

On Combining Support Vector Machines and Simulated Annealing in Stereovision Matching

Gonzalo Pajares and Jesús M. de la Cruz

Abstract—This paper outlines a method for solving the stereovision matching problem using edge segments as the primitives. In stereovision matching, the following constraints are commonly used: *epipolar*, *similarity*, *smoothness*, *ordering*, and *uniqueness*. We propose a new strategy in which such constraints are sequentially combined. The goal is to achieve high performance in terms of correct matches by combining several strategies. The contributions of this paper are reflected in the development of a similarity measure through a support vector machines classification approach; the transformation of the smoothness, ordering and epipolar constraints into the form of an energy function, through an optimization simulated annealing approach, whose minimum value corresponds to a good matching solution and by introducing specific conditions to overcome the violation of the smoothness and ordering constraints. The performance of the proposed method is illustrated by comparative analysis against some recent global matching methods.

Index Terms—Epipolar, matching, ordering, similarity, simulated annealing, smoothness, stereovision, support vector machines, uniqueness.

I. INTRODUCTION

A MAJOR PORTION of the research efforts of the computer vision community has been directed toward the study of the three-dimensional (3-D) structure of objects using machine analysis of images [1]. According to [2], 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, which is 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 focal length, the orientation of the optical axis, and the sampling interval of each camera are known, the depth can be established by triangulation.

Manuscript received August 28, 2003; revised February 29, 2004. This work was supported in part under projects CICYT DPI2002-02924 and CICYT TAP94-0832-C02-01. This paper was recommended by Associate Editor X. Jiang.

The authors are with the Departamento Arquitectura de Computadores y Automática, Facultades de Informática y Físicas, Universidad Complutense, 28040 Madrid, Spain (e-mail: pajares@dacya.ucm.es).

Digital Object Identifier 10.1109/TSMCB.2004.827391

A. 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]. Area-based stereo techniques use correlation between brightness (intensity) patterns in the local neighborhood of a pixel in one image with brightness patterns in the local neighborhood of the other image [5]–[7], 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 [8]–[10], or the corresponding edges themselves [4], [11]–[17]. 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 a small number of features to be considered. We select a feature-based method with edge-segments as features, as they are abundant in the environment where our mobile robot equipped with the stereovision system navigates. They have been studied in terms of reliability [3] and robustness [18].

B. 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 of the following constraints [6], [8], [10]:

- 1) *epipolar*, given two segments one in the left image and a second in the right one, if we slide one of them along a parallel direction to the epipolar line, they would intersect (overlap) (Fig. 1);
- 2) *similarity*, matched edge segments have similar local properties or attributes;
- 3) *smoothness*, disparity values in a given neighborhood change smoothly, except at a few depth discontinuities;
- 4) *ordering*, the relative position among two edge-segments in the left image is preserved in the right one for the corresponding matches;
- 5) *uniqueness*, each edge-segment in one image should be matched to a unique edge-segment in the other image.

The similarity and uniqueness constraints are associated to a *local* matching process, the smoothness and ordering constraints to a *global* matching process, and the epipolar is with both, local and global processes. The major difficulty of stereo processing arises due to the need to make *global correspondences*.

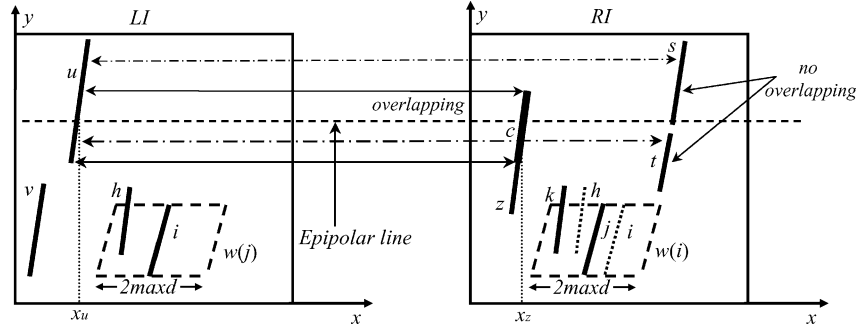


Fig. 1. Overlapping concept, edge-segments interactions and neighborhood for the pair (i, j) .

C. Relaxation in Stereovision

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 state in order to achieve a globally consistent interpretation of image data [19]. Two main relaxation processes can be distinguished in stereo image matching, i.e., 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 constraints. The following papers use a relaxation technique: 1) probabilistic/merit [8], [14], [20]–[23] and 2) optimization through a Hopfield neural network [9], [10], [13], [24]. We also use an optimization approach based on energy minimization through *simulated annealing* (SA) where a network of nodes (each one with its state) is built and the states of the nodes are updated. Such states are previously obtained through a classification approach based on the framework of *support vector* (SV) machines. The final states determine the correspondences.

D. Motivational Research and Contribution

For a given pair of features the application of the epipolar constraint during the edge extraction and the similarity constraint makes local correspondences, but the smoothness, ordering, and epipolar (a second application) constraints make global correspondences based on neighboring interactions. In the local process, we propose the mapping of the similarity constraint through the SV classification approach and in the global process the mapping of the smoothness, ordering, and epipolar constraints through the SA optimization approach for solving the edge-segment stereovision matching problem. This is based on the following reasons:

- 1) we incorporate the recent advances and the power of the SV machines for classification, where the similarity constraint can be easily mapped;
- 2) we use the well-known optimization SA strategy, allowing the direct mapping of the smoothness, ordering and epipolar constraints;
- 3) we have used successfully classification approaches for local correspondences based on the similarity constraint [15] among others and also global matching relaxation (optimization and probabilistic) approaches mapping the smoothness and ordering constraints [13], [14];
- 4) in the literature relaxation matching approaches have been also successfully used (see previous section).

The main contribution of the paper is the combination of both SV and SA strategies and the exploitation of their properties for solving the stereovision matching problem in order to achieve high performance in terms of correct matches. The performance for real time applications is left open for parallel implementations. We have modified appropriately the SV and SA frameworks in order to fit them to our stereovision matching problem, and we have found a better performance of the proposed approach with respect the above mentioned strategies. Each pair of edge-segments to be matched is a node in a network and each node has its own state, which determines the strength of the correspondence. Through the SV, the nodes are loaded with an initial matching estimation, which is updated through the SA optimization process. The correspondences are established based on the final nodes' states. We consider the possible smoothness and ordering constraints violations due to occlusions, camera geometry, objects near the cameras, etc. Such considerations make up another contribution to the stereovision matching problem. We have designed our sequential approach (SV followed by SA) based on the assumption that in relaxation methods (labeling/optimization) the final solution depends heavily on the starting point and tend to relax in local optima [25]. So, a good performance in SV will contribute positively to the overall performance. Therefore, our method integrates several strategies. The performance of the SV and SA local and global matching strategies, respectively, and the coupling for the final matching process, as well as the results obtained against other existing strategies justify the choice.

E. Paper Organization

The paper is organized as follows. In Section II the sequential processing is described. The performance of the method is il-

illustrated in Section III, where a comparative study against other existing global optimization/probabilistic matching methods is carried out. Finally in Section IV there is a discussion of some related topics.

II. SEQUENTIAL PROCESSING

The two cameras have equal focal lengths and are aligned so that their viewing direction is parallel. The plane formed by the viewed 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. Thus each point in an edge segment lies along the same horizontal (epipolar) line crossing both images. This is part of the system calibration process.

This paper proposes a sequential combination of four methods for solving the stereovision matching problem. Each method is implemented by a module. The system receives as inputs a pair of stereo images left (*LI*) and right (*RI*). This pair is processed in order to extract *features and attributes* in the *FA* module, each pair of extracted features (i, j) is to be matched, the features i and j come from *LI* and *RI*, respectively. For each pair (i, j) an attribute difference vector \mathbf{x} is computed. All extracted \mathbf{x} vectors are supplied to the *support vector* module (*SV*), which computes a matching state s_{ij} for each pair of features. This module uses the similarity constraint and requires a previous training process in order to achieve each current matching state. Until this stage, only a local matching process has been carried out. Once all matching states are obtained, they are supplied to the *simulated annealing* (*SA*) module, which updates the states s_{ij} by the optimization process. The *SA* module implements the global matching process. After this stage, perhaps there are still ambiguous matches which are solved by the *unambiguous* (*UA*) module based on the strength of each state. The output of the system is a set of matches. We give details about the behavior of each module.

A. Feature and Attribute Extraction (FA)

The contour edges in both images are extracted using the Laplacian of Gaussian filter in accordance with the zero-crossing criterion [26]. For each zero-crossing in a given image, its gradient vector (magnitude and direction) [27], Laplacian [28], and variance [29] values are computed from the gray levels of a central pixel and its eight immediate neighbors. The edges are obtained by joining adjacent zero-crossings following the algorithm of [30], where 1) a margin of deviation of $\pm 20\%$ in gradient magnitude and of $\pm 45^\circ$ in gradient direction is allowed, and 2) each detected contour is approximated by a series of piecewise linear line segments [31]. 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 pair of features has two associated four-dimensional (4-D) vectors \mathbf{x}_i and \mathbf{x}_j , where the components are the attribute values, and the sub-indices i and j denote features belonging to the left and right images, respectively. A 4-D difference measurement vector \mathbf{x} is then also obtained from the above \mathbf{x}_i and

\mathbf{x}_j vectors, $\mathbf{x} = \mathbf{x}_i - \mathbf{x}_j = \{x_m, x_d, x_l, x_v\}$. The components of \mathbf{x} are the corresponding differences for module and direction gradient, Laplacian and variance values. Only those pairs verifying the following three initial conditions will be processed:

- 1) their absolute value of the difference in the gradient direction is below a specific threshold, fixed to 25° ;
 - 2) their absolute value in the gradient magnitude is also below a fixed threshold, set to 15;
 - 3) their overlap rate surpasses a certain value, fixed to 0.5.
- The remaining pairs that do not meet such conditions are directly considered as false correspondences.

The overlap is a concept introduced in [4], two segments u and z , overlap if by sliding one of them in a direction parallel to the epipolar line, they would intersect. In our approach, the overlapping is a first implementation of the epipolar constraint in the local matching process. Then, the epipolar constraint will be also mapped as an energy function to be minimized in the global SA approach.

Fig. 1 clarifies the overlapping concept. Indeed, segment u in the left image overlaps with segment s in the right image, but segment v does not overlap with s . The overlap rate between edge segments (u, z), α_{uz} is defined as the percentage of coincidence, ranging in $[0, 1]$, when two segments u and z overlap, and it is computed taken into account the common overlap length l_c defined by c and the two lengths for the involved edge segments l_u and l_z , respectively. All lengths are measured in pixels

$$\alpha_{uz} = \frac{2l_c}{(l_u + l_z)}. \quad (1)$$

Taking into account the above initial conditions 1) and 3) and the parallel optical axis geometry of the stereovision system, we compute the disparity between two edge-segments (u and z in Fig. 1) as follows: trace epipolar lines (four) crossing the common overlapping segment (c), for each line compute x_u and x_z , so the disparity is $(x_u - x_z)$. Then the disparity for edge segments u and z is the averaged disparity for the four pairs of points x_u and x_z .

B. Support Vector (SV) Machines: Local Matching Strategy

The stereovision matching problem is viewed as a two classification problem, where a pair of edge-segments is classified as a true or false match (true and false classes). This method requires a previous training process, where the goal is to obtain a decision function from the training patterns. This function will be used during the current matching process for computing the states s_{ij} from the pairs of features supplied by the *FA* module.

1) *Training Process*: The SV is based on the observation of a set X of n pattern samples. The outputs are two symbolic values $y \in \{+1, -1\}$, $y = +1$ corresponds the class of true matches [32]–[34]. The finite sample (training) set is denoted by (\mathbf{x}_l, y_l) , $l = 1, \dots, n$, where each \mathbf{x}_l vector is a training element and $y_l \in \{+1, -1\}$ the class it belongs to. In our stereovision matching problem \mathbf{x}_l is a 4-D difference vector, computed as specified in the above Section II-A. The goal of *SV* is to find,

from the training sample set, a decision function capable of separating the data into two groups. Such SV decision function has the general form

$$f(\mathbf{x}) = \sum_{l=1}^n \alpha_l y_l H(\mathbf{x}_l, \mathbf{x}). \quad (2)$$

Equation (2) establishes a representation of the decision function $f(\mathbf{x})$ as a linear combination of kernels centered in each data point. In this paper, we have used Gaussian Radial Basis functions $H(\mathbf{x}, \mathbf{y}) = \exp\{-|\mathbf{x} - \mathbf{y}|^2/\sigma^2\}$ where σ defines the width of the kernel set to 3.0 after different experiments. The parameters α_l , $l = 1, \dots, n$, in (2) are the solution for the following quadratic optimization problem. Maximize the functional

$$Q(\alpha) = \sum_{l=1}^n \alpha_l - \frac{1}{2} \sum_{l,m=1}^n \alpha_l \alpha_m y_l y_m H(\mathbf{x}_l, \mathbf{x}_m)$$

$$\text{subject to } \sum_{l=1}^n y_l \alpha_l = 0, \quad 0 \leq \alpha_l \leq \frac{c}{n}, l = 1, \dots, n \quad (3)$$

and given the training data (\mathbf{x}_l, y_l) , $l = 1, \dots, n$, the inner product kernel H , and the regularization parameter c . As stated in [33], at present, there is not a well-developed theory on how to select the best c , although in several applications it is set to a large fixed constant value, set to 2000 in this paper. The data points \mathbf{x}_l associated with the nonzero α_l are called *support vectors*.

2) *Local Matching Process*: Once the system has been trained, we have available the decision function $f(\mathbf{x})$ generated by the SV module. In our matching classification problem the $f(\mathbf{x})$ polarity, sign of $f(\mathbf{x})$, determines the class membership for a node (i, j) with a difference vector \mathbf{x} . We interpret the magnitude of $f(\mathbf{x})$ as a measure of certainty in the decision made. But for each node we need its state s_{ij} , which is computed by a warping function (4) that modifies the sigmoid function in [9], ranging in $[-1, +1]$. In order to avoid severe bias in the distances for the training data, the parameter a is estimated experimentally and set to 0.2 in our experiments

$$s_{ij}(\mathbf{x}) = \frac{2}{1 + \exp(-af(\mathbf{x}))} - 1. \quad (4)$$

C. Simulated Annealing (SA): Global Matching Strategy

The SA receives the network with each node initialized according to (4). The goal of the optimization process is to increase the consistency of a given pair of edge segments among three constraints (smoothness, ordering, and epipolar) so that the state of a node representing a correct match can be increased and the state of any incorrect match can be decreased during the optimization process. Suppose a network has N nodes. The simulated annealing optimization problem is to modify the values of the s_{ij} so as to minimize the energy

$$E = -\frac{1}{2} \sum_{i,j=1}^N \sum_{h,k=1}^N w_{(ij)(hk)} s_{ij} s_{hk} \quad (5)$$

where $w_{(ij)(hk)}$ is a symmetric weight interconnecting two nodes (i, j) and (h, k) . We require the self-feedback terms to vanish (i.e., $w_{(ij)(ij)} = 0$) because the nonzero merely add an unimportant constant to E , independent of the s_{ij} . The optimization task is to find the states with the most stable configuration, the one with lowest energy. The energy function is built so that it represents three stereovision constraints, i.e., *smoothness*, *ordering*, and *epipolar*. 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. The compatibility coefficient makes global consistency between neighbors pairs of edge segments based on such constraints.

D. Mapping the Smoothness Constraint

The smoothness constraint assumes that neighboring edge segments have similar disparities, except at a few depth discontinuities [4]. Generally, when the smoothness constraint is applied, it is assumed there is a bound on the disparity range allowed for any given segment. We denote this limit as *maxd*, fixed at 15 pixels in this paper (see Fig. 1). According to the procedure described in [4] for example, 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 segment “ j ” in the right image, we define a window $w(j)$ in the left image in which corresponding edge segments from the left image must lie. It is said that “a segment h must lie in $w(j)$ ” if at least the 30% of the length of the segment “ h ” is contained in the $w(j)$ window. The shape of this window is a parallelogram; one side is “ i ,” for left to right match, and the other a horizontal vector of length $2 \cdot \text{maxd}$. 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 $|d_{ij} - d_{hk}|$, 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 [4], [10], [13], [14], and [24] among others. We define a compatibility coefficient derived from [10] and [24] given by

$$c_{(ij)(hk)}(D) = \frac{2}{1 + \exp\left[\gamma \left(\frac{D}{m(D)} - 1\right)\right]} - 1 \quad (6)$$

where $D = |d_{ij} - d_{hk}|$, $m(D)$ denotes the average of all values D in the pair of stereo images (*LI* and *LR*) under processing. The slope of the compatibility coefficient in (6) is expressed by γ and varies for each pair of stereo images. To determine γ , it is assumed that the probability distribution function of D is Gaussian with average $m(D)$ and standard deviation $\sigma(D)$, i.e., $p(D) = [1 + \exp[\gamma(D_{(ij)(hk)}/m(D) - 1)]]^{-1}$. Under this assumption and following [35], [36], to set the possibility value to 0.1 when the value of cumulative distribution function is 0.9, γ value is calculated by $\gamma = \ln 9((m(D))/((1.282\sigma(D))))$. In our

experiments, typical values of γ , $m(D)$, and $\sigma(D)$ are about 6, 9, and 2, respectively. So, values of D near 0 should give high values in the compatibility coefficient $c_{(ij)(hk)}(\cdot) \approx +1$, but near 25 they give low values, $c_{(ij)(hk)}(\cdot) \approx -1$ and intermediate values should give values near 0, as expected. Note that $c_{(ij)(hk)}(\cdot)$ ranges in $[-1, 1]$. So, a compatibility coefficient of +1 is obtained for a good consistency between nodes (i, j) and (h, k) ($D = 0$) and a compatibility of -1 for a bad consistency ($D \gg 0$). The energy function embedding the smoothness constraint must be minimum when $D = 0$ (i.e., high compatibility coefficient) and high states values. We define an energy function assuming this as follows:

$$E_s = -\frac{A}{2} \sum_{ij=1}^N \sum_{hk=1}^N c_{(ij)(hk)} s_{ij} s_{hk} \quad (7)$$

where A is a positive constant to be defined later.

E. Mapping the Ordering Constraint

We define the ordering coefficient $\bar{O}_{(ij)(hk)}$ for the edge-segments according to (8), which measures the relative average position of edge segments “ i ” and “ h ” in the left image with respect to “ j ” and “ k ” in the right image, it ranges from 0 to 1:

$$\begin{aligned} \bar{O}_{(ij)(hk)} &= \frac{1}{N} \sum_N o_{(ij)(hk)} \\ \text{where } o_{(ij)(hk)} &= |S(x_i - x_h) - S(x_j - x_k)| \\ \text{and } S(r) &= \begin{cases} 1, & \text{if } r > 0 \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (8)$$

We trace S scanlines (four) along the common overlapping length, each scanline produces a set of four intersection points (i_S and h_S in LI and j_S and k_S in the RI) with the four edge-segments. Hence, the lower-case o_{ijhk} can be computed as in [10] considering the above four edge points, and it takes 0 and 1 as two discrete values. As $c_{(ij)(hk)}(\cdot)$ ranges in $[-1, +1]$, in order to achieve similar contributions, we re-scale the $\bar{O}_{(ij)(hk)}$ values to $[-1, +1]$ as follows: $O_{(ij)(hk)} = 2\bar{O}_{(ij)(hk)} - 1$.

To satisfy the ordering constraint, the energy function should have its minimum value when the nodes constituting each pair of nodes, for which the corresponding edges do not satisfy the ordering constraint, have low states values simultaneously. The energy could be written as

$$E_o = \frac{B}{2} \sum_{ij=1}^N \sum_{hk=1}^N O_{(ij)(hk)} s_{ij} s_{hk} \quad (9)$$

where B is a positive constant to be defined later.

F. Mapping the Epipolar Constraint (Overlapping Concept)

The epipolar constraint is mapped through the overlapping concept in [4], by the overlapping coefficient

$$\bar{\lambda}_{(ij)(hk)} = 0.5(\alpha_{ij} + \alpha_{hk}) \quad (10)$$

where α is the overlap rate defined in (1). Under the epipolar constraint we can assume that correct matches should have high overlap rates and $\bar{\lambda}_{(ij)(hk)}$ for neighborhoods should be high increasing the consistency. The overlapping criterion is justified

by the fact that the edge segments are reconstructed by piecewise linear line segments as described in Section II-A. As before, we re-scale the $\bar{\lambda}_{(ij)(hk)}$ values to the interval $[-1, +1]$ as follows: $\lambda_{(ij)(hk)} = 2\bar{\lambda}_{(ij)(hk)} - 1$. The energy function should have its minimum value when the nodes constituting each pair of nodes, for which the corresponding edges satisfy the overlapping concept, have high $\lambda_{(ij)(hk)} (\approx 1)$ and high states values simultaneously. The energy could be written as

$$E_e = -\frac{C}{2} \sum_{ij=1}^N \sum_{hk=1}^N \lambda_{(ij)(hk)} s_{ij} s_{hk}. \quad (11)$$

G. Considering Smoothness and Ordering Constraints Violations

There are complex images in which the ordering and smoothness constraints can be violated. In systems with parallel geometry, objects close to the cameras, occlusions and also the definition of a neighborhood (disparity limit) could lead to such violations. So, some excellent neighborhoods could be excluded.

When can we say whether or not there are violations? The ordering constraint is not violated if $O_{(ij)(hk)} > U_0$, with U_0 set to 0.85 in this paper. The smoothness constraint is not violated when: 1) the pairs (i, j) and (h, k) are to be matched, being h a neighborhood of i in $w(j)$ and k a neighborhood of j in $w(i)$, Fig. 1, according to the fixed disparity limit (minimum disparity criterion [4]); 2) s_{hk} has a high positive value although it is not the maximum of all matches (h, k') where h is involved. This is formally expressed as follows: $s_{hk} \geq H^*(s_{hk'})_{\max} \forall k' \in RI$ with $(s_{hk'})_{\max} = \max\{s_{hk'} \mid \forall k' \in RI\}$, H is set to 0.85 in this paper. The compatibility coefficient is maximum if the matching states are maximum (minimum differential disparity) and vice versa. When no violations occur, the compatibility and ordering coefficients are computed through (6) and (8), respectively.

Depending on which of the above conditions is not fulfilled, we say that the ordering, smoothness or both are violated. When this occurs, global consistency is not applicable, and we must fall back on local consistency. This is to avoid that true pairs of edge segments are not penalized during the relaxation process due to the violation of such constraints. Depending on which constraint is violated, the corresponding coefficient is

$$v_{(ij)(hk)} = 0.5(s_{ij} + s_{hk}) \quad \text{where } v = c, O. \quad (12)$$

H. Total Energy Function

The total energy function can be obtained as $E = E_s + E_o + E_e$. By comparison of expressions (7), (9), and (11) and (5), by multiplying the constant term by -1 , it is easy to derive the connection weights

$$w_{(ij)(hk)} = (Ac_{(ij)(hk)} - BO_{(ij)(hk)} + C\lambda_{(ij)(hk)} - \delta_{(ij)(hk)}) \quad (13)$$

where the delta function $\delta_{(ij)(hk)} = 1$ for $(i, j) = (h, k)$ and 0 otherwise. To ensure the convergence to stable state, symmetrical inter-connection weights and no self-feedback are required, i.e., we see that by setting $A = B = C = 1$ both conditions are fulfilled.

I. Deterministic Simulated Annealing

Originally developed in [37] and [38], in this paper we have implemented the approach described in [39], [40]. According to [39], we have chosen deterministic simulated annealing because the stochastic one is slow. We have verified this assertion by implementing both versions (deterministic and stochastic) obtaining very similar solutions and identical performance in terms of percentage of correct matches. Nevertheless, the deterministic version has been faster than the stochastic, by two orders of magnitude. This agrees with [39]. Moreover, we have not found problems to reach the global minima under the deterministic version; this is because the SA is initially guided (not randomly) by the SV procedure.

In the original SA algorithm, the forces exerted by the other nodes are summed to find an analogue value s_{ij} without the intervention of the state of the node which is being updated. We modify this in order to include the contribution of its own state, so that the power of the similarity constraint is considered. The temperature (T) also plays a very important role in the optimization process.

Let $F_{(ij)} = \sum_{(hk)} w_{(ij)(hk)} s_{hk}$ be the force exerted on node (i, j) by the other nodes (h, k) , then the new state $s_{ij}(t)$ is obtained by adding the fraction $f(\cdot, \cdot)$ to the previous one

$$\begin{aligned} s_{ij}(t) &= f(F_{(ij)}(t), T(t)) + s_{ij}(t-1) \\ &= \tanh\left(\frac{F_{(ij)}(t)}{T(t)}\right) + s_{ij}(t-1) \end{aligned} \quad (14)$$

where t represents the iteration index. The fraction $f(\cdot, \cdot)$ depends upon the temperature. At high T , the value of $f(\cdot, \cdot)$ is lower for a given value of the forces F . Details about the behavior of T are given in [39]. We have verified that this fraction must be small as compared to $s_{ij}(t-1)$ in order to avoid that the updating is controlled by this fraction exclusively and that the similarity constraint is cancelled. Under the above considerations and based on [17], [41], the following annealing schedule suffices to obtain a global minimum: $T(t) = T_0/\log(t+1)$, with T_0 being a sufficiently high initial temperature. We have computed T_0 as follows [42]:

- 1) we select four stereo images, previously the SV has been trained and the support vectors obtained; now we compute the initial energy;
- 2) we choose an initial temperature, that permits about 80% of all transitions to be accepted (i.e., transitions that decrease the energy function), and this value is changed until this percentage is achieved (in [42], a two-fold change is proposed);
- 3) we compute the M transitions ΔE_i and we look for a value for T for which $1/M \sum_{i=1}^M \exp(-\Delta E_i/T) = 0.8$, after rejecting the higher order terms of the Taylor expansion of the exponential, $T = 5\langle \Delta E_i \rangle$, where $\langle \cdot \rangle$ is the mean value.

In our experiments, we have obtained $\langle \Delta E_i \rangle = 6.10$, giving $T_0 = 30.5$ (with a similar order of magnitude as the reported in [17], [41]). We have also verified that a value of $t_{\max} = 100$ suffices, although the expected condition $T(t) = 0$, $t \rightarrow +\infty$ in the original algorithm is not fully fulfilled. This last requirement and a possible overly rapid cooling, only occur when

simulated annealing is applied for achieving the solid thermal equilibrium but not in our approach in which there is not a solid. Moreover, the above cooling scheduling is justified by the fact that our initial state has reached a certain equilibrium as a result of the SV local matching process and it is unnecessary to heat at high temperature, hence we have a prior knowledge about the system before it is relaxed by SA.

The proposed deterministic SA algorithm derived from [39] including the modifications mentioned is

1. *Initialization*: $t = 0$, $T(0) = T_0$, $w_{(ij)(hk)}$ as given by (14), $s_{ij} = 1, \dots, N$ the state values received from the SV module
2. *SA process*: set $t = t + 1$ and $np = 0$; for each node (i, j) update $s_{ij}(t)$ according to (14) and if $|s_{ij}(t) - s_{ij}(t-1)| > \varepsilon$ then $np = np + 1$; when all (i, j) nodes are updated, if $np \neq 0$ or $t < t_{\max}$ then go to step 2, else stop.
3. *Output*: s_{ij} updated

where, np is the number of nodes for which the matching states are modified by the updating procedure, N is the number of nodes, $T(t)$ is the annealing schedule, ε is a constant to accelerate the convergence, set to 0.01.

J. Unambiguous Module (UA): Uniqueness Constraint

This stage represents the mapping of the *uniqueness* constraint, which completes the set of matching constraints used for solving our stereovision matching problem.

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 is made by choosing the greater state value (in the unambiguous case there is only one) whenever it surpasses a previous fixed threshold $U_1 (= 0)$, intermediate value for s_{ij} ranging in $[-1, +1]$. A true match should have $s_{ij} = +1$.

The ambiguities produced by broken edge segments are allowed. Therefore, we make a provision for broken segments resulting in possible multiple correct matches. The following pedagogical example from Fig. 1 clarifies this. The edge segment u in LI matches with the broken segment represented by s and t in RI , but under the condition that s and t do not overlap, that the s and t orientations do not differ by more than $U_2 (\pm 10^\circ)$ and both s_{us} , s_{ut} are greater than U_1 .

III. VALIDATION, COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION

A. Design of a Test Strategy

In order to assess the validity and performance of the proposed method, we have selected 82 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 82 stereo pairs are classified into three groups: SP1, SP2, and SP3 with 28, 31, and 23 pairs of stereo images, respectively. Previously, we used an

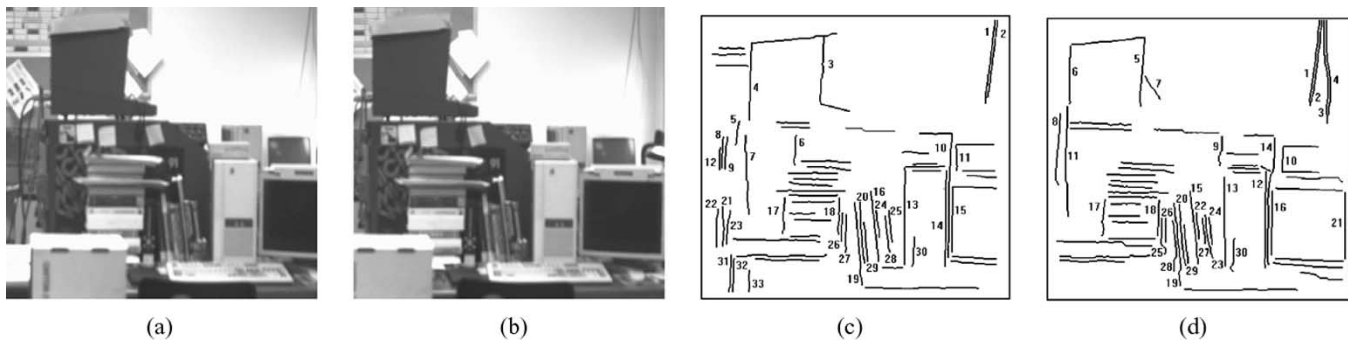


Fig. 2. (a) Group SP2 and (b) original left and right stereo images; (c) and (d) labeled edge segments in left and right images.

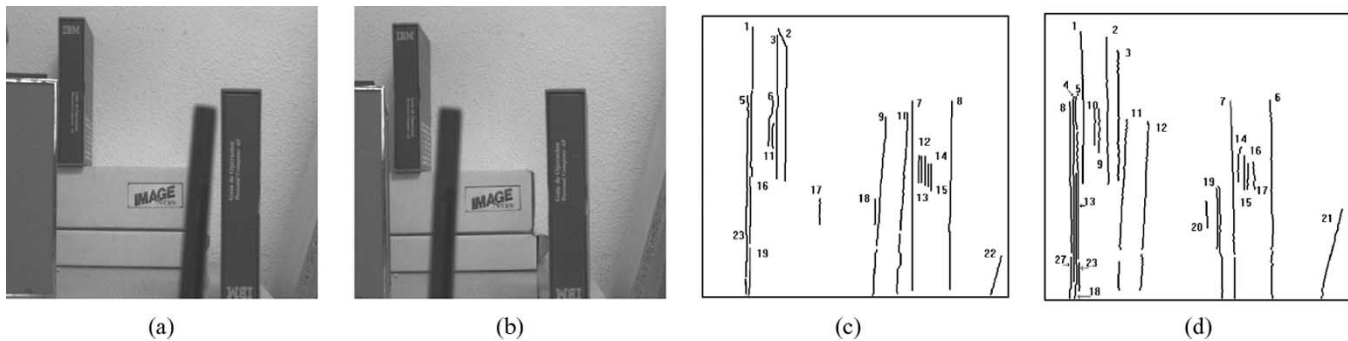


Fig. 3. (a) Group SP3 and (b) original left and right stereo images; (c) and (d) labeled edge segments in left and right images.

additional set of 15 stereo images to obtain the support vectors which are to be used in the decision function, (2), required by the SV. Then, during the current stereo matching process and after the processing of each stereo pair of images, the set of support vectors is updated according to the procedure described in Section II-B1. Group SP1 consists of stereo images without apparent complexity.

Group SP2 corresponds to scenes where a repetitive structure (vertical books) has been captured, Fig. 2 show a stereo pair representative of SP2. Finally, group SP3 contains objects close to the cameras, that produce a high range of disparity violating the smoothness and ordering constraints. Fig. 3 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. The SP2 and SP3 are of special interest as they are complex images containing structures that appear with a high degree of difficulty. This kind of images have been studied in depth in [13], [14] where similar considerations have been also introduced to deal with the violation of the smoothness and ordering constraints. Therefore, we compare the support vector and simulated annealing (SVSA) we propose in this paper with the method described in [14] which is a relaxation labeling (RELB) approach and the method described in [13], which is an optimization approach based on the Hopfield neural network (HNNB1).

Both, RELB and HNNB1 apply the similarity constraint by computing a matching probability based on the estimation of a probability density function through the Bayes's theory. The

matching probabilities are used as the inputs for the relaxation and optimization processes, respectively. From these processes, RELB performs an iteration procedure by applying smoothness, ordering and uniqueness constraints. HNNB1 performs the optimization process by mapping the smoothness and uniqueness in an energy function which is to be minimized. From HNNB1, we have implemented a new version HNNB2, by mapping the ordering constraint as an energy function to be minimized and applying the similarity constraint as the 4-D difference null vector \mathbf{x} . HNNB2 can be considered a very close approach to the described in [10], although this work uses edge pixels as features, we have modified the original method in [10] to use edge-segments as in SVSA. A variant of the HNNB1 is the mapping of the similarity constraint by estimating the probability density function through the Parzen's window as described in [15].

We have also compared our approach with the stochastic stereovision matching method (SSVM) in [43], also used in [44]. This method uses the regularization criterion proposed in [45], where an energy functional is minimized based on a penalty functional which measures the dissimilarity between corresponding features (similarity constraint) and a stabilizing functional by which the smoothness constraint is imposed (see also [46]). The energy minimization is carried out through the simulated annealing algorithm, we have used a value of 50 as in [43] for the regularization parameter λ (this works well for images quantized in eight-bit values) and the same neighborhood criteria as that used in this paper. Two differences are considered in this implementation with respect to our implementation. 1) The edge-segments disparities are the outputs obtained in SSVM, which are used to obtain the correspondences and 2) the hierarchical coarse-to-fine control structure with reheating in [44] is not used in our implementation.

TABLE I
SUMMARY OF STEREOVISION MATCHING METHODS AND CONSTRAINTS

	<i>Stereovision matching constraints</i>					<i>constraints violation</i>	
	<i>Similarity</i>	<i>Smoothness</i>	<i>Ordering</i>	<i>Epipolar</i>	<i>Uniqueness</i>	<i>Smoothness</i>	<i>Ordering</i>
SVSA	Support Vector Machines	mapped as an energy minimized by Simulated Annealing	mapped as an energy minimized by Simulated Annealing	mapped as an energy minimized by Simulated Annealing	applied by selecting the highest state values	Yes	Yes
RELB	Bayes probability density estimation	Probabilistic relaxation	Probabilistic relaxation	mapped under the overlapping concept	applied by selecting the highest probabilities	Yes	Yes
HNNB1	Bayes probability density estimation	mapped as an energy minimized by Hopfield	No	mapped under the overlapping concept	mapped as an energy minimized by Hopfield	Yes	No
HNNB2	Euclidean distance without estimation	mapped as an energy minimized by Hopfield	mapped as an energy minimized by Hopfield	mapped under the overlapping concept	mapped as an energy minimized by Hopfield	No	No
HNNP	Parzen's window probability density estimation	mapped as an energy minimized by Hopfield	No	mapped under the overlapping concept	mapped as an energy minimized by Hopfield	Yes	No
SSVM	mapped as an energy minimized by regularization	mapped as an energy minimized by regularization	No	implicit application by image registration	No	No	No
MDDA	qualitative boolean function	merit function relaxation	No	implicit application by image registration	applied by selecting the highest merits	No	No

TABLE II
TRAINING PATTERNS AND SUPPORT VECTORS

	Initially		SP1		SP2		SP3	
<i># Training patterns</i>	125	180	643	781	1227	1435	1701	1921
<i># Support Vectors</i>	18	31	32	49	51	73	66	85

Finally we have chosen the minimum differential disparity algorithm [4] (MDDA) for comparative purposes for the following reasons:

- 1) it is a merit relaxation approach;
- 2) it applies the commonly used constraints (similarity, smoothness and uniqueness);
- 3) it uses edge segments as features and the contrast and orientation of the features as attributes;
- 4) some concepts of MDDA, such as minimum differential disparity, overlapping concept, disparity limit or average disparity are used in our SVSA approach.

Table I summarizes the main differences between the six methods compared. All methods use edge-segments as features and the same four attributes.

We have used the same set of training patterns for estimating the decision function in (2) for SVSA and the probabilities densities in RELB, HNNB1, and HNNB2. Once each set SP1, SP2, and SP3 is processed, we use the matches as new training patterns, which are added to the old ones, for new estimations.

Table II displays the number of pairs of edge-segments used as training patterns and the number of support vectors obtained. Two columns are shown for each group: the left with the thresholds and settings of the Section II-A and the right when such settings are relaxed to allow a greater number of pairs (next section).

An important reduction is achieved in the number of support vectors with respect to the training patterns. This is an impor-

tant improvement because we only need to store and handle the support vectors in SVSA.

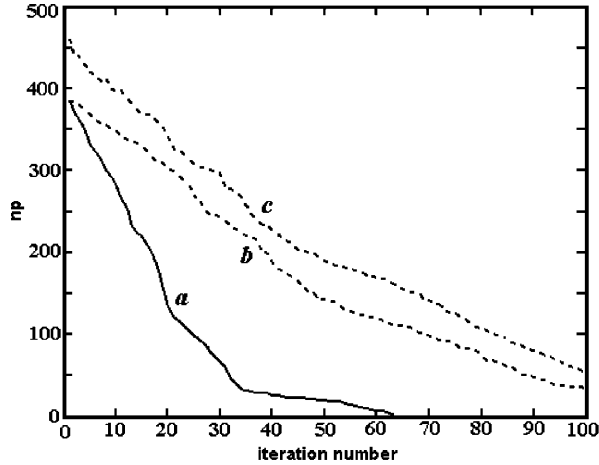
B. Comparative Analysis

The system processes the SP1, SP2, and SP3 groups. Of all the possible combinations of pairs of matches formed by segments of left and right images only 643, 584, and 474 pairs are considered for SP1, SP2, and SP3, respectively, under the thresholds and settings given in Section II-A. These are the incremental training patterns shown in the most left column in Table II for each group. The computed results with the thresholds and settings given in Section II-A are summarized in Table III. It shows the percentage of successes for each group (SP1, SP2, and SP3) and for each method (SVSA, RELB, HNNB1, HNNB2, HNNP, SSVM, and MDDA) as a function of the number of iterations. Iteration 0 corresponds with results for the local matching process (SV in this paper), the starting point for the optimization/relaxation process (SA in this paper).

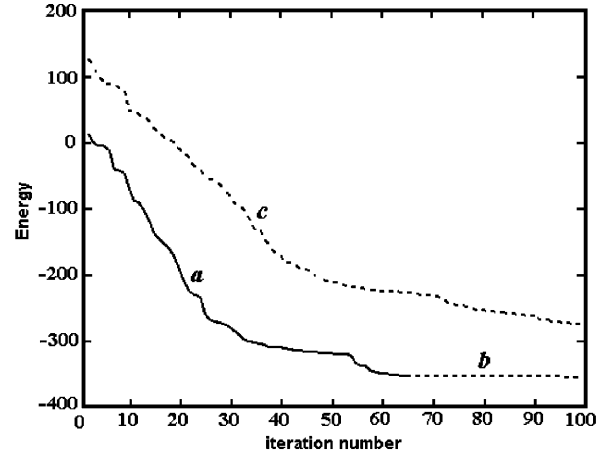
1) *Decision Process*: When the optimization/relaxation processes have finished, there are still both unambiguous pairs of segments and ambiguous ones, depending on either one, and only one, or several right image segments corresponding to a given left image segment. In any case, the decision about whether a match is correct or not is made by choosing the result of greater state in SVSA, greater probability in RELB, HNNB1, HNNB2, and HNNP, and the best merit value for MDDA. The results in SSVM are directly obtained from the

TABLE III
PERCENTAGE OF SUCCESSES FOR THE GROUPS OF STEREO-PAIRS SP1, SP2, AND SP3

Iteration #	0			30			65		
SP#	SP1	SP2	SP3	SP1	SP2	SP3	SP1	SP2	SP3
SVSA	83.9	67.6	78.3	88.4	82.5	86.3	97.5	94.3	95.8
RELB	77.1	65.9	62.6	86.1	82.6	87.9	95.3	93.1	94.3
HNNB1	77.1	65.9	62.6	85.9	80.7	86.8	93.8	92.3	94.0
HNNB2	73.1	58.5	57.8	84.8	70.1	66.1	92.5	75.9	70.3
HNNP	82.0	67.6	66.3	87.2	83.0	86.9	95.6	93.9	95.0
SSVM	0	0	0	68.2	55.2	37.2	80.1	79.1	58.1
MDDA	72.1	57.1	56.9	80.4	69.9	61.9	86.1	76.8	59.5



(a)



(b)

Fig. 4. (a) Number of pairs (np) of edge segments for which the matching states are modified by SA as a function of the iteration number. (b) Energy values as a function of the iteration number averaged for all SA processes.

disparity values. In the unambiguous case there is only one. The values must be greater than the corresponding intermediate value: 0.5 for the probabilities in RELB, HNNB1, HNNB2, and HNNP and 0 for the states in SVSA.

According to values from Table III, the following conclusions may be inferred.

2) *Local Matching Process: Iteration 0* (the optimization/relaxation has not started), the best performance is achieved with SVSA. This means that the SV approach appears as a good local matching method. The results obtained by HNNP are also good. The methods without estimation (HNNB2 and MDDA) obtain the worst results at this phase.

3) Global Matching Process:

Group SP1: The best performance is achieved with SVSA. The probabilistic/optimization process improves the local matching results. The ordering constraint is not decisive: HNNB2 (with ordering) obtains worse results than HNNB1 and HNNP (without ordering). There are no violations in the smoothness and ordering constraints. As shown in Fig. 4, SVSA reaches its equilibrium with an average of 65 iterations (this number is used for the remainder methods) under the thresholds and settings of this paper. SSVM only reaches the 96.0 in percentage with 400 iterations, as reported in [43]. MDDA obtains the worst results.

Group SP2 (Containing Repetitive Structures): SVSA achieves the best results. The ordering constraint, once again, is not relevant (HNNP achieves better results than RELB). The consideration of the smoothness constraint violation is decisive

(HNNB2 achieves poor results as compared to SVSA, RELB, HNNB1, or HNNB2) and the ordering constraint violation has low relevance (HNNP obtains better results than RELB). The behavior of SSVM is similar to that of group SP1. MDDA obtains the worst results.

Group SP3 (Structures Violating the Smoothness and Ordering Constraints): The best performance is achieved with SVSA. As expected, the methods that take into account the smoothness and the ordering violation (SVSA and RELB) achieve better results than HNNB1 and HNNP (only smoothness) and of course better than HNNB2 or SSVM (without any consideration). The results obtained with the MDDA decrease as the number of iterations increases, because the merit of false matches increases (pairs (3,3) or (2,2) in Fig. 3).

C. Experimental Results and Thresholds Settings

Table IV summarizes the set of thresholds and settings with a description, their values (used in this paper for obtaining the above results), relevance (high, medium or low) and comments.

We have carried out some additional experiments in order to determine the influence of the thresholds. Fig. 4(a) shows the number of pairs (np) of edge segments for which the matching states are modified during all SA processes as a function of the iteration number. Fig. 4(b) shows the average energy values for all SA processes as a function of the iteration number, and they are computed according to (5), (7), (9), and (11). Each figure shows three lines labeled as *a*, *b*, and *c*. The continuous line *a* corresponds to the settings used in this paper (Table IV). The

TABLE IV
THRESHOLDS AND SETTINGS USED IN THIS PAPER

Module	Description	Value(s)	Relevance	Comments
FA	Gradient edge pixel joining (magnitude, direction)	$\pm 20\%, \pm 45^\circ$	Low	Could range in $\pm 30\%$ and $\pm 90^\circ$
	Gradient edge segment difference for matching (magnitude, direction)	15, 25°	Medium	Could range between $[0,30]$ and $[0^\circ,45^\circ]$ respectively. If the values increase, the number of pairs to be matched increases.
	Overlap rate	0.5	Medium	Ranges in $[0,1]$. High values decreases the number of pairs and do not allow broken edge segments
SV	Width kernel, σ in the Gaussian Radial Basis	3.0	Low	More data less value, could range in $[0.5,10]$
	Regularization parameter, c in equation (3)	2000	Low	Large values ranging in $[1000, 5000]$, from reference [33]
	α in equation (5)	0.2	Low	Could range in $[0.05, 0.5]$
SA	Disparity limit, $maxd$ Determines the neighborhood	15	High	Depends on the baseline in the optics system geometry. In our system the baseline is 20cms. Could range in $[5,25]$
	Compatibility coefficient slope, γ in eq. (7)	6	High	Derived from references [35, 36]. Could range in $[2,15]$
	Constraint violations, U_0, H	0.85	Medium	High values are required for states of matched edge segments
	Initial schedule temperature, T_0	30.5	High	Computed from reference [42]. Its effect is minimized, as the initial states are loaded by the SV module. Could range $[20,50]$
	Maximum number of iterations, t_{max}	100	High	Must be sufficient to ensure the convergence, greater than 65
	Constant to accelerate the convergence, ε	0.01	High	Near the equilibrium the states change very slowly
UA	Disambiguation, U_1	0	Low	This is the intermediate state value. It ranges in $[-1,1]$
	Difference in segments orientations, U_2	$\pm 10^\circ$	Low	The system geometry limits such difference. It could reach $\pm 20^\circ$

dotted line b represents the results obtained by setting ε to 0 in SA, i.e., the convergence is not accelerated, leaving the remainder of thresholds and parameters as above (unchanged). The dotted line c represents the results obtained by changing the thresholds in the module FA and leaving the remainder unchanged, also ε which is set to 0.01 as in the Table IV. The modification in this last case has been oriented to increase the number of edge segments to be matched, namely: gradient edge pixel joining magnitude and direction to $\pm 30\%$ and $\pm 90^\circ$; gradient edge segment difference for matching in magnitude and direction to 30° and 45° and the overlap rate to 0.3.

In all experiments the critical parameters γ and T_0 have been set to their values in Table IV, they are computed as described in Sections II-D and II-J, respectively according to [35], [36] and [42]. We have varied their values inside of their ranges in Table IV, leaving the remainder values unchanged. The results are slightly worse than those obtained with their values in Table IV. This means that the methods used for their computations are robust enough. We have assigned different values in its range (Table IV) for the also critical parameter $maxd$ (defining the neighborhood). Leaving the remainder unchanged.

The results are once again similar to the obtained with its value in Table IV. This means that this parameter is well controlled under the consideration of the smoothness constraint violation, so that its possible negative effect is minimized under the prevision of smoothness violation. The above reflection is applicable for the parameters U_0 and H . The parameters involved in the module SV have little relevance, their values vary within the ranges in Table IV, and so the results are always similar to the obtained with their fixed values. This is also applicable for the thresholds in UA, this is explained because the SA relaxation process achieves a stable equilibrium and the parameters in UA have small relevance and a great number of decisions are made without ambiguity. From the above experiments, the following general conclusions can be inferred:

1) *Experiments under the settings of this paper:* We can see from Fig. 4(b) that there are not local minima in the averaged

energy. The stopping criterion is achieved before the maximum number of iterations is reached ($t_{max} = 100$). The absence of local minima in the energy is also reported in [13] and [43]. From Fig. 4, the number of pairs that change their states and the energy drop in a relatively smooth fashion. This means that the number of correct matches grows slowly during the SA global matching process. This is because initially (iteration 0) there is a relatively high percentage of true matches established by the SV module, this is also reported in [23]. From iteration number 33, np and the energy vary slowly achieving a high degree of performance.

2) *Effect of the ε threshold:* By setting $\varepsilon = 0$, the maximum number of iterations ($t_{max} = 100$) is reached. This means that the equilibrium is not achieved before such a limit. From Fig. 4(a) (dotted line b), we can see that the number of pairs changing varies slowly as compared with the continuous line a . Nevertheless, the energy remains practically unchanged from the iteration 65, Fig. 4(b) (dotted line b). No improvement is achieved with this setting, justifying its value in this paper.

3) *Effect of the thresholds in the module FA:* Once again the maximum number of iterations is achieved, there are a greater number of pairs of edge segments involved. The slope of the dotted line c in Fig. 4(a) is similar to the dotted line b . The energy in figure Fig. 4(b) (dotted line c) is greater than before. This means that the states, on average, have values smaller than in the previous cases. Hence, the final decision is more complicated, resulting in a worse performance in percentage of successes.

IV. CONCLUDING REMARKS

The stereo correspondence problem is formulated as two linked processes: local (classification) and global (optimization). A global process after a local process improves the results substantially. In global strategies, the consideration of the smoothness and ordering violations also improves the results when violations occur. The proposed SVSA approach performs better than the remainder methods, although it obtains results

very close to other techniques where violations are considered. The ordering constraint is less relevant than the smoothness, perhaps this is the reason by why ordering is not applied in classical stereovision approaches [4], [8], and [24], although it is useful in structural matching methods [11].

Although the number of thresholds is high we have carried out different experiments achieving a control over them. So, in the SV method they have little relevance and in the SA the values for the critical coefficients are selected based on robust existing procedures or its risk is minimized (*maxd*). This means that the values in this approach can be used in other applications under perhaps slight modifications. This fact and the performance of the proposed SVSA approach make an important contribution for solving the stereovision matching problem. Any sequential implementation of relaxation approaches is computational intensive. Although the time for each iteration depends on the number of edge segments to be matched, on average, each iteration takes about 1.2 s under an Intel Pentium IV at 2.5 GHz. The machine dependence is high and the recommendation is: its implementation under a parallel architecture.

Since in the SA the current state of a system that has experienced a transition depends only on the previous state, it follows that SA has the Markov property [40]. Hence, the Markov random fields could be considered in the future as a potential mathematical theory, where the sum of clique potentials is the energy function to be minimized as in [47].

ACKNOWLEDGMENT

The authors wish to acknowledge the constructive recommendations provided by the reviewers.

REFERENCES

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1/2/3, pp. 7–42, 2002.
- [2] S. Barnard and M. Fishler, "Computational stereo," *ACM Comput. Surv.*, vol. 14, pp. 553–572, 1982.
- [3] D. H. Kim and R. H. Park, "Analysis of quantization error in line-based stereo matching," *Pattern Recognit.*, vol. 8, pp. 913–924, 1994.
- [4] G. Medioni and R. Nevatia, "Segment based stereo matching," *Comput. Vis., Graph., Image Process.*, vol. 31, pp. 2–18, 1985.
- [5] J. Banks and M. Bannamoun, "Reliability analysis of the rank transform for stereo matching," *IEEE Trans. Syst., Man, Cybern. B*, vol. 31, pp. 870–880, Dec. 2001.
- [6] L. Tang, C. Wu, and Z. Chen, "Image dense matching based on region growth with adaptive window," *Pattern Recognit. Lett.*, vol. 23, pp. 1169–1178, 2002.
- [7] W. E. L. Grimson, "Computational experiments with a feature-based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 17–34, Jan. 1985.
- [8] D. Marr and T. Poggio, "A computational theory of human stereovision," in *Proc. R. Soc. Lond.*, vol. B207, 1979, pp. 301–328.
- [9] M. S. Mousavi and R. J. Schalkoff, "ANN implementation of stereo vision using a multilayer feedback architecture," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, pp. 1220–1238, Aug. 1994.
- [10] Y. Ruichek and J. G. Postaire, "A neural network algorithm for 3-D reconstruction from stereo pairs of linear images," *Pattern Recognit. Lett.*, vol. 17, pp. 387–398, 1996.
- [11] K. L. Boyer and A. C. Kak, "Structural stereopsis for 3-D vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 144–166, Mar. 1988.
- [12] W. Hoff and N. Ahuja, "Surface from stereo: Integrating feature matching, disparity estimation, and contour detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 121–136, Mar. 1989.
- [13] G. Pajares, J. M. Cruz, and J. Aranda, "Relaxation by Hopfield network in stereo image matching," *Pattern Recognit.*, vol. 31, no. 5, pp. 561–574, 1998.
- [14] G. Pajares, J. M. Cruz, and J. A. López-Orozco, "Relaxation labeling in stereo image matching," *Pattern Recognit.*, vol. 33, pp. 53–68, 2000.
- [15] G. Pajares and J. M. Cruz, "The nonparametric Parzen's window in stereo vision matching," *IEEE Trans. Syst., Man, Cybern. B*, vol. 32, pp. 225–230, Aug. 2002.
- [16] F. Candocia and M. Adjouadi, "A similarity measure for stereo feature matching," *IEEE Trans. Image Processing*, vol. 6, pp. 1460–1464, Oct. 1997.
- [17] J. P. P. Starink and E. Backer, "Finding point correspondences using simulated annealing," *Pattern Recognit.*, vol. 28, no. 2, pp. 231–240, 1995.
- [18] D. M. Wuescher and K. L. Boyer, "Robust contour decomposition using a constraint curvature criterion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 41–51, Jan. 1991.
- [19] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision Vols. I and II*. Reading, MA: Addison-Wesley, 1992, 1993.
- [20] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene labeling by relaxation operation," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 420–453, Nov./Dec. 1976.
- [21] N. M. Nasrabadi, "A stereo vision technique using curve-segments and relaxation matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 566–572, Jan. 1992.
- [22] K. E. Price, "Relaxation matching techniques—a comparison," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 5, pp. 617–623, 1985.
- [23] W. J. Christmas, J. Kittler, and M. Petrou, "Structural matching in computer vision using probabilistic relaxation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 749–764, Aug. 1995.
- [24] N. M. Nasrabadi and C. Y. Choo, "Hopfield network for stereovision correspondence," *IEEE Trans. Neural Networks*, vol. 3, pp. 123–135, Jan. 1992.
- [25] S. Anily and A. Federgruen, "Simulated annealing methods," *J. Appl. Probabil.*, vol. 28, pp. 657–666, 1987.
- [26] A. Huertas and G. Medioni, "Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 651–664, May 1986.
- [27] J. G. Leu and H. L. Yau, "Detecting the dislocations in metal crystals from microscopic images," *Pattern Recognit.*, vol. 24, pp. 41–56, 1991.
- [28] M. S. Lew, T. S. Huang, and K. Wong, "Learning and feature selection in stereo matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 869–881, Sept. 1994.
- [29] E. P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*. New York: Springer-Verlag, 1989.
- [30] S. Tanaka and A. C. Kak, "A rule-based approach to binocular stereopsis," in *Analysis and Interpretation of Range Images*, R. C. Jain and A. K. Jain, Eds. Berlin, Germany: Springer-Verlag, 1990, pp. 33–139.
- [31] R. Nevatia and K. R. Babu, "Linear feature extraction and description," *Comput. Vis., Graph., Image Process.*, vol. 13, pp. 257–269, 1980.
- [32] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 2000.
- [33] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory and Methods*. New York: Wiley, 1998.
- [34] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [35] Y. S. Kim, J. J. Lee, and Y. H. Ha, "Stereo matching algorithm based on modified wavelet decomposition process," *Pattern Recognit.*, vol. 30, no. 6, pp. 929–952, 1997.
- [36] E. Kreszig, *Advanced Engineering Mathematics*. New York: Wiley, 1983.
- [37] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [38] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *J. Stat. Phys.*, vol. 34, pp. 975–984, 1984.
- [39] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.
- [40] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan, 1994.
- [41] B. Hajek, "Cooling schedules for optimal annealing," *Math. Oper. Res.*, vol. 13, pp. 311–329, 1988.
- [42] P. M. J. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Norwell, MA: Kluwer, 1989.
- [43] S. T. Barnard, "Stochastic stereo matching over scale," *Int. J. Comput. Vis.*, vol. 3, no. 1, pp. 17–32, 1989.
- [44] S. Hattori, A. Okamoto, and H. Hasegawa, "Stereo matching by simulated annealing incorporating a diffusion equation," in *Proc. ASPRS Annu. Conf.*, 1998, pp. 1030–1041.

- [45] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.
- [46] N. Yokoya, T. Shakunaga, and M. Kanbara, "Passive range sensing techniques: Depth from images," *IEICE Trans. Inform. Syst.*, vol. E82, no. 3, pp. 523–533, 1999.
- [47] A. Fusiello, U. Castellani, and V. Murino, "Relaxing symmetric multiple windows stereo using Markov random fields," in *Proc. Energy Minimization Methods Computer Vision Pattern Recognition*, Berlin, Germany, 2001, pp. 91–105.



Gonzalo Pajares received the M.Sc and Ph.D. degrees in physics from the Universidad Nacional de Educación a Distancia, Madrid, Spain, in 1987 and 1995, respectively.

He worked at Indra, Madrid, in critical real-time software development since 1990. He also was working in Indra Space, Madrid, developing remote sensing applications. He joined the Complutense University, Madrid, as Associate Professor in 1995 and Professor in 2004 teaching robotics and computer visual perception, including image

processing and computer vision tasks. His current research interests include robotics vision systems, medical imaging, and remote sensing applications in registering, fusion, change detection, and superresolution in video sequences.



Jesús Manuel de la Cruz received the M.Sc and Ph.D. degrees in physics from the Complutense University, Madrid, Spain in 1979 and 1984, respectively.

From 1985 to 1990 he worked in the Department of Automatic Control at the Universidad Nacional de Educación a Distancia, Madrid, and from October 1990 to 1992, the Department of Electronic, University of Santander, Madrid.. In October 1992, he joined the Department of Computer Science and Automatic Control, Complutense University,

where he is currently a Professor. His research interest include robotics vision systems, fusion sensors, and applications of automatic control to robotics and flight control.