
ELSEVIER

Pattern Recognition Letters 22 (2001) 1457–1473

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Local stereovision matching through the ADALINE neural network

Gonzalo Pajares *, Jesús M. de la Cruz

Dpto. Arquitectura de Computadores y Automática, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

Received 24 September 1999; received in revised form 20 March 2001

Abstract

This paper presents an approach to the local stereovision matching problem using edge segments as features with four attributes. Based on these attributes we compute a matching probability between pairs of features of the stereo images. A correspondence is said to be true when this probability is maximum. The probability value is a weighted sum of the attributes. We use two combined ADALINE neural networks to compute the weight for each attribute. A comparative analysis among other recent matching methods is illustrated. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Neural networks; ADALINE; Delta rule; Training; Learning; Stereovision; Matching; Similarity

1. Introduction

A number of research efforts in the computer vision community have been directed towards the study of the three-dimensional (3-D) structure of objects using machine analysis of images (Dhond and Aggarwal, 1989). The analysis of stereo images has become an important passive method for extracting the 3-D structure of a scene.

The key step in stereovision is image matching, namely, the process of identifying the corresponding points in two images that are generated by the same physical point in the 3-D scene. This paper is devoted solely to this problem. The stereo correspondence problem can be defined in terms of finding pairs of true matches that satisfy three competing constraints: similarity, smoothness and uniqueness (Marr and Poggio, 1979). The similarity constraint is associated to a local matching process where a metric is used to measure the degree of correspondence between features through the attribute values. The results computed in the local process are later used by a global matching process where other constraints such as smoothness (Marr and Poggio, 1979), minimum differential disparity (Medioni and Nevatia, 1985), and figural continuity (Pollard et al., 1981) are imposed. A good choice of a local matching strategy is the key for good results in the global matching process.

To measure the similarity between a pair of features we use the matching probability, in Eq. (1), taken from the classical work of Kim and Aggarwal (1987) (KA).

* Corresponding author. Tel.: +34-91-394-4477; fax: +34-91-892-2217.

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

PII: S0167-8655(01)00097-6

$$p_{\mathrm{lr}} = \sum_{j=1}^{n} w_{\mathrm{S}j} \frac{1}{1 + |av_{\mathrm{l}j} - av_{\mathrm{r}j}|},$$
$$\text{where } \sum_{j=1}^{n} w_{\mathrm{S}j} = 1; \quad 0 \leqslant w_{\mathrm{S}j} \leqslant 1, \tag{1}$$

where *av* denotes the *attribute value* corresponding to each feature; the subscripts l and r denote features belonging to the left and right images, respectively, in the pair of features to be matched; $w_{\mathrm{S}j}$ is the associated specific weight for each attribute j and n is the number of attributes. The choice of this matching probability is because it considers the relative importance of each attribute and this is a key issue in our approach.

The criterion used in KA is that the weights in Eq. (1) are fixed arbitrarily to a constant value. Instead of fixing the specific weights with the KA criterion, we develop a supervised learning strategy based on two combined ADALINE (ADAptive LINear Element) neural networks (Kohonen, 1989, 1995; Patterson, 1996; Wu, 1994) to compute and update the specific weights during the training process. The learning of the weights through the two ADALINEs makes up the finding and contribution to the matching problem. Additionally, the computation of the specific weights allows us to take into account the relative importance of each attribute during the computation of the matching probabilities. This represents an important improvement with respect to the idea outlined in the method based on the perceptron criterion function in (Cruz et al., 1995b), where a unique perceptron was used and the computation of the weights was a consequence, but not the key issue.

1.1. Techniques in stereovision matching

Two types of techniques have been broadly used for stereovision matching (Dhond and Aggarwal, 1989; Ozanian, 1995; Pajares, 1995; Trucco and Verri, 1998), namely the correlation-based and the feature-based methods. In the correlation-based method, the elements to be matched are image windows of fixed size and the similarity criterion is a measure of the correlation between windows in the two images (Fua, 1993). The corresponding element is given by the window that maximizes the similarity criterion within a search region. The number of pairs of features to be considered becomes high, because all pixels in the left image must be matched with all pixels in the right one. The feature-based methods use sets of pixels with similar attributes, usually either pixels belonging to edges (Kim and Aggarwal, 1987; Marr and Poggio, 1979; Mousavi and Schalkoff, 1994; Pollard et al., 1981) or the corresponding edges themselves (Ayache and Faverjon, 1987; Ayache, 1991; Kim et al., 1992; Cruz et al., 1995a,b; Hoff and Ahuja, 1989; Medioni and Nevatia, 1985; Pajares, 1995). Instead of image windows, they use numerical and symbolic properties of features. As shown in (Dhond and Aggarwal, 1989), feature-based methods lead to a sparse depth map only, leaving the rest of the object surface to be reconstructed by interpolation. They are faster than area-based methods, because there are fewer points (features) to be considered.

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

There are *intrinsic* and *extrinsic* factors affecting the stereovision matching system. The extrinsic factors are due to external interactions with the system, since, in a practical stereovision system, the left and right images are obtained with two cameras placed at different positions/angles. Although they both capture the same scene, each camera may receive different illumination and also incidentally different reflections. In contrast, the intrinsic factors come from internal interactions in the system; this is because 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, and hence each camera may convert the same illumination level into a different gray level.

Hence, due to the above-mentioned factors, the corresponding attributes in the two images may display different values. This may lead to incorrect matches. Thus, it is very important to find features in both images which are independent of possible variations in the images (Wuescher and Boyer,

1991). Our experiment has been carried out in an indoor space where the edge segments are abundant, making suitable such features (Trucco and Verri, 1998). Moreover, they have been studied in terms of reliability (Breuel, 1996) and robustness (Wuescher and Boyer, 1991) and, as mentioned before, have also been used in previous stereovision matching works. This fact justifies our choice of features, although such features are produced by intensity changes. Four attribute values (module and direction of the gradient vector, Laplacian and variance) are computed for each edge segment. The Laplacian and variance are such functions of the gray levels within a small neighborhood of a given pixel; then they are computed for each edge segment as explained in Section 2.1.

1.3. Some learning strategies in stereovision matching

As mentioned in the previous section, there are extrinsic and intrinsic factors affecting the stereovision matching system. The effects of the extrinsic factors have been broadly considered in the literature. But this is not the case for the effects of the intrinsic ones. This paper deals with both kinds of effects but it is mainly concerned with the effects produced by the intrinsic factors because we have verified their importance and, as a result, a research line is open, including: (1) statistical unsupervised learning (SUL) strategies (Cruz et al., 1995a; Pajares, 1995); (2) self-organizing feature mapping (SOM) (Pajares et al., 1998b); (3) supervised stereovision matching based on learning vector quantization (LVQ) (Pajares et al., 1998c); (4) stereovision matching based on Hebbian learning (SHL) (Pajares et al., 1999); (5) the perceptron criterion function (PCF) (Cruz et al., 1995b).

In SUL, SOM, LVQ and SHL, two four-dimensional vectors $\boldsymbol{x}_l$ and $\boldsymbol{x}_r$ are associated to each pair of features, where the vector components are the attribute values and the subscripts l and r are as defined previously. A four-dimensional difference measurement vector $\boldsymbol{x}$ is obtained from the above $\boldsymbol{x}_l$ and $\boldsymbol{x}_r$ vectors, where its components are the differences for the module, and the direction of the gradient vector, the Laplacian and the variance, respectively. Such vectors, denoted $\boldsymbol{x} = \{x_m, x_d, x_l, x_v\}^t$, where the superscript t indicates the transpose operation, are the inputs for the stereovision matching system. The $\boldsymbol{x}$ vectors for true matches cluster around a center. This center should be the origin, i.e. the null vector, for an ideal system. Unfortunately this rarely occurs and such a center is learned by the system. The dispersion of these difference vectors in the cluster is measured by the covariance matrix. Once the system has been trained with all available samples, the center and the covariance matrix are used to compute the Mahalanobis distance (Duda and Hart, 1973; Maravall, 1993) for each incoming $\boldsymbol{x}$ difference vector associated with a new stereo pair of images. We use the minimum distance criterion to classify the pair of features as a true match when such a distance value is less than a previously defined threshold. Otherwise, it is considered as a false match. A comparative analysis between LVQ, SUL and SOM shows that LVQ yields better results than SUL and SOM due to the use of the supervised learning scheme.

The PCF is based on the delta rule, as is the method of this paper, but a single neuron is used instead of the two neurons of this approach and the correspondence is directly established with the computed weights instead of the probability used here.

In a different way the global stereovision matching algorithm of Lew et al. (1994) (LHW) integrates learning, feature selection and surface reconstruction, using points as matching primitives. Lew et al. (1994) propose the following possible attributes associated to each point (x, y): intensity, magnitude and orientation of the gradient vector, Laplacian and curvature. The goal is to establish the correspondence between a template point (x_p, y_p) and a matching point (x_c, y_c). The central idea in LHW is to find a subset of these attributes for (x_p, y_p) that will uniquely define the point. But, from LHW, we only exploit the conclusion about the importance of the attributes for the matching, as in this paper.

1.4. Paper organization

This paper is organized as follows. In Section 2 the stereo matching system is considered. A

training phase is outlined through the two ADALINEs with the embedded learning law and a subsequent current stereo matching process is proposed. In Section 3 a test strategy is designed and a comparative analysis is performed. Finally, the conclusion is presented in Section 4.

2. Two ADALINEs for stereovision matching

Our local stereo matching system is designed with a parallel optical axis geometry and is composed of three basic modules: (1) *image analysis*, (2) *training* and (3) *current stereo matching*. The function of the image analysis module is to extract the features and their attributes from the scene and to make this information available for training or for current stereo matching. The image analysis module also realizes an initial correspondence of pairs of features.

In Section 2.1 we explain how the features are extracted and how their attributes are computed. A pair of edge segments is a potential match if the two segments verify the following three initial conditions: (1) the absolute value of the difference in the gradient vector direction between the two segments is below a specific threshold, set at 20°; (2) the absolute value of the difference in the gradient magnitude is also below a fixed threshold, set at 10; (3) the overlap rate surpasses a given value, set at 0.8. The "overlapping" is a concept introduced in (Medioni and Nevatia, 1985) as follows: "*two segments overlap if by sliding one of them in a direction parallel to the epipolar line*, *i.e. to an horizontal line*, *they would intersect*". We apply this concept to a stereo pair of edge segments. This stereo pair is made up of an edge segment of the left image and an edge segment of the right image. The axes of the two cameras of the stereo image system are previously aligned, so that the images are also aligned, i.e. the vertical positions of the images of any point in the scene are approximately similar in both images. Fig. 1 clarifies this overlapping concept. Indeed, segment *a* in the left

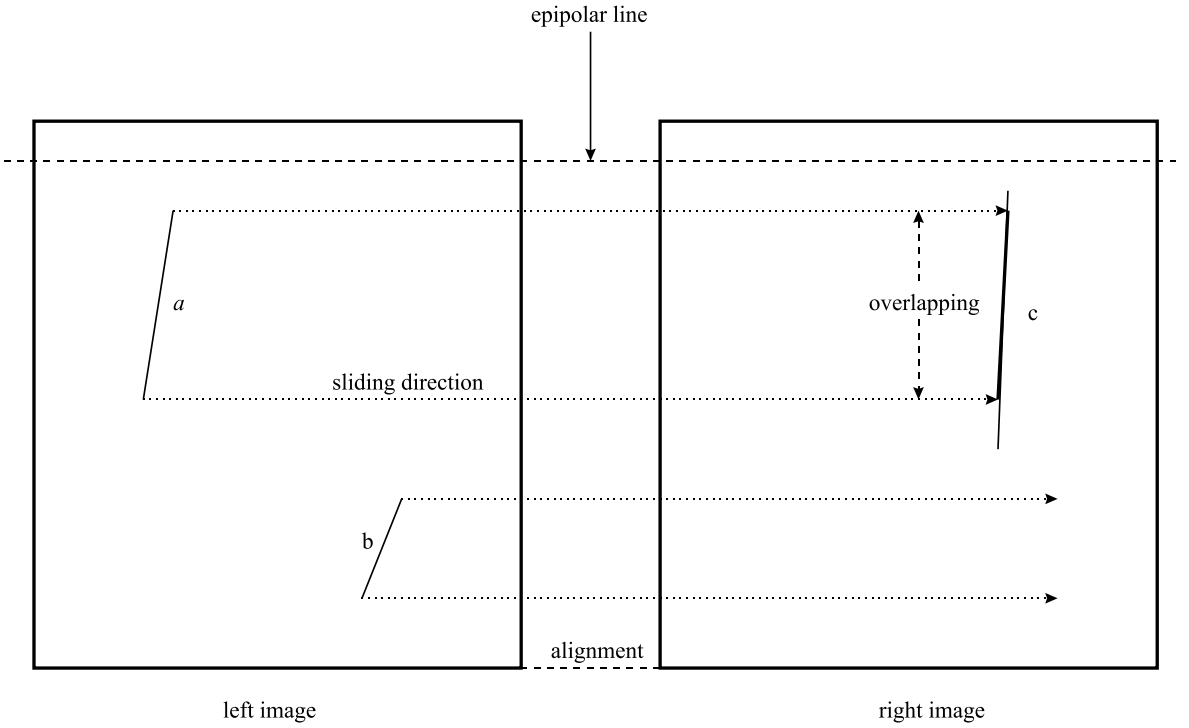

Fig. 1. Overlapping concept between edge segments.

image overlaps with segment c in the right image, but segment b does not overlap. The overlap rate α, given in Eq. (2), is defined as the percentage of coincidence when two segments overlap. It is computed by taking into account the common overlap length and the two lengths of the involved edge segments.

$$\alpha = \frac{2l_c}{l_l + l_r}, \tag{2}$$

where l_c is the common overlap length, and l_l, l_r are the corresponding lengths for the left and the right edge segments under matching, respectively. All lengths are measured in pixels. In Fig. 1, l_l and l_r are the lengths of segments a and c, and l_c is the common length between a and c.

The system works in two mutually exclusive modes: OFF-LINE or training process and ON-LINE or current stereo matching process. In the two modes, the features and their corresponding attributes are extracted from the images. In the OFF-LINE mode the system updates, through the corresponding training process, the synaptic weight vectors attached to each ADALINE. This is a supervised learning process because an unknown set of weights is estimated from known input and output samples (see Section 2.2). During the ON-LINE process, an incoming pair of features with their associated attributes is presented and processed by the system. The updated synaptic weight vectors are used to decide if it is a true or a false match. This operating mode is similar to the one proposed in the PCF and LVQ methods.

2.1. Feature and attribute extraction

For clarity and completeness, we describe the method to extract feature and attributes which can also be found in SUL, PCF, SOM, LVQ or SHL. The edge pixels in both images are extracted using the Laplacian of Gaussian filter in accordance with the zero-crossing criterion described in (Huertas and Medioni, 1986). At each zero-crossing in a given image we compute the magnitude and the direction of the gradient vector as in (Leu and Yau, 1991), the Laplacian as in (Lew et al., 1994) and the variance as in (Krotkov, 1989). These four attributes are computed from the gray levels of a central pixel and its eight immediate neighbors. The gradient magnitude is obtained by taking the largest difference in gray levels of two opposite pixels in the corresponding eight-neighborhood of a central pixel. The gradient direction points from the central pixel towards the pixel with the maximum absolute value of the two opposite pixels with the largest difference. It is measured in degrees, quantified by multiples of 45. The Laplacian is computed by using the corresponding Laplacian operator over the eight neighbors of the central pixel. The variance indicates the dispersion of the nine gray-level values in the eight-neighborhood of the same central pixel. These attributes are selected because we are using edge segments as features, and therefore pixels close to an edge segment are characterized by high local gray-level variations. The gradient, the Laplacian as operators and the statistical variance provide an appropriate criterion to measure these local changes in gray levels (Krotkov, 1989).

Our stereo matching system is part of a general robotics navigation system, which must cope with undesired events, such as objects very close to the cameras (see Fig. 7, object labeled as 9–10 in the left image and 11–12 in the right image). Such out-of-focus objects can be detected by measuring the grade of focus. A criterion to measure the focus quality is provided by the Laplacian, the variance and the gradient magnitude.

The gradient magnitude and/or the gradient direction are generally used for local stereovision procedures involving edge pixels or edge segments (Kim and Aggarwal, 1987; Medioni and Nevatia, 1985; Ayache, 1991; Kim et al., 1992; Kahn et al., 1990; Lew et al., 1994; Cruz et al., 1995a,b; Pajares, 1995; Mousavi and Schalkoff, 1994; Ruichek and Postaire, 1996; Pajares et al., 1998a,b,c), although some other global strategies could be used later in global stereovision matching approaches for improving the local matching results, when these local results are considered unsatisfactory, e.g. *relaxation* as in (Kim and Aggarwal, 1987; Medioni and Nevatia, 1985; or Pajares, 1995); *structural stereopsis* as in (Ayache, 1991); *optimization by the Hopfield neural network* as in (Ruichek and Postaire, 1996; Mousavi and Schalkoff, 1994; or Pajares et al., 1998a). From an analysis of

the bibliography, we have found that the gradient is in most cases the unique attribute used in stereovision similarity metrics (i.e. in local stereovision matching approaches). The Laplacian, used by Lew et al. (1994), has interesting discriminatory properties. But, as pointed out by Maravall (1993), it is noise sensitive. Hence, due to the Laplacian noise sensitivity and under the consideration of avoiding undesired events, we use the variance value as an additional attribute to increase the reliability of our stereo matching system, so that the measurement for the grade of focus is reliable. This is valid even for attributes of the same nature; this is the case for the gradient vector and the Laplacian, which are both based on derivative operations of first and second order, respectively. We conclude that the four selected attributes have all discriminatory properties after an exhaustive study based on the correlation derived from the covariance matrix for the differences in attributes for the cluster of true matches in LVQ, SOM and SHL. Note that the covariance matrix provides cross-correlation information between attribute values. We call undesired events, under the robotic navigation perspective, when an object is close to the robot and a collision is possible. Objects very close to the stereovision system (i.e. to the robot) are detected by measuring the grade of focus.

Once the edge-pixels are detected by means of the zero-crossings determination procedure, we use the following two strategies for extracting the edge segments or features:

1. Adjacent zero-crossings are connected if their corresponding differences in gradient magnitude and gradient direction do not overpass the quantities of $\pm 20\%$ and $\pm 45°$, respectively (Tanaka and Kak, 1990).
2. Each detected contour according to the preceding algorithm is approximated by a series of piecewise linear line segments (Nevatia and Babu, 1980).

We consider the four attributes for all the pixels standing along each piecewise linear line segment and for each attribute an average value is finally obtained. For computing the mean gradient direction we apply circular statistics as described by Mardia (1972). All average attribute values are scaled, so that they fall within the same range. Finally, these averaged values are the attributes associated to the given edge segment and their contribution during the computation of the specific weights are the same. Moreover, each edge segment is identified with its initial and final pixel coordinates, its length and its label.

Hence, given a stereo pair of edge segments, one coming from the left image and the other from the right image, we have four associated attributes for each of them (i.e. two groups of four attributes). With these two groups of attributes, we make up two four-dimensional vectors $\boldsymbol{x}_\mathrm{l}$ and $\boldsymbol{x}_\mathrm{r}$, with components equal to the averaged attribute values associated with each edge segment. An ideal true match fulfills $\boldsymbol{x}_\mathrm{l} = \boldsymbol{x}_\mathrm{r}$. Nevertheless, in any real system and due to the intrinsic and extrinsic factors, $\boldsymbol{x}_\mathrm{l}$ differs at least lightly from $\boldsymbol{x}_\mathrm{r}$.

2.2. Training process

We base our approach on the Widrow–Hoff *delta rule* for learning (Patterson, 1996; Wu, 1994; Hilera and Martinez, 1995) which is applied to two single-layer feedforward linear networks, with two real-valued outputs (cf. Fig. 2). Each single neural network model is an ADALINE.

The inputs of each neuron are the components of the four-dimensional $\boldsymbol{x}_\mathrm{l}$ and $\boldsymbol{x}_\mathrm{r}$ vectors, respectively (hereafter written as $\boldsymbol{x}_\mathrm{m}$ where the sub-index m denotes either l or r). The input vectors in a neural network are the training patterns to perform the updating process. Each component $x_{\mathrm{l}i}$ or $x_{\mathrm{r}i}$ is an input attached to a synapse. Each element has four synapses, as we are processing pairs of features with four attributes. As usual, a weight is associated to each synapse. Hence, in our approach we also define the corresponding four-dimensional synaptic weight vector as $\boldsymbol{w}_\mathrm{m} = \{w_{\mathrm{m}1}, w_{\mathrm{m}2}, w_{\mathrm{m}3}, w_{\mathrm{m}4}\}$, which is updated during the training process. The sub-indices 1, 2, 3 and 4 in the components of $\boldsymbol{x}_\mathrm{m}$ and $\boldsymbol{w}_\mathrm{m}$ are associated to the module and the direction of the gradient vector, the Laplacian and the variance, respectively. With this convention, our model consists of two ADALINEs and performs two weighted sums as an inner product between each input vector and the

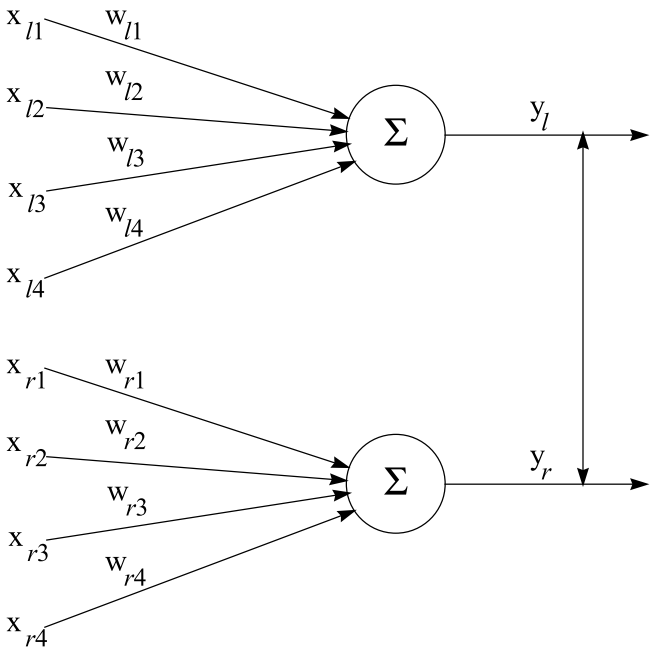

Fig. 2. Two ADALINEs applied to stereovision matching. The vertical double arrow means that the target output for each actual output is the other (i.e. the target output for y_l is y_r and vice versa).

corresponding synaptic weight vector to compute the analogue output, y_m, as

$$y_m = \boldsymbol{x}_m \boldsymbol{w}_m^t = \sum_{i=1}^{4} x_{mi} w_{mi}. \quad (3)$$

As mentioned previously, for a given ideal pair of true matches with their corresponding $\boldsymbol{x}_l$ and $\boldsymbol{x}_r$ attribute vectors, we can assert that $\boldsymbol{x}_l$ and $\boldsymbol{x}_r$ are equal (i.e. the two features of a pair of true matches have identical attribute values). Hence, we can initially choose the weight vectors $\boldsymbol{w}_m$ as unit vectors in (3).

Under the hypothetical ideal case, it is assumed that y_l is equal to y_r, so we give proper form to this fact: "for an input pattern vector $\boldsymbol{x}_l$ the desired or target analogue output value is y_r but y_l is the actual value computed by the neuron and vice versa". The vertical double arrow in Fig. 2 represents this output relation. This is the supervised concept: a known output for a given input. But as such input/output relation is an ideal case, this assertion will not be truth and we must assume $y_l \neq y_r$ and $\boldsymbol{w}_m \neq \mathbf{1}$. Hence, the goal is to learn each $\boldsymbol{w}_m$, so that the differences between y_l and y_r are minima for a given number of training patterns. To do that, we apply the delta rule where the training patterns are pairs of true matches with their corresponding attribute vectors. In (Patterson, 1996), we can find the derivation of the delta rule for the ADALINE neuron, where the goal is to adjust the weights to minimize the total squared error (E_{tot}) between the desired output and the actual weighted sum for the given number P of training patterns, so p ranges from 1 to P:

$$E_{tot}(\boldsymbol{w}_l, \boldsymbol{x}_l) = \sum_{p=1}^{P} (y_l^p - y_r^p)^2 = \sum_{p=1}^{P} (\boldsymbol{x}_l^p \boldsymbol{w}_l^t - y_r^p)^2. \quad (4)$$

As specified in (Duda and Hart, 1973), we have sought the weight vectors $\boldsymbol{w}_l$ and $\boldsymbol{w}_r$, making the inner products $\boldsymbol{x}_l^p \boldsymbol{w}_l^t$ and $\boldsymbol{x}_r^p \boldsymbol{w}_r^t$ positive, respectively, as the weight vector components and the attributes are all positive.

According to Hush and Horne (1992), E_{tot} generally defines a bowl-shaped error surface in the weight space with the solution (or family of solutions) at the bottom of the bowl. The way to

the bottom of the bowl can be found using a descent technique known as *gradient search* technique. The procedure starts from the initial weight vector $\boldsymbol{w}_{\mathrm{m}} = \mathbf{1}$, which is the ideal case, and then iteratively searches for the solution by taking small "downhill" steps on the error surface. In a gradient search procedure, the downhill direction is found by computing the negative gradient of E_{tot} with respect to the synaptic weights. Using this technique, the synaptic weights are adjusted at each step according to the Widrow–Hoff rule:

$$\boldsymbol{w}_{\mathrm{l}}(k+1) = \boldsymbol{w}_{\mathrm{l}}(k) - \eta \nabla_{\boldsymbol{w}_{\mathrm{l}}} E_{\mathrm{tot}}(\boldsymbol{w}_{\mathrm{l}}, \boldsymbol{x}_{\mathrm{l}})|_{\boldsymbol{w}_{\mathrm{l}}(k)}, \tag{5}$$

where η is a positive constant that controls the amount of adjustment and k is the iteration index. A practical choice for the parameter η, which is often referred to as the *convergence parameter* or the *learning rate*, is $0.1 < \eta < 1.0$ (Wu, 1994).

Computing the gradient in (5) and introducing a normalization factor by dividing each summation term by the squared Euclidean norm of the pattern $\boldsymbol{x}_{\mathrm{l}}^{p}$, so that the contribution of each term to the weight updating is independent of the magnitude of the inputs, finally leads to Eq. (4), which is rewritten in the form

$$\boldsymbol{w}_{\mathrm{l}}(k+1) = \boldsymbol{w}_{\mathrm{l}}(k) - \eta \sum_{p=1}^{P} \frac{(y_{\mathrm{l}}^{p}(k) - y_{\mathrm{r}}^{p})\boldsymbol{x}_{\mathrm{l}}^{p}}{\|\boldsymbol{x}_{\mathrm{l}}^{p}\|}. \tag{6}$$

The process of computing the gradient and adjusting the synaptic weights is repeated until a minimum, or a point sufficiently close to the minimum, is found. In practice, several stopping criteria can be considered (Hush and Horne, 1992). The first one is based on the magnitude of the gradient vector. The algorithm stops when this magnitude is sufficiently small since by definition the gradient will be zero at the minimum. One might also consider stopping the algorithm when E_{tot} falls below a fixed threshold. However, this requires some knowledge about the minimal value of E_{tot} which is not always available. Finally one might consider stopping when a fixed number of iterations have been performed, although there is little guarantee that the algorithm will stop at a minimum with this condition. Experimental results suggest that this last stopping condition is acceptable in our stereovision matching approach. Indeed, when the minimum is still not achieved, we have verified that each weight achieves about 98.8% of its final value with only 15 iterations. Additional iterations only modify the value at the thousandth level but make worse the time for the OFF-LINE process. This slight modification does not improves the final matching results. Hence, we apply the first stopping condition, but if a number of 15 iterations has been reached, the algorithm stops before the first condition is satisfied. These tests have been carried out with the set of images proposed in Section 3. By interchanging the subindex l by r in Eqs. (4)–(6), we obtain the same set of equations when the desired output is y_{l} and the actual weighted sum is y_{r}. There is no interrelation between $\boldsymbol{w}_{\mathrm{l}}$ and $\boldsymbol{w}_{\mathrm{r}}$.

2.2.1. Computation of the specific weights associated with the four attributes

When the OFF-LINE training process is completed, the specific weight for each attribute $w_{\mathrm{S}j}$, is updated and is available for the following ON-LINE process.

As mentioned before, the initial synaptic weight vectors $\boldsymbol{w}_{\mathrm{m}}$ are unit vectors. We define a metric to measure the distance between 1 and each left (l) or right (r) weight for each attribute. Let us consider the expression $D = \sum_{k=1}^{4} D_k$, where $D_k = |1 - w_{\mathrm{l}k}| + |1 - w_{\mathrm{r}k}| + \varepsilon$ and $k = 1, \ldots, 4$. This expression is derived from the following reasoning: "a maximum deviation from 1 corresponds to a minimum specific weight and vice-versa". Note that ε is a small quantity to avoid divisions by 0 in (7), when $w_{\mathrm{l}k}$ and $w_{\mathrm{r}k}$ are equal to one; ε is set to 0.01 in this paper.

Then, as the specific weights must satisfy the constraints in Eq. (1) we obtain the following expression:

$$w_{\mathrm{S}j} = \frac{1 - D_j/D}{\sum_{h=1}^{4}(1 - D_h/D)} = \frac{D - D_j}{3D}. \tag{7}$$

With our approach the specific weights are computed from the learned synaptic weight vectors during the OFF-LINE training process unlike these proposed by Kim and Aggarwal (1987), which are fixed ad hoc without learning. This is a fundamental improvement of our method.

2.3. Stereo matching process

The stereo matching is an ON-LINE process in which the contents of a pair of new stereo images are to be matched. As mentioned before in Section 2.2, the image analysis system extracts pairs of features and determines their corresponding four-dimensional associated attribute vectors $\boldsymbol{x}_l$ and $\boldsymbol{x}_r$. Once this information is obtained, the updated $\boldsymbol{w}_S$ vector is used. This vector becomes available after a previous OFF-LINE training process. This result is used in Eq. (1) to assign a matching probability to each incoming pair of features.

3. Comparative analysis and performance evaluation

We design a test strategy in order to: (1) compare our ADALINE (ADL)-based method against the PCF; (2) show the effectiveness of the learning process against classical criteria where the weights are constant as in KA; (3) validate the ADL method with respect to other more recent local learning methods like SHL and LVQ [it is unnecessary to compare the ADL with SOM and SUL as in (Pajares et al., 1998c) a comparison is explicitly given]; (4) compare our most significant attributes with the set of attributes used in the algorithm of Lew et al. (1994).

The experimental set up is carried out into an indoor space (a robotics laboratory) where all objects are manmade, and the edge segments are the dominant features. The range of distances, for the different objects, varies from 30 cm (objects labeled as 9–10 in the left image and 11–12 in the right image in Fig. 7) to 15 m (blackboard in Fig. 5).

3.1. Design of a test strategy

The objective is to prove the interest of the method by varying indoor environmental conditions in two ways: by using new images with different features (different objects) and by changing the illumination. With this aim in mind a set SP0 of 12 pairs of stereo images captured with natural illumination are used to extract initial input vectors. Figs. 3–6 show four representative left images of this set. Three other different sets of stereo images, SP1, SP2 and SP3, are used for the test. They are composed of 10, 10 and 15 stereo images, respectively. The total number of pairs of edge segments extracted from these images is 2132. This number of pairs of edge segments is representative of the environment where our mobile robot, equipped with our stereovision system, navigates. A representative stereo image pair is shown for each SP set in Figs. 7(a) and (b), 8(a) and (b) and 9(a) and (b). The set SP1 of stereo images has been captured with natural illumination, as the initial SP0 stereo image samples, and the sets SP2 and SP3 with artificial illumination.

The process can be summarized as follows:

STEP 0:
OFF-LINE → Initialize $\boldsymbol{w}_m$ to $\{1, 1, 1, 1\}$
OFF-LINE → Update $\boldsymbol{w}_m$ through (6) with the true matches in the set SP0
STEP 1:
ON-LINE → Classify the pairs of features as true or false matches in the sets SP1 and SP3
OFF-LINE → Update $\boldsymbol{w}_m$ through (6) with the true matches in the set SP1
STEP 2:
ON-LINE → Classify the pairs of features as true or false matches in the set SP2
OFF-LINE → Update $\boldsymbol{w}_m$ through (6) with the true matches in the set SP2
STEP 3:
ON-LINE → Classify the pairs of features as true or false matches in the set SP3
OFF-LINE → Update $\boldsymbol{w}_m$ through (6) with the true matches in the set SP3

Any unsupervised stereovision matching method can be used for supplying as many true matches as possible during STEP 0. We have used the SUL, as it provides good results. When this is not possible, the use of a minimum distance criterion is appropriate and the Euclidean distance is sufficient. Also, a human expert could provide such true matches, although this implies that the system loses its automatic capability at this stage.

Note that in STEP 1 the pairs of features from SP1 and SP3 are classified, but only those coming from SP1 are used during the OFF-LINE process.

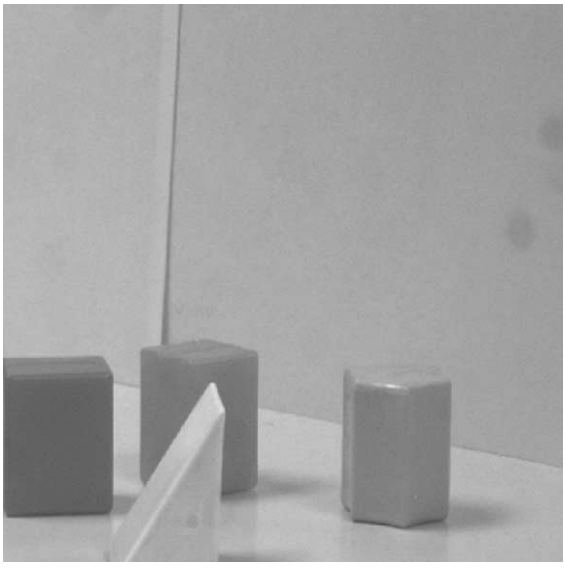

Fig. 3. Left original training image (blocks).

Fig. 4. Left original training image (furniture).

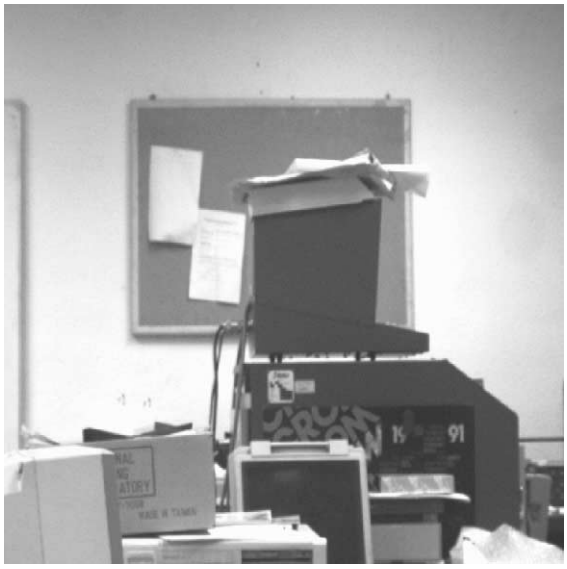

Fig. 5. Left original training image (computers I).

Fig. 6. Left original training image (computers II).

This is because the pairs of features for the set SP3 are again ON-LINE-processed in STEP 3. Otherwise, the knowledge acquired by the system in STEP 1 through SP3 should be used to classify this same set SP3 in STEP 3. This is undesired, since, in STEP 3, it is intended to show better results than those obtained in STEP 1 for the set SP3 because the system has learned, in STEP 2 through SP2, about images with the same type of illumination as that used for capturing SP3. Table 1 shows the specific weight vectors $\boldsymbol{w}_S$, obtained at the completion of each step.

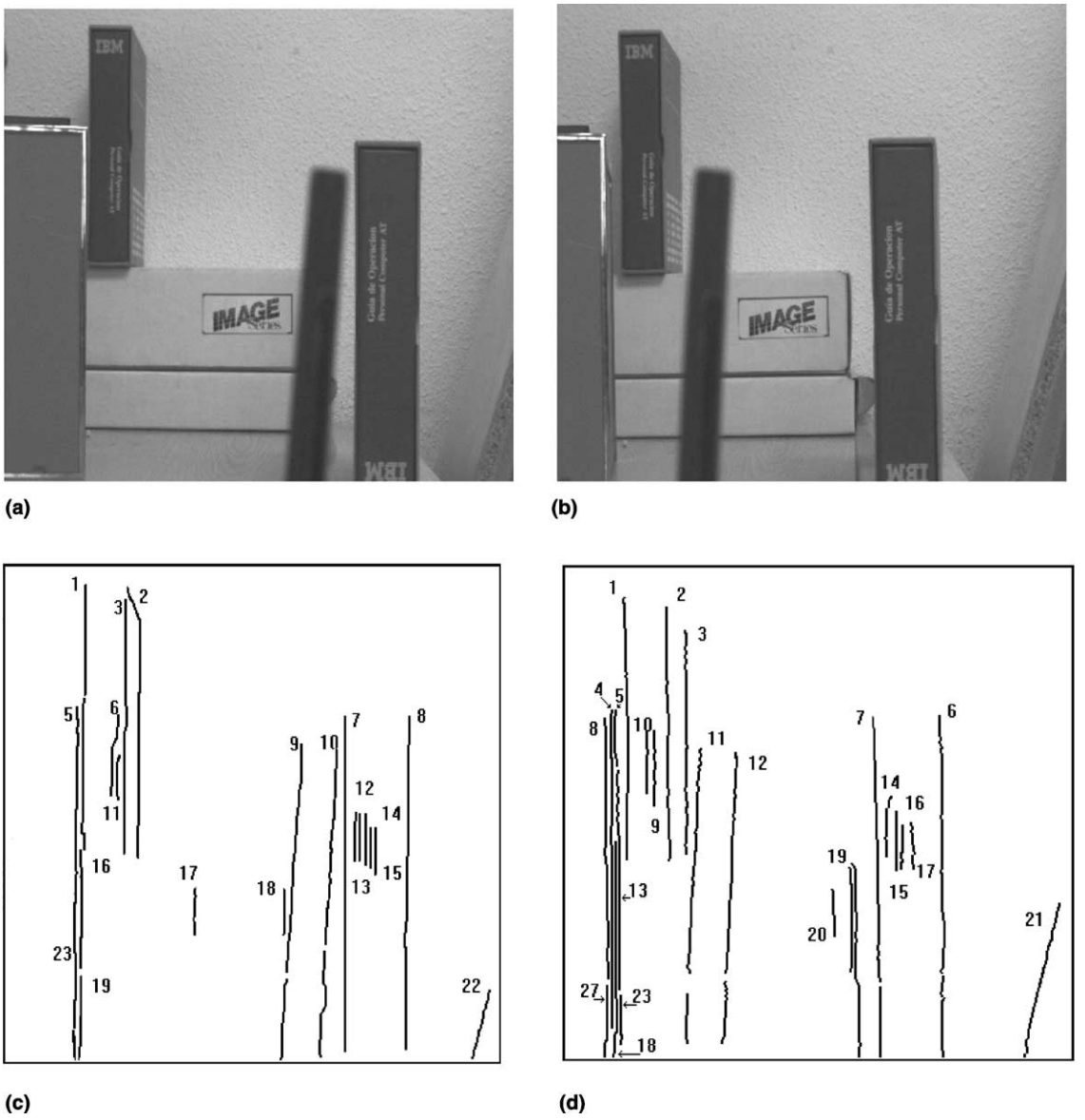

Fig. 7. (a) SP1: original left stereo image. (b) SP1: original right stereo image. (c) SP1: labeled segments left image. (d) SP1: labeled segments right image.

The importance of the attributes for matching, through the specific weights, can be derived from the results in the Table 1 as follows. According to Eq. (7), each specific weight contains information about the relative importance of each attribute when it is compared to the other three attributes. Hence, according to the computed w_S in the four steps, the most important attribute for matching is the gradient direction, followed by the gradient magnitude. These two attributes have been used by Medioni and Nevatia (1985), Lew et al. (1994), Ruichek and Postaire (1996), Mousavi and Schalkoff (1994), Cruz et al. (1995a,b), Pajares (1995) and Pajares et al. (1998a,b,c, 1999) among

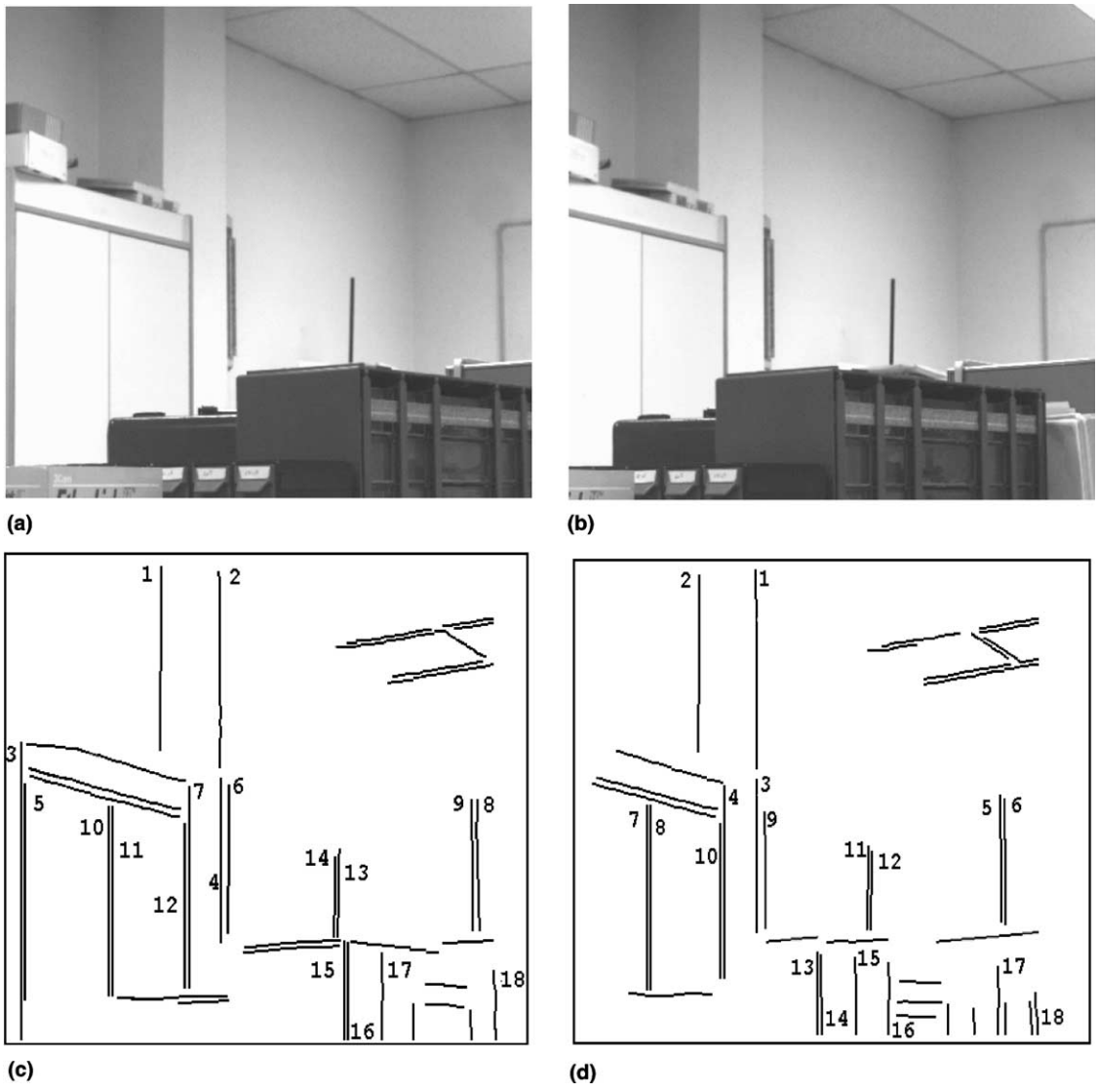

Fig. 8. (a) SP2: original left stereo image. (b) SP2: original right stereo image. (c) SP2: labeled segments left image. (d) SP2: labeled segments right image.

others. In contrast, the variance and the Laplacian are, in that order, the less important attributes for matching.

3.2. Comparative analysis

Based on Eq. (1), we compute the matching probabilities between edge segments. These probability values allow us to compare the effectiveness of our ADL against the classical criterion used by KA where no learning is involved and where the specific weights components are fixed to 0.25, so that the four attributes have the same relative importance. During the decision process, there are unambiguous and ambiguous pairs of features, depending on whether a given left image segment corresponds to one and only one, or to several right image segments, respectively. In any case, the

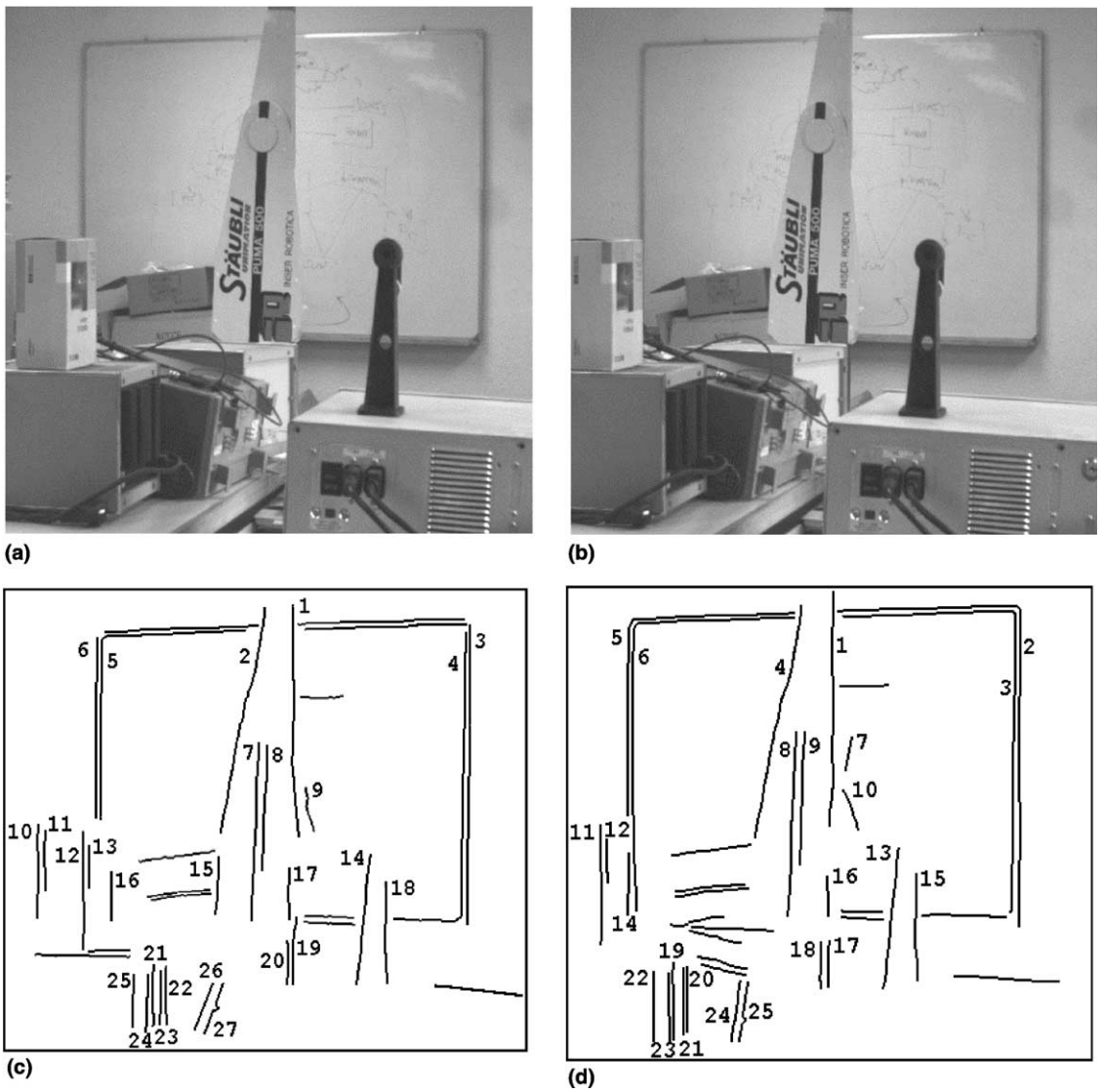

Fig. 9. (a) SP3: original left stereo image. (b) SP3: original right stereo image. (c) SP3: labeled segments left image. (d) SP3: labeled segments right image.

Table 1
Specific weight vectors w_S obtained at the completion of each step

	STEP 0	STEP 1	STEP 2	STEP 3
w_S	$\{0.26, 0.30, 0.24, 0.20\}$	$\{0.26, 0.33, 0.23, 0.18\}$	$\{0.25, 0.36, 0.22, 0.17\}$	$\{0.28, 0.41, 0.19, 0.11\}$

decision about the correct match is made by choosing the pair with the greater probability value (in the unambiguous case, there is only one) as long as it surpasses the threshold of 0.50 (intermediate probability value).

We have tested a large number of values for η in the range $[0, 1]$ and we have verified that the best results are obtained with η in the range $[0.2, 0.3]$. The best results are obtained with η set to 0.24. Therefore this is the value used in our experiments.

Table 2
Percentage of successes obtained by the KA and ADL criteria for the stereo pairs representing sets SP1, SP2 and SP3 with η fixed to 0.24 and coefficient μ

	$SP1_{KA}$	$SP1_{ADL}$	$SP2_{KA}$	$SP2_{ADL}$	$SP3_{KA}$	$SP3_{ADL}(1)$	$SP3_{ADL}(3)$
Percentage of successes	54	74	62	83	58	76	94
μ	0.08	0.16	0.10	0.19	0.09	0.22	0.36

Obviously, the corresponding results for the KA criterion are not affected by the choice of η.

The second row in Table 2 shows the percentage of successes obtained with the ADL and the KA criteria. As we have not used known testing stereo image models, the evaluation of the matching results is carried out by a semi-automatic process. Each matched pair of edge segments is displayed over the corresponding pair of original stereo images. Then, a human expert verifies the quality of the match and updates the score of the true or the false matches accordingly. The percentage of successes is computed from the values in both scores.

The third row in Table 2 shows the values for the coefficient μ, which provides a decision margin when ambiguities arise. It is computed as follows:

(a) Without loss of generality, assume the following set of pairs of edge segments as an ambiguous case. The left edge segment l matches with n right edge segments, $r = 1, 2, \ldots, n$; with matching probabilities p_{lr}. Let a be one of the n right edge segments, so that $p_{la} = \max_{r=1,2,\ldots,n}\{p_{lr}\}$. As mentioned before, the match la is considered a correct match.

(b) Compute $m = \min\{|p_{la} - p_{lr}|\}$ for $r = 1, 2, \ldots, n$; $r \neq a$.

(c) For each ambiguous case j on each set of stereo image pairs SPh, where $h = 1, 2, 3$, compute m_j as in (b).

(d) For each SPh at each step 1, 2 or 3, compute the coefficient μ as follows:

$$\mu = \frac{\sum_{j=1}^{k} m_j}{k}, \tag{8}$$

where k is the number of ambiguous cases in SPh.

As the set of stereo images is processed two times in STEPs 1 and 3 (see Section 3.1), we denote this as $SP3_{ADL}(1)$ and $SP3_{ADL}(3)$, respectively.

From the results in Table 2, we see that μ increases with the learning. As μ provides a decision margin for ambiguities, the degree of confidence for taking decisions in ambiguous cases also increases with the learning.

Table 3 shows the percentage of successes for STEPs 1–3 obtained by our ADL approach and the learning methods PCF, LVQ and SHL. This comparison can be established because ADL, PCF, LVQ and SHL use the same test strategy, described in Section 3.1, with the same pairs of stereo images. As above, SP3(1) and SP3(3) mean results computed for the set SP3 of all its stereo image pairs in STEPs 1 and 3, respectively.

Although ADL, LVQ and SHL methods correspond to different local matching strategies, their results are similar. This allows us to validate ADL against the other two methods. It is very difficult to surpass the percentages obtained with the ADL method by using only the similarity constraint. It is possible to achieve better results with global matching strategies including similarity, uniqueness and smoothness constraints [see (Pajares et al., 1998a, 2000)]. More significant differences are obtained between ADL and PCF. As they both use the delta rule, the improvement of ADL with respect to PCF is found in the model of two combined ADALINEs in ADL

Table 3
Percentage of successes in STEPs 1–3 for ADL, PCF, LVQ and SHL

Percentage of successes	SP1	SP2	SP3(1)	SP3(3)
ADL	81	89	77	94
PCF	72	80	72	81
LVQ	79	86	78	92
SHL	80	88	78	94

against the unique one used by PCF. As mentioned in the introduction, good results in a local matching strategy contribute to good results in a global one, when the first is used for mapping the similarity constraint in the global matching strategy.

In LHW the most dominant attributes in order of importance are intensity, x derivative of intensity and gradient orientation. Gradient magnitude, Laplacian and curvature were rarely used. These results differ from those obtained with our ADL method. We have found an explanation for such differences in the different features because, as mentioned before, the gradient provides an effective measurement for edge segments (ADL), where the contrast in gray level is important, unlike the intensity that is basic in the LHW. The contrast is the gray-level difference between gray-level values of two pixels, each one at a different region of the two regions separated by an edge.

Analyzing all the results, the following two general conclusions are inferred:

1. The results obtained with our ADL local approach are at least similar to those resulting from other local strategies, so the ADL method is validated.
2. The learning process improves the matching results. This is supported by the following:
   - 2.1. As the training increases, the results are better.
   - 2.2. The decision margin increases with the training (i.e. better decisions are taken when the learning increases).
   - 2.3. Our learning ADL approach produces better results than the classical KA criterion. As we use the same four attributes, we conclude that the improvement is due to the learning of the weights.
   - 2.4. As shown in Tables 2 and 3, the decision margin and the results of the SP3 set during STEP 3 are better than those during STEP 1. As the SP3 set is the same in both steps, we conclude that this improvement is due to a greater degree of learning in STEP 3 than in STEP 1 (i.e. the system increases its knowledge of the environment in which it is moving).

4. Concluding remarks

The ADL stereo matching learning approach has been validated against other learning strategies. It improves the results as compared against classical similarity matching approaches. Also, the results improve with respect to the PCF. This validation and improvement is due to the design of the ADL learning strategy, in which the synaptic weight vector moves slightly from the unit vector as the training increases, i.e. as the system improves its knowledge of the environment. Such a behavior is not affected by the nature of the different objects or by the illumination conditions. The extrinsic factors do not contribute to the change in the synaptic weight vector. The following justifies this assertion. Since the stereo images are captured at a large number of spatial positions in the environment, if the synaptic weight vector moves away from the unit vector at a given position, this movement is compensated when the system captures another stereo pair at the same spatial position but rotated 180° with respect to the first position. At this rotated position the synaptic weight vector moves towards the unit vector. In contrast, the intrinsic factors contribute to the behavior of the synaptic weight vector. This is justified by the fact that the system is equipped with two physical cameras, always placed at the same relative position (left and right). Although the system is rotated, this relative position is always preserved when the difference measurement vector $\boldsymbol{x} = \boldsymbol{x}_l - \boldsymbol{x}_r$ is computed. Note that these measurement vectors are used for updating the synaptic weight vectors.

The mismatches could be eliminated by applying global matching constraints, which are out of the scope of this work (Pajares et al., 1998a, 2000).

Moreover, the range of distances of the different objects do not affect the final results.

Although we treat the special case of edge segments as matching features, we argue that the method introduced in this paper can be easily generalized to most features (i.e. pixels, curves, regions) where a set of attributes should be defined. At this moment, other types of environments are out of our interest, such as outdoor

scenes or aerial images, where probably edge segments are not appropriate features and therefore different features and attributes should be suitable.

Acknowledgements

Part of the work has been performed under project CICYT TAP94-0832-C02-01. The constructive recommendations provided by the reviewers are also gratefully acknowledged.

References

Ayache, N., 1991. Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception. MIT Press, Cambridge, MA.

Ayache, N., Faverjon, B., 1987. Efficient registration of stereo images by matching graph descriptions of edge segments. Internat. J. Comput. Vision 1, 107–131.

Breuel, T.M., 1996. Finding lines under bounded error. Pattern Recognition 29 (1), 167–178.

Cruz, J.M., Pajares, G., Aranda, J., 1995a. A neural network approach to the stereovision correspondence problem by unsupervised learning. Neural Networks 8 (5), 805–813.

Cruz, J.M., Pajares, G., Aranda, J., Vindel, J.F.V., 1995b. Stereo matching technique based on the perceptron criterion function. Pattern Recognition Lett. 16, 933–944.

Dhond, A.R., Aggarwal, J.K., 1989. Structure from stereo – a review. IEEE Trans. Systems Man Cybernet. 19, 1489–1510.

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis. Wiley, New York.

Fua, P., 1993. A parallel algorithm that produces dense depth maps and preserves image features. Machine Vision Appl. 6, 35–49.

Hilera, J.R., Martinez, V.J., 1995. Redes Neuronales Artificiales. RA-MA, Madrid.

Hoff, W., Ahuja, N., 1989. Surface from stereo: integrating feature matching, disparity estimation, and contour detection. IEEE Trans. Pattern Anal. Machine Intell. 11, 121–136.

Huertas, A., Medioni, G., 1986. Detection of intensity changes with subpixel accuracy using Laplacian–Gaussian masks. IEEE Trans. Pattern Anal. Machine Intell. 8 (5), 651–664.

Hush, D.R., Horne, B., 1992. An overview of neural networks, part I: static networks. Inf. Autom. 25 (1), 19–36.

Kahn, P., Kitchen, L., Riseman, E.M., 1990. A fast line finder for vision-guided robot navigation. IEEE Trans. Pattern Anal. Machine Intell. 12 (11), 1098–1102.

Kim, Y.C., Aggarwal, J.K., 1987. Positioning three-dimensional objects using stereo images. IEEE J. Robot. Autom. 3 (4), 361–373.

Kim, D.H, Choi, W.Y., Park, R.H., 1992. Stereo matching technique based on the theory of possibility. Pattern Recognition Lett. 13, 735–744.

Kohonen, T., 1989. Self-organization and Associative Memory. Springer, New York.

Kohonen, T., 1995. Self-organizing Maps. Springer, Berlin.

Krotkov, E.P., 1989. Active Computer Vision by Cooperative Focus and Stereo. Springer, Berlin.

Leu, J.G., Yau, H.L., 1991. Detecting the dislocations in metal crystals from microscopic images. Pattern Recognition 24 (1), 41–56.

Lew, M.S, Huang, T.S., Wong, K., 1994. Learning and feature selection in stereo matching. IEEE Trans. Pattern Anal. Machine Intell. 16 (9), 869–881.

Maravall, D., 1993. Reconocimiento de Formas y Visión Artificial. RA-MA, Madrid.

Mardia, K.V., 1972. Statistics of Directional Data. Academic Press, London.

Marr, D., Poggio, T., 1979. A computational theory of human stereo vision. Proc. Roy. Soc. London Ser. B 207, 301–328.

Medioni, G., Nevatia, R., 1985. Segment based stereo matching. Comput. Vision Graphics Image Process. 31, 2–18.

Mousavi, M.S., Schalkoff, R.J., 1994. ANN implementation of stereovision using a multi-layer feedback architecture. IEEE Trans. Systems Man Cybernet. 24 (8), 1220–1238.

Nevatia, R., Babu, K.R., 1980. Linear feature extraction and description. Comput. Vision Graphics Image Process. 13, 257–269.

Ozanian, T., 1995. Approaches for stereo matching – a review. Model. Ident. Control 16 (2), 65–94.

Pajares, G., 1995. Estrategia de Solución al Problema de la Correspondencia en Vision Estereoscópica por la Jerarquía Metodológica y la Integración de Criterios. Ph.D. Thesis, Dpto. Informática y Automática, Facultad Ciencias, UNED, Madrid.

Pajares, G., Cruz, J.M., Aranda, J., 1998a. Relaxation by Hopfield network in stereo image matching. Pattern Recognition 31 (5), 561–574.

Pajares, G., Cruz, J.M., Aranda, J., 1998b. Stereo matching based on the self-organizing feature-mapping algorithm. Pattern Recognition Lett. 19, 319–330.

Pajares, G., Cruz, J.M., López-Orozco, J.A., 1998c. Improving stereo vision matching through supervised learning. Pattern Anal. Appl. 1, 105–120.

Pajares, G., Cruz, J.M., López-Orozco, J.A., 1999. Stereo matching using Hebbian learning. IEEE Trans. Systems Man Cybernet. 29B (4), 553–559.

Pajares, G., Cruz, J.M., López-Orozco, J.A., 2000. Relaxation labeling in stereo image matching. Pattern Recognition 33 (1), 53–68.

Patterson, D.W., 1996. Artificial Neural Networks. Prentice-Hall, Singapore.

Pollard, S.B., Mayhew, J.E.W., Frisby, J.P., 1981. PMF: a stereo correspondence algorithm using a disparity gradient limit. Perception 14, 449–470.

Ruichek, Y., Postaire, J.G., 1996. A neural matching algorithm for 3-D reconstruction from stereo pairs of linear images. Pattern Recognition Lett. 17, 387–398.

Tanaka, S., Kak, A.C., 1990. A rule-based approach to binocular stereopsis. In: Jain, R.C., Jain, A.K. (Eds.), Analysis and Interpretation of Range Images. Springer, Berlin.

Trucco, E., Verri, A., 1998. Introductory Techniques for 3-D Computer Vision. Prentice-Hall, Upper Saddle River.

Wuescher, D.M., Boyer, K.L., 1991. Robust contour decomposition using a constraint curvature criterion. IEEE Trans. Pattern Anal. Machine Intell. 13 (1), 41–51.

Wu, J.K., 1994. Neural Networks and Simulation Methods. Marcel Dekker, New York.