct is made by choosing the greater probability value (in the unambiguous case there is only one) whenever it surpasses a previous fixed probability threshold T , set to 0.5. The ambiguities under the criterion given in Section 2.1 are allowed as they can be produced by broken edge segments. This step represents the mapping of the *uniqueness* constraint, which completes the set of matching constraints in stereovision matching.

4. Validation, comparative analysis and performance evaluation

4.1. Design of a test strategy

In order to assess the validity and performance of the proposed method (i.e. how well our approach works in images with and without additional complexity), we have selected 48 stereo pairs of realistic stereo images from an indoor environment. Each pair consists of two left and right original images and two left and right images of labeled edge segments. All tested images are 512×512 pixels in size, with 256 gray levels. The two cameras have equal focal lengths and are aligned so that their viewing direction is parallel. The plane formed by the viewed

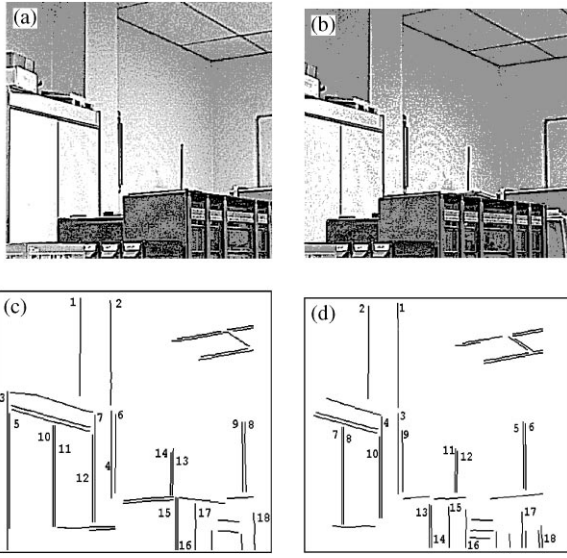


Fig. 4. Group SP1: (a) original left stereo image, (b) original right stereo image, (c) labeled segments left image, (d) labeled segments right image.

point and the two focal points intersects the image plane along a horizontal line known as the epipolar line. A grid pattern of parallel horizontal lines is used for camera alignment. The cameras are adjusted so that the images of these lines coincide [71]. Thus, each point in an edge segment lies along the same horizontal or epipolar line crossing both images of the stereo pair. The 48 stereo pairs are classified into three groups: SP1, SP2 and SP3 with 18, 18 and 12 pairs of stereo images, respectively. For space reasons a unique representative stereo pair is shown for each SP set.

Group SP1 consists of stereo images without apparent complexity. Figs. 4a–d (original images and labeled edge segments as stated) show a stereo pair representative of SP1. Group SP2 corresponds to scenes where a repetitive structure has been captured, Figs. 5a–d, show a stereo pair representative of SP2, where in this case the repetitive structure is provided by the vertical books in the foreground of the scene. Finally, group SP3 contains objects close to the cameras, that produce a high range of disparity violating the smoothness and ordering constraints. Figs. 6a–d is a stereo pair representative of SP3, where we can see the object labeled as 9, 10 in left image and 11, 12 in right image as a characteristic example of an object close to the cameras occluding the edge segment 19 in the right image. Although this last type of images is unusual, its treatment is very important as they could produce fatal errors in navigation systems for example, where the nearest objects must be processed immediately.

As mentioned during the introduction, we base our method on the relaxation work of Christmas et al. [1], in which a stereovision matching problem is solved. This

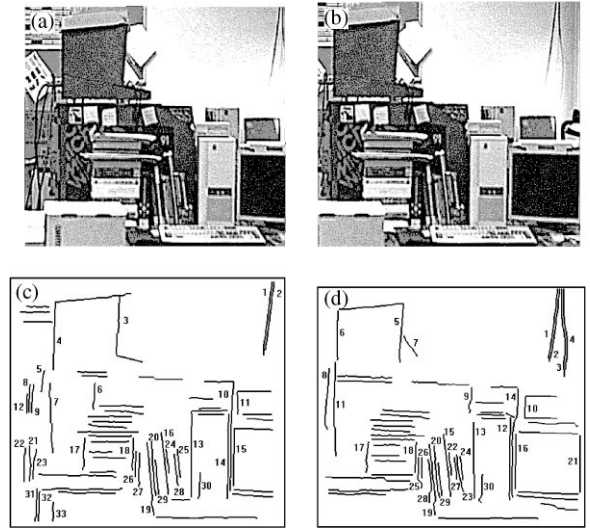


Fig. 5. Group SP2: (a) original left stereo image, (b) original right stereo image, (c) labeled segments left image, (d) labeled segments right image.

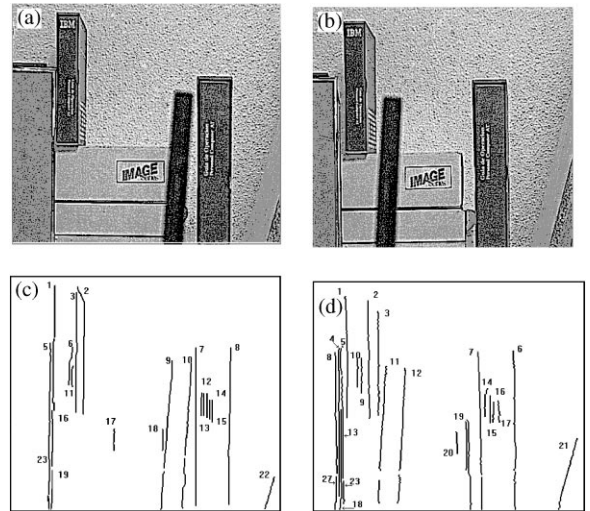


Fig. 6. Group SP3: (a) original left stereo image, (b) original right stereo image, (c) labeled segments left image, (d) labeled segments right image.

algorithm uses straight-line segments as features for the matching, a unique unary measurement of each line segment (orientation of the scene relative to the model) and four binary relations. These binary relations are: (a) the angle between line segments a_i and a_j ; (b) the angle between segment a_i and the line joining the center-points of segments a_i and a_j ; (c) the minimum distance between the endpoints of line segments a_i and a_j ; and (d) the

distance between the midpoints of line segments a_i and a_j . It is a structural matching method based on geometrical relations. These are the fundamental and main differences with regard to our approach. Indeed, we use four unary measurements (module and direction gradient, Laplacian and variance) to map the similarity constraint and two binary relations to map the smoothness and ordering constraints. Only the direction gradient in our approach could be identified with the unary measurement in Ref. [1]. These reasons impede us from performing the appropriate comparative analysis between our approach and the structural method in Ref. [1].

Moreover, the geometrical relationships would be only applicable with great difficulty to our complex images. According to Section 1.4, there are different relaxation approaches: probabilistic, merit and optimization. This paper develops a probabilistic relaxation (PR) method. To compare our PR technique, we choose a merit relaxation and two optimization relaxation methods.

4.1.1. The merit relaxation approach (MR)

The minimum differential disparity algorithm of Medioni and Nevatia [8] is the merit relaxation method chosen for comparison purposes for the following reasons: (1) it applies the commonly used constraints, similarity, smoothness and uniqueness; (2) it uses edge segments as features and the contrast and orientation of the features as attributes; (3) it is implied in the derivation of PR; and (4) Medioni and Nevatia conclude that several improvements to the matching method may be possible. A brief description of the MR algorithm is given below.

MR uses straight-line segments as features because they believe that the feature-based stereo systems have strong advantages over area-based correlation systems. The method defines two windows around every segment in the left and right images and assumes a bound on the disparity range allowed. We have applied these concepts to our algorithm and they are exhaustively explained in Section 2.4.2.

The MR local matching process defines a boolean function indicating whether two segments are potential matches, such a function is true iff the segments overlap and have similar contrast and orientation (similarity constraint). The overlap concept in PR is also taken from this method and the contrast and orientation could be identified with our module and direction gradient, although we believe that gradient provides better performance. To each pair of segments, the MR associates an average disparity which is the average of the disparity between the two involved segments along the length of their overlap. This is also applied in our method.

Thus, a global relaxation matching process is triggered, where a merit function is updated iteratively, and changes if the segments in the neighborhood are assigned new preferred matches. The preferred matches are selected according to the boolean function and the neigh-

borhood concept, this latter defined on the basis of the disparity bound. The MR applies the smoothness constraint through the minimum differential disparity criterion. We have also applied this criterion in PR.

The fundamental differences between MR and PR can be summarized as follows: (1) PR uses quantitative probability values (matching probabilities) and MR uses a qualitative boolean function; (2) PR introduces specific conditions to overcome the possible violation of the smoothness constraint and to avoid the problem of the fixation of the disparity limit, which is missing in MR.

4.1.2. Two optimization relaxation approaches (OR1 and OR2)

Two appropriate techniques to compare our PR are the approaches based on the Hopfield neural network [36] (OR1) and [27] (OR2). These techniques take into account other previous works in this area [12,14,24,55, 56]. They are chosen for the following reasons: (1) in OR1 the similarity and smoothness constraints are mapped as in PR, moreover both use the same set of attributes and features, (2) the compatibility coefficient in OR1 only takes into account the violation of the smoothness constraint, whereas in PR the violation of both the smoothness and ordering constraints are considered, and (3) although OR2 uses edge pixels as features it is considered as a guideline to map the similarity, smoothness and ordering constraints for edge segments, this is made as in PR but removing the specific conditions C_1 and C_2 , that is, with respect to C_1 the preferred match “ k ” of “ h ” fulfills $P_{hk}^{(n)} = P_{\max_{hk'}}^{(n)} \forall k' \in RI$ and with respect to C_2 the violation of the smoothness and ordering constraints is not considered.

The comparison between PR and OR1 allows us to check the effectiveness of the ordering constraint because they use a similar compatibility coefficient and the comparison between PR and OR2 allows us to prove the effectiveness of the compatibility coefficient itself, and more specifically when the smoothness and ordering constraints are violated.

4.2. Experimental results

In this study, the proposed PR method for solving the stereovision matching problem was executed on a VAX 8810 workstation. During the test, 48 relaxation processes were built-in and executed, one for each stereo image pair. The number of pairs of edge segments in the different relaxation processes is variable and ranges from 36 to 104. This variability depends on the number of edge segments extracted and the initial conditions established in Section 2.1.

Eq. (3) describes the evolution of the matching probabilities for our stereo global matching approach. Fig. 7 shows the number of pairs (n_{pair}) of edge segments for

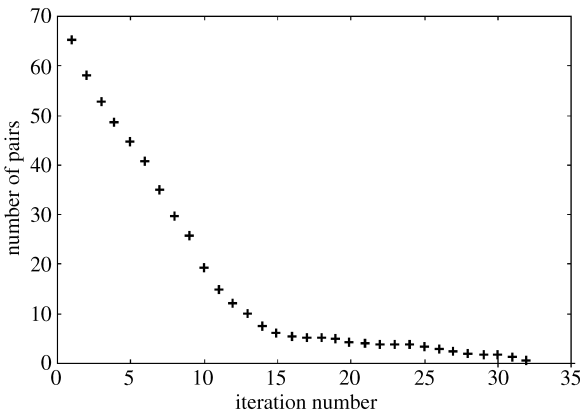


Fig. 7. Number of pairs ($npair$) of edge segments for which the matching probabilities are modified by the updating procedure as a function of the iteration number, averaged for all relaxation processes.

Table 1
Percentage of successes for the group of stereo-pairs SP1

SP1	0	8	16	24	32
PR	79.1	86.1	91.2	94.3	97.3
OR1	79.1	84.3	88.7	92.9	96.2
OR2	79.1	80.3	83.5	87.6	89.9
MR	79.1	79.8	82.6	85.2	86.7

which the matching probabilities are modified by the updating procedure as a function of the iteration number. This number is averaged for the 48 executed relaxation processes. It is computed according to point 6 in Section 3.

The number of pairs that change their matching probabilities drops in a relatively smooth fashion, with the maximum number of pairs that change between an iteration and the subsequent one being 9. This means that the number of correct matches grows slowly during the global matching process, and this is due to the following two reasons: (a) initially (iteration 0) there is a relatively high percentage of true matches (as we will see later in Table 1), this is a direct consequence of the local stereovision matching process used, and is the same idea that the used in Christmas et al. [1], where the initial probabilities are computed based on the unary measurements, and (b) the coefficient ε is introduced in Section 3, point 6, so that the criterion to determine whether $npair$ changes is relaxed. Finally, we can see that from iteration number 15, $npair$ varies slowly and it can be assumed that all analyzed stereo images reach an acceptably high degree of performance at iteration 15.

We can point out that due to the variability in the number of pairs of edge segments to be matched for each

stereo image and to different complexity of the stereo images the time for convergence is also variable. We have verified in all tested stereo images that an average time for convergence is about 7 min. This time is measured from iterations 0 to 32, the start and end of the global relaxation process, which is the objective of this paper.

4.3. Comparative analysis

As we already know, the stereo-matching process can be split into local and global processes. This paper develops a global relaxation process, probabilistic based, which is now of interest to us, since we have already exhaustively studied the local process in previous works [4,31,32]. Therefore, the comparative analysis we perform is only concerned with the relaxation process.

The system processes the SP1, SP2 and SP3 groups of stereo images. Of all the possible combinations of pairs of matches formed by segments of left and right images only 468, 359 and 294 pairs are considered for SP1, SP2 and SP3, respectively, because the remainder of the pairs do not meet the initial conditions.

For each pair of features we compute the local matching probability following the learning procedure described in Section 2.2. This is the starting point for all relaxation processes. For PR, OR1 and OR2 the local matching probabilities are the initial matching probabilities. MR requires the definition of a boolean function to indicate whether two segments are potential matches. We have chosen the criterion that if the local matching probability is greater than 0.5, the intermediate probability value in $[0,1]$, the pair of segments is a potential match and the boolean function is true.

The computed results are summarized in Tables 1–3. They show the percentage of successes for each group of stereo-pairs (SP1, SP2 and SP3) and for each method (PR, OR1, OR2 and MR) as a function of the number of iterations. Iteration 0 corresponds with results for local matching probabilities, the starting point for the relaxation process. As shown in Fig. 7, 32 iterations suffice to halt the PR process. This is the number of iterations used for the remainder of the methods. We have performed experiments with window-width values ($maxd$, see Section 2.4b) in the range 3–15 pixels, where the best results were obtained when $maxd$ was set to 8 pixels.

Table 2
Percentage of successes for the group of stereo-pairs SP2

SP2	0	8	16	24	32
PR	61.3	75.6	84.8	90.1	96.7
OR1	61.3	72.7	81.9	88.9	95.3
OR2	61.3	67.8	70.4	76.9	80.3
MR	61.3	66.5	69.8	73.7	77.2

Table 3
Percentage of successes for the group of stereo-pairs SP3

SP3	0	8	16	24	32
PR	62.8	78.3	87.9	91.4	95.7
OR1	62.8	76.5	86.8	90.7	95.1
OR2	62.8	64.6	67.1	69.3	70.5
MR	62.8	62.1	61.6	60.7	59.8

4.3.1. Decision process

When the relaxation processes have finished, there are still both unambiguous pairs of segments and ambiguous ones, depending on whether one and only one, or several right image segments correspond to a given left image segment. In any case, the decision about whether a match is correct or not is made by choosing the result of greater probability for PR, OR1 and OR2 and the best merit value for MR. In the unambiguous case there is only one. For PR, OR1 and OR2, it is necessary that this probability surpasses the previous fixed threshold probability ($T = 0.5$).

It is allowed that an edge segment l in the left image matches with r_1 and r_2 in the right image if r_1 and r_2 do not overlap, their directions differ a value smaller than a fixed threshold (set to 10 in this paper) and the matching probabilities between l and r_1 and l and r_2 are greater than T .

According to values from Tables 1–3, the following conclusions may be inferred:

(1) *Group SP1*: All methods progress in a similar fashion in the number of successes. The best results are obtained with PR and OR1. In both cases this is due to the definition of C_1 in the compatibility coefficient.

(2) *Group SP2*: The best results are once again obtained with PR and OR1 where the difference with respect to OR2 and particularly with respect to MR are significant. The explanation can be found by the presence of repetitive structures in the images. These structures produce false matches, which are taken as true by MR and OR2 and false by PR and OR1, hence PR and OR1 work as expected. The reason is the definition of condition C_2 applied to the compatibility coefficient.

(3) *Group SP3*: The number of successes for PR and OR1 evolves as expected whereas for OR2 remains practically stable and for MR decreases as the number of iterations increases. This last case is a surprising result. This is a direct consequence of images with some objects close to the cameras violating the smoothness and ordering constraints (in Fig. 3a–d, see object bounded by labels 9 and 10 in left image and 11 and 12 in right image). MR increases the merit of false matches, such as (3, 3) or (2, 2), and PR and OR1 handle this phenomenon correctly through condition C_2 .

In general, the best results are always obtained with PR. The small difference with respect to OR1 is due to the convergence criterion used in the Hopfield neural network, which requires a greater number of iterations (about 22) than PR to obtain a similar percentage of successes. We have verified that the ordering constraint is not relevant, this is because the impact of such a constraint is assumed by the smoothness constraint. We think that this is the reason for which the ordering constraint is not considered in classical stereovision matching methods [8,14,18,20,23,55] among others. Nevertheless, it is useful in structural or geometric matching approach methods [1,30,37].

(4) The difference in the results between PR and OR2 tells us that the definition of the compatibility coefficient with its two conditions is appropriate to deal with the violation of the smoothness and ordering constraints and also to fix the disparity limit.

(5) Additionally, we can point out that the global relaxation matching process substantially improves the local matching results (compare results between iteration 0 and 32). Hence the relaxation process becomes particularly necessary for complex stereo images.

5. Concluding remarks

An approach to stereo correspondence using a probabilistic relaxation method is presented. The stereo correspondence problem is formulated as a relaxation task where a matching probability is updated through the mapping of four constraints (similarity, smoothness, ordering and uniqueness). The similarity constraint is doubly applied: (a) during the computation of the local matching probability (see Section 2.2), and (b) during the mapping of the corresponding constraint (see Section 2.4a). We have also introduced, through the compatibility coefficient of Eq. (6), specific conditions to overcome the violation of the smoothness and ordering constraints and to avoid the serious problem caused by the setting of the disparity limit. The advantage of using a relaxation process is that a global match is automatically achieved, taking into account each stereo pair of edge segments to be matched: (a) in an isolated fashion (unary measurements), and (b) as a member of a neighborhood (binary measurements).

A comparative analysis is performed against other methods OR1, OR2 and MR. We have shown that the proposed method PR gives a better performance. We first started our experiments with normal images and then used more complex images with repetitive structures, objects violating the smoothness and ordering constraints and occlusions, always obtaining satisfactory results. We have also shown the necessity of using a global relaxation strategy.

We hope that the above results show the potential of PR. However, our method does not yet provide a full performance as it does not reach the 100% success level and several improvements to the matching method may be possible. We are currently investigating some of these: (1) hierarchical processing, starting at region level and (2) using corners as complementary features [54].

References

- [1] W.J. Christmas, J. Kittler, M. Petrou, Structural matching in computer vision using probabilistic relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (8) (1995) 749–764.
- [2] A.R. Dhond, J.K. Aggarwal, Structure from stereo – a review, *IEEE Trans. Systems Man Cybernet.* 19 (1989) 1489–1510.
- [3] T. Ozanian, Approaches for stereo matching – a review, *Modeling Identification Control* 16 (2) (1995) 65–94.
- [4] G. Pajares, Estrategia de solución al problema de la correspondencia en visión estereoscópica por la jerarquía metodológica y la integración de criterios, Ph.D. Thesis, Facultad de C.C. Físicas, U.N.E.D., Department of Informática y Automática, Madrid, 1995.
- [5] S. Barnard, M. Fishler, Computational stereo, *ACM Comput. Surveys* 14 (1982) 553–572.
- [6] K.S. Fu, R.C. González, C.S.G. Lee, *Robótica: Control, Detección, Visión e Inteligencia*, McGraw-Hill, Madrid, 1988.
- [7] D.H. Kim, R.H. Park, Analysis of quantization error in line-based stereo matching, *Pattern Recognition* 8 (1994) 913–924.
- [8] G. Medioni, R. Nevatia, Segment based stereo matching, *Comput. Vision Graphics Image Process.* 31 (1985) 2–18.
- [9] H.H. Baker, Building and Using Scene Representations in Image Understanding, AGARD-LS-185 Machine Perception. 1982, 3.1–3.11.
- [10] I.J. Cox, S.L. Hingorani, S.B. Rao, B.M. Maggs, A maximum likelihood stereo algorithm, *Comput. Vision Image Understanding* 63 (2) (1996) 542–567.
- [11] P. Fua, A Parallel Algorithm that produces dense depth maps and preserves image features, *Mach. Vision Appl.* 6 (1993) 35–49.
- [12] J.J. Lee, J.C. Shim, Y.H. Ha, Stereo correspondence using the Hopfield neural network of a new energy function, *Pattern Recognition* 27 (11) (1994) 1513–1522.
- [13] Y. Shirai, *Three-Dimensional Computer Vision*, Springer, Berlin, 1983.
- [14] Y. Zhou, R. Chellappa, *Artificial Neural Networks for Computer Vision*, Springer, New York, 1992.
- [15] W.E.L. Grimson, Computational experiments with a feature-based stereo algorithm, *IEEE Trans. Pattern Anal. Mach. Intell.* 7 (1985) 17–34.
- [16] A. Khotanzad, A. Bokil, Y.W. Lee, Stereopsis by constraint learning feed-forward neural networks, *IEEE Trans. Neural Networks* 4 (1993) 332–342.
- [17] Y.C. Kim, J.K. Aggarwal, Positioning three-dimensional objects using stereo images, *IEEE J. Robot. Automat.* 3 (1987) 361–373.
- [18] V.R. Lutsiv, T.A. Novikova, On the use of a neurocomputer for stereovision processing, *Pattern Recognition Image Anal.* 2 (1992) 441–444.
- [19] D. Maravall, E. Fernandez, Contribution to the matching problem in stereovision, *Proceedings of the 11th IAPR: International Conference in Pattern Recognition*, The Hague, 1992, pp. 411–414.
- [20] D. Marr, *La Vision*, Alianza Editorial, Madrid, 1985.
- [21] D. Marr, *Vision*, Freeman, San Francisco, 1982.
- [22] D. Marr, T. Poggio, A computational theory of human stereovision, *Proc. Roy. Soc. London B* 207 (1979) 301–328.
- [23] D. Marr, T. Poggio, Cooperative computation of stereo disparity, *Science* 194 (1976) 283–287.
- [24] M.S. Mousavi, R.J. Schalkoff, ANN Implementation of stereo vision using a multi-layer feedback architecture, *IEEE Trans. Systems Man Cybernet.* 24 (8) (1994) 1220–1238.
- [25] P. Rubio, Análisis comparativo de Métodos de Correspondencia estereoscópica, Ph.D. Thesis, Facultad de Psicología, Universidad Complutense, Madrid, 1991.
- [26] P. Rubio, RP: un algoritmo eficiente para la búsqueda de correspondencias en visión estereoscópica, *Informática y Automática* 26 (1993) 5–15.
- [27] Y. Ruycheck, J.G. Postaire, A neural network algorithm for 3-D reconstruction from stereo pairs of linear images, *Pattern Recognition Lett.* 17 (1996) 387–398.
- [28] N. Ayache, B. Faverjon, Efficient registration of stereo images by matching graph descriptions of edge segments, *Int. J. Comput. Vision* 1 (1987) 107–131.
- [29] H.H. Baker, T.O. Binford, Depth from edge and intensity based stereo, in: *Proceedings of the 7th International Joint Conference Artificial Intelligence*, Vancouver, Canada, 1981, pp. 631–636.
- [30] K.L. Boyer, A.C. Kak, Structural Stereopsis for 3-D vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 10 (2) (1988) 144–166.
- [31] J.M. Cruz, G. Pajares, J. Aranda, A neural network approach to the stereovision correspondence problem by unsupervised learning, *Neural Networks.* 8 (5) (1995) 805–813.
- [32] J.M. Cruz, G. Pajares, J. Aranda, J.L.F. Vindel, Stereo matching technique based on the perceptron criterion function, *Pattern Recognition Letters.* 16 (1995) 933–944.
- [33] W. Hoff, N. Ahuja, Surface from Stereo: Integrating feature matching, disparity estimation, and contour detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (1989) 121–136.
- [34] D.H. Kim, W.Y. Choi, R.H. Park, Stereo matching technique based on the theory of possibility, *Pattern Recognition Lett.* 13 (1992) 735–744.
- [35] Y. Ohta, T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. Pattern Anal. Mach. Intell.* 7 (2) (1985) 139–154.
- [36] G. Pajares, J.M. Cruz, J. Aranda, Relaxation by Hopfield network in stereo image matching, *Pattern Recognition* 31 (5) (1998) 561–574.
- [37] L.G. Shapiro, R.M. Haralick, Structural descriptions and inexact matching, *IEEE Trans. Pattern Anal. Machine Intell.* 3 (5) (1981) 504–519.

- [38] M.S. Wu, J.J. Leou, A bipartite matching approach to feature correspondence in stereo vision, *Pattern Recognition Letters* 16 (1995) 23–31.
- [39] D.M. Wuescher, K.L. Boyer, Robust contour decomposition using a constraint curvature criterion, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (1) (1991) 41–51.
- [40] R.M. Haralick, L.G. Shapiro, *Computer and Robot Vision*, Vols. I and II, Addison-Wesley, Reading, MA, 1992, 1993.
- [41] M. Sonka, V. Hlavac, R. Boyle, *Image Processing, Analysis and Machine Vision*, Chapman & Hall, London, 1995.
- [42] A. Rosenfeld, R. Hummel, S. Zucker, Scene labelling by relaxation operation, *IEEE Trans. Systems Man Cybernet.* 6 (1976) 420–453.
- [43] S.T. Barnard, W.B. Thompson, Disparity analysis of images, *IEEE Trans. on Pattern Anal. Mach. Intell.* 13 (1982) 333–340.
- [44] R. Hummel, S. Zucker, On the foundations of relaxation labeling processes, *IEEE Trans. Pattern Anal. Mach. Intell.* 5 (1983) 267–287.
- [45] A. Laine, G. Roman, A parallel algorithm for incremental stereo matching on SIMD machines, *IEEE Trans. Robot. and Automat.* 7 (1991) 123–134.
- [46] S. Lloyd, E. Haddow, J. Boyce, A parallel binocular stereo algorithm utilizing dynamic programming and relaxation labelling, *Comput. Vision Graphics Image Process.* 39 (1987) 202–225.
- [47] N.M. Nasrabadi, A stereo vision technique using curve-segments and relaxation matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (5) (1992) 566–572.
- [48] S. Peleg, A new probabilistic relaxation scheme, *IEEE Trans. Pattern Anal. Mach. Intell.* 2 (1980) 362–369.
- [49] K. Prazdny, Detection of binocular disparities, *Biol. Cybernet.* 52 (1985) 93–99.
- [50] K.E. Price, Relaxation matching techniques – a comparison, *IEEE Trans. Pattern Anal. Mach. Intell.* 7 (5) (1985) 617–623.
- [51] S. Ranade, A. Rosenfeld, Point pattern matching by relaxation, *Pattern Recognition* 12 (1980) 269–275.
- [52] D. Sherman, S. Peleg, Stereo by incremental matching of contours, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (1990) 1102–1106.
- [53] C. Stewart, C. Dyer, Simulation of a connectionist stereo algorithm on a memory-shared multiprocessor, in: V. Kumar (Ed.), *Parallel Algorithms for Machine Intelligence and Vision*, Springer, Berlin, 1990, pp. 341–359.
- [54] C.Y. Wang, H. Sun, S. Yada, A. Rosenfeld, Some experiments in relaxation image matching using corner features, *Pattern Recognition* 16 (2) (1983) 167–182.
- [55] N.M. Nasrabadi, C.Y. Choo, Hopfield network for stereovision correspondence, *IEEE Trans. Neural Networks* 3 (1992) 123–135.
- [56] Y.-H. Tseng, J.-J. Tzen, K.-P. Tang, S.-H. Lin, Image-to-image registration by matching area features using Fourier descriptors and neural networks, *Photogrammetric Eng. Remote Sensing* 63 (8) (1997) 975–983.
- [57] U.R. Dhond, J.K. Aggarwal, Stereo matching in the presence of narrow occluding objects using dynamic disparity search, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (7) (1995) 719–724.
- [58] T. Pavlidis, Why progress in machine vision is so slow, *Pattern Recognition Lett.* 13 (1992) 221–225.
- [59] A. Huertas, G. Medioni, Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks, *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (5) (1986) 651–664.
- [60] D. Marr, E. Hildreth, Theory of edge detection, *Proc. Roy. Soc. London B* 207 (1980) 187–217.
- [61] J.G. Leu, H.L. Yau, Detecting the dislocations in metal crystals from microscopic images, *Pattern Recognition* 24 (1991) 41–56.
- [62] M.S. Lew, T.S. Huang, K. Wong, Learning and feature selection in stereo matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (9) (1994) 869–881.
- [63] E.P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*, Springer, New York, 1989.
- [64] S. Tanaka, A.C. Kak, A rule-based approach to binocular stereopsis, in: R.C. Jain, A.K. Jain (Eds.), *Analysis and Interpretation of range images*, Springer, Berlin, 1990, pp. 33–139.
- [65] R. Nevatia, K.R. Babu, Linear feature extraction and description, *Comput. Vision Graphics Image Process.* 13 (1980) 257–269.
- [66] R.O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
- [67] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.
- [68] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice-Hall Englewood Cliffs, NJ, 1992.
- [69] A.M.N. Fu, H. Yan, A new probabilistic relaxation method based on probability space partition, *Pattern Recognition* 30 (11) (1997) 1905–1917.
- [70] S.Z. Li, Matching: invariant to translations, rotations and scale changes, *Pattern Recognition* 25 (1992) 583–594.
- [71] J. Majumdar, Seethalakshmy, Efficient parallel processing for depth calculation using stereo, *Robot. Autonomous Systems* 20 (1997) 1–13.

About the Author—G. PAJARES received the B.S. and Ph. D degrees in physics from U.N.E.D. (Distance University of Spain) (1987, 1995) discussing a thesis on the application of pattern recognition techniques to stereovision. Since 1990 he has worked at ENOSA in critical software development. He joined the Complutense University in 1995 as an Associate Professor in Robotics. His current research interests include robotics vision systems and applications of automatic control to robotics.

About the Author—J.M. CRUZ received a M.Sc. degree in Physics and a Ph.D. from the Complutense University in 1979 and 1984 respectively. From 1985 to 1990 he was with the department of Automatic Control, UNED (Distance University of Spain), and from October 1990 to 1992 with the Department of Electronic, University of Santander. In October 1992, he joined the Department of Computer Science and Automatic Control of the Complutense University where he is professor. His current research interests include robotics vision systems, fusion sensors and applications of automatic control to robotics and flight control.

About the Author—J.A. LOPEZ-OROZCO received an MSc degree in Physics from the Complutense University in 1991. From 1992 to 1994 he had a Fellowship to work on Inertial Navigation Systems. Since 1995 he has been an assistant researcher in robotics at the Department of Computers Architecture and Automatic control, where he is working for his Ph.D. His current interest are robotics vision systems and multisensor data fusion.