ELSEVIER

Fuzzy Sets and Systems 110 (2000) 413–427

FUZZY sets and systems

www.elsevier.com/locate/fss

A new learning strategy for stereo matching derived from a fuzzy clustering method

Gonzalo Pajares, Jesús M. de la Cruz*

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

Received September 1996; received in revised form November 1997

Abstract

This paper presents an approach to the local stereo correspondence problem. The primitives or features used are groups of collinear connected edge points called segments. Each segment has several associated attributes or properties. We have verified that the differences of the attributes for the true matches cluster in a cloud around a center. Then for each current pair of primitives we compute a distance between the difference of its attributes and the cluster center. The correspondence is established in the basis of the minimum distance criterion (similarity constraint). We have designed an image understanding system to learn the best representative cluster center. For such purpose a new learning method is derived from the Fuzzy c-Means (FcM) algorithm where the dispersion of the true samples in the cluster is taken into account through the Mahalanobis distance. This is the main contribution of this paper. A better performance of the proposed local stereo-matching learning method is illustrated with a comparative analysis between classical local methods without learning.

Keywords: Cluster analysis; Decision making; Pattern recognition; Learning; Stereo matching; Similarity; Mahalanobis distance

1. Introduction

The number of research efforts of the computer vision community have been directed towards the study of the three-dimensional (3-D) structure of objects using machine analysis of images [10, 38]. Analysis of video images in stereo has become an important passive method for extracting the 3-D structure of a scene.

The basic principle involved in the recovery of depth using passive imaging is triangulation. In stereopsis, triangulation must be achieved with the help of existing environmental lighting alone. Hence, a correspondence needs to be established between features from two images that correspond to some physical feature in space. Then, if the position of the centers of projection, the effective focal length, the orientation of the optical axis and the sampling interval of each camera are known, the depth can be established using triangulation [14].

The process of stereo analysis consists of the following steps: image acquisition, camera modeling, feature acquisition, image matching and depth determination. The key step is that of image matching, namely, the process of identifying the corresponding points in two images that are cast by the same

* Corresponding author.

0165-0114/00/$ - see front matter
PII: S0165-0114(97)00382-5

physical point in 3-D space. 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 [34, 35]. Generally a former local matching process is associated with the similarity constraint where a minimum difference attribute (properties of features) criterion is applied, the results are later used by a global matching process where the smoothness constraint is imposed. A good choice of the local matching strategy is the key to good results in the global matching process.

This paper presents an approach to the local stereopsis correspondence problem developing a learning strategy (see [11, 16, 25, 32]) derived from a fuzzy clustering method (see [19, 20, 49, 54]). As the learning strategy is only concerned with the local stereo correspondence no global contraints (e.g., smoothness [34, 35], minimum differential disparity [36], coherence principle [43] or figural continuity [41, 44]) are applied, although they could improve the final results.

Two sorts ot techniques have been broadly used for stereo matching [24, 36, 37]; area-based and feature-based. 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 in the other image [2, 13, 46, 53], where the number of pairs of features to be considered becomes high, while feature-based methods use sets of pixels with similar attributes normally either pixels belonging to edges [12, 15, 21, 22, 30, 31, 33–35, 41, 44, 46] or the corresponding edges themselves [1, 8, 9, 17, 23, 36, 39, 40]. As shown in [38], these latter methods lead only to a sparse depth map, leaving the rest of the surface to be reconstructed by interpolation; but they are faster than area-based methods, because there are many fewer points (features) to be considered.

There are intrinsic and extrinsic factors affecting the stereovision matching system:

(a) *Extrinsic*, in a practical stereo vision system, the left and right images are obtained at different positions/angles. The matching is made difficult, in part, by changes in the images of corresponding points due to different perspective view points. The amount of change is dependent on the stereo angle.

(b) *Intrinsic*, the stereovision system is equipped with two different physical cameras (i.e. with different components), which are always placed at the same relative position (left and right). A systematic noise appears for each one.

Due to the above mentioned factors, the corresponding features in both images may display different values. A correspondence is established between features when such differences in feature values are assumed to be small, but the differences are sometimes too large to be considered, and matching is then rejected on this assumption. This may lead to incorrect matches. Thus, it is very important to find features in both images which are unique or independent of possible variation of the images [52]. Our experiment has been carried out in an artificial environment where the edge segments are abundant. Such features have been studied in terms of reliability [6, 24] and robustness [52] and as mentioned before, they have also been used in previous stereovision matching works. This fact justifies our choice of features, although they may be too local. Four average attribute values (module and direction gradient, variance and Laplace) are computed for each edge-segment as we will see later. Generally, the methods that use edge-segments as features compute an average gradient vector or its equivalent [1, 23, 27, 36]. We have added the Laplacian and variance because the stereo correspondence is safer.

The extrinsic factors have been broadly considered in the literature. This paper deals with both kinds of factors but it is mainly concerned with the intrinsic factors because we have verified their significance and, as a result, a research line has been opened including learning strategies. Hence, works [8, 9, 39] have been produced and both supervised and unsupervised learning methods implemented. Despite the fact that they show good results, they do not reach full performance for which reason new learning strategies are still being researched.

In stereovision matching we are only concerned with true matches, namely, pairs of features from left and right images that correspond to the same physical reality in the 3-D scene. In a hypothetical ideal system, the differences in attribute values should be null, but in a real system, and due to the above mentioned extrinsic and intrinsic factors, such differences are at least lightly off the null value (as we have verified in [8, 9, 39]) and this no null value is the goal to learn,

which is the center around the true correspondences should tend to cluster in one cloud.

All that really happens in such stereovision systems can be summarized as follows:

(a) physical cameras are placed at different positions and also they are built-in with different physical components, so they display different values in grey levels.

(b) The difference vectors for the true matches tend to cluster around a center.

Hence and according to the above conclusions, we will attempt to design and optimize our stereo matching system as per the following requirements:

(a) the features must be extracted and their attribute values computed;

(b) only true matches are of interest since the false ones do not contribute to the 3-D scene reconstruction;

(c) a clustering method is to be selected to deal with the true correspondences;

(d) a learning strategy must be implemented in order to detect the center around the true matches cluster;

(e) a similarity measure between such a center and a current pair of features must be computed and the current pair classified as a true or false match;

(f) the dispersion of the samples in the cluster must be taken into account.

The main contribution of this paper is concerned with the clustering method and the learning strategy (i.e. a pattern classification technique applied to stereovision matching). We have chosen a *Fuzzy c-Means* (FcM) algorithm because it is a member of a family of algorithms for clustering data points [19, 20, 49, 54], as we need cluster attribute difference vectors for true matches, and because with certain amount of manipulation it can be fitted to carry out a learning process. Only true matches are of interest to us, hence a unique cluster and a supervised learning setting is considered.

This paper is organized as follows: in Section 2 an image understanding system with three basic components (image analysis system, supervised learning system and stereo-matching system) is designed. Also the image analysis system is explained. In Section 3, the FcM algorithm is fitted to carry out the supervised learning strategy applied to stereovision matching, where only the cluster of true correspondences is taken into account. The current stereo matching system is explained in Section 4. To show the effectiveness of the proposed method, a comparative analysis is performed in Section 5 against a classical method where no learning is involved. We have developed other similar learning strategies applied to stereovision [8, 9, 39] with satisfactory global results, hence confirming the suggested method. Finally, in Section 6, the conclusion is presented.

2. The image understanding system

The image understanding system, with a parallel optical axis geometry, consists of three basic components. Following the diagram shown in Fig. 1, these components are [48]: an *image analysis system*, *a supervised learning system and a current stereo-matching system*. The function of the image analysis system is to extract information (features and attributes) from the scene and to make this information available to the supervised learning process during the training process or to the stereo-matching process during the current stereo image process. The image analysis system is also responsible for carrying out an initial selection of pairs of feature, supplying either of the other two systems with only those pairs that fulfill two conditions: the absolute value in the difference in the direction of the gradient is below a specific threshold and the overlap rate surpasses a certain value, that will be specified later. For the two involved segments the overlap rate is defined as the proportion expressed as a percentage between the intersection length and the longest length when one segment slides over another in a direction parallel to the epipolar line (note that no calibration is required).

The supervised learning process has to learn a mean difference measurement vector (cluster center) representative of all true correspondences and the dispersion of the true samples in the cluster (covariance matrix) when a number of labeled samples (pairs of true matches) from an environment are given to it and then processed.

The current stereo-matching system has to match the features from a new pair of stereo-images computing a distance between the current measurement difference vector for the new pair and the learned cluster center, so that the correspondences can be established.

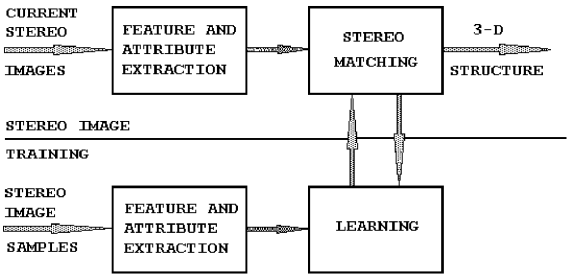

Fig. 1. The image understanding system.

As a result, the method here developed increases the number of correct matches as compared to the classical methods that, as a similarity measurement, use a minimum criterion distance, computing the minimum difference attribute value, generally through the Euclidean distance.

2.1. Feature and attribute extraction

As mentioned before, we believe that the feature-based stereo systems have strong advantages over area based correlation systems. However, detection of such boundaries is a complex time consuming scene analysis task. The contour edges in both images are extracted using the Laplacian of Gaussian filter in accordance with the zero-crossing criterion [18]. For each zero-crossing in a given image, its gradient vector, Laplacian and variance are computed from the gray levels of a central pixel and its eight immediate neighbors. The gradient vector (magnitude and direction) is computed as in [28], the Laplacian as in [29] and the variance as in [26]. To find the gradient magnitude of the central pixel, we compare the gray level differences from the four pairs of opposite pixels in the 8-neighbourhood and the largest difference is taken as the gradient magnitude. The gradient direction of the central pixel is the direction out of the eight principal directions whose opposite pixels yield the largest gray level difference and also points in the direction which the pixel gray level is increasing. We use a chain-code and we assign eight digits to represent the eight principal directions, such digits are integer numbers from 1 to 8. This approach allows the normalization of the gradient direction, so its values fall in the same range as the remainder of the attribute values. In order to avoid noise problems in edge-detection that can lead to later mismatches in realistic images, the following two global consistent methods are used: (a) the edges are obtained by joining adjacent zero-crossings following the algorithm of [47]; it is allowed a margin of deviation of $\pm 20\%$ in gradient magnitude and of $\pm 45^\circ$ in gradient direction; (b) each detected contour is approximated by a series of piecewise linear line segments as in [37]. Finally, for every segment, an average value of the four attributes is obtained from all computed values of its zero-crossings. All average attribute values are normalized in the same range. Each segment is identified with its initial and final pixel coordinates, its length and its label.

Therefore, each pair of features has associated two 4-dimensional vectors $\boldsymbol{x}_\mathrm{l}$ and $\boldsymbol{x}_\mathrm{r}$, where the components are the attribute values and the sub-indices l and r denote features belonging to the left and right images, respectively.

Thus a four-dimensional difference measurement vector $\boldsymbol{x}$ is also obtained from the above $\boldsymbol{x}_\mathrm{l}$ and $\boldsymbol{x}_\mathrm{r}$ vectors, where its components, $\boldsymbol{x} = \{x_\mathrm{m}, x_\mathrm{d}, x_\mathrm{l}, x_\mathrm{v}\}$, are now the corresponding original component differences for module and direction gradient, Laplacian and variance, respectively. Such an $\boldsymbol{x}$ vector will be the input for both the involved processes: *training and stereo matching*.

3. Learning derived from the Fuzzy c-Means algorithm

According to the above definitions, the difference measurement vector for an ideal true match should be the null vector as the corresponding attributes for each edge segment should be identical. If we consider the ideal true correspondences as a class, the null vector is its best representative mean attribute difference vector. Obviously, the above consideration is only feasible in an ideal world because the computed attributes for the corresponding features have a certain variability as the left and right images are obtained at different positions and both cameras are physically different, and thus even the corresponding features in both images may display different attribute values.

But, when we are working in the real world, the representative mean attribute difference vector for the true correspondences will differ from the ideal

null vector. The goal is to compute the best real representative mean attribute difference vector, named $\boldsymbol{z}$ (cluster center) for the true correspondences and a measure of dispersion of the samples around $\boldsymbol{z}$ (covariance matrix $\boldsymbol{C}$), so that during the *stereo matching* process a distance between $\boldsymbol{z}$ and the associated $\boldsymbol{x}$ difference measurement vector for a given pair of features can be established in order to decide if such pair is a true or false correspondence according to a minimum criterion distance. Really, all that we are trying to do is to **learn** the best representative vector for all true stereo pairs of features. Therefore a **learning** process is involved, where a number of labeled samples will be supplied in order to learn $\boldsymbol{z}$ and $\boldsymbol{C}$ during the corresponding training process.

The training process is designed as a supervised learning strategy using a **fuzzy c-means** algorithm [19, 20, 49, 54] as it includes both: (a) a clustering method and (b) a learning law; the first one allows us to consider the true correspondences as a cloud clustered around the cluster center vector $\boldsymbol{z}$ and the second one to learn this center.

In order to deal with the dispersion and consistency of the samples in the cluster a probability density function can be associated, without loss of generality a Gaussian one, with two parameters to be estimated: (a) the representative mean difference vector, that agrees with $\boldsymbol{z}$ and (b) the covariance matrix $\boldsymbol{C}$. Hence, the squared Mahalanobis distance, given by Eq. (1), is chosen as a metric [11, 32],

$$d_{\mathrm{M}} = (\boldsymbol{x} - \boldsymbol{z})^{\mathrm{t}} \boldsymbol{C}^{-1} (\boldsymbol{x} - \boldsymbol{z}), \tag{1}$$

where t stands for the transpose.

To develop our proposed method, we present a brief review of the *fuzzy clustering problem*. Following [19,20,49 or 54] a objective function of the form given by Eqs. (2) and (3) below is minimized

$$J_m(W, Z) = \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^m d_{ij}^2 \tag{2}$$

subject to

$$\begin{aligned} &\sum_{j=1}^{c} w_{ij} = 1, \quad 1 \leqslant i \leqslant n, \\ &w_{ij} \geqslant 0, \quad 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant c, \end{aligned} \tag{3}$$

where n is the number of samples to be clustered; c is the number of clusters, $1 < c < n$, m is a scalar, $m > 1$; $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{z}_j\|$ the Euclidean distance between samples $\boldsymbol{x}_i$ and the center $\boldsymbol{z}_j$ (other distances could be used [7, 43], however the Euclidean distance appears to be used more often), $X = \{x_1, x_2, \ldots, x_n\} \in R^s$ is a finite set of samples, $\boldsymbol{z}_j \in R^s$, $1 \leqslant j \leqslant c$ are the cluster centers. The w_{ij} factor is the membership grade of pattern i with cluster j, $W = \{w_{ij}\}$ an $n \times c$ matrix and $\boldsymbol{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_c]$ an $s \times c$ matrix. In order to minimize the objective function (2), the cluster center and membership grades are chosen so that high memberships occur for samples close to the corresponding cluster center. The number m is called the *exponent weight* [50, 51, 54]. The higher the value of m used, the less those samples whose memberships are low contribute to the objective function. Consequently, such samples tend to be ignored in determining the cluster centers and membership grades [51]. Bezdek [4] proposed the FcM algorithms to solve the above mathematical program which can also be found with an exhaustive treatement in [19, 20, 53, 54].

Our stereo-matching problem is similar to a fuzzy clustering problem, because we have two clusters associated with the true and false matches. Hence, the stereo matching problem could be solved by using the FcM algorithm. This algorithm computes the center for each cluster based on the membership grades of the samples with the cluster. As we are only concerned with the true matches, our method is exclusively restricted to the cluster associated with the true correspondences. This requires an amount of manipulation from the original FcM algorithm (explained below). The following is a statement of the proposed algorithm:

Initialization

Select a set of n true matches (labeled samples), $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in \boldsymbol{R}^4$.

Choose a scalar $\varepsilon > 0$ and also the initial center $\boldsymbol{z}$ and the covariance matrix $\boldsymbol{C}$; if training has not been performed previously set $\boldsymbol{z}$ to $\boldsymbol{0}$ and set $\boldsymbol{C}$ to the identity matrix ($\boldsymbol{I}$) otherwise set both data items to the last computed and stored values. Also set $k = 1$ and fix m.

Step 1: For each sample $\boldsymbol{x}_i$ compute at the iteration k the membership grade w_i of such sample with the

cluster of true matches as follows:

$$w_i(k) = \frac{1}{1 + [d_i(k)]^{2/(m-1)}}, \quad (4)$$

where $d_i(k)$ is the Mahalanobis distance between the sample $\boldsymbol{x}_i$ and $\boldsymbol{z}$. Therefore, w_i is considered a neighbourhood function through $d_i(k)$, $w_i \in (0, 1]$.

Step 2: Compute $\boldsymbol{z}(k+1)$ using the formula

$$\boldsymbol{z}(k+1) = \frac{\sum_{i=1}^{n} w_i^m(k)\boldsymbol{x}_i}{\sum_{i=1}^{n} w_i^m(k)}. \quad (5)$$

Step 3: If $\|\boldsymbol{z}(k) - \boldsymbol{z}(k+1)\| < \varepsilon$ or a number K of iterations has been reached, stop. Otherwise set $k = k+1$ and go to Step 1.

Step 4: From $i = 1$ to n compute the covariance matrix $\boldsymbol{C}$ according to Eq. (6)

$$C(i+1) = C(i) + w_i[(\boldsymbol{x}_i - \boldsymbol{z})^{\mathrm{t}}(\boldsymbol{x}_i - \boldsymbol{z}) - C(i)]. \quad (6)$$

End

At the end of Section 5 and after experimentation a discussion over the thresholds ε and K involved in the stopping condition is provided. Also, a cluster validity analysis, according to [50, 51], is performed to choose the exact value for the exponential weight m.

The differences between our algorithm and the original FcM algorithm are summarized as follows:

(a) *Initialization*: We set the cluster center to a known value, the FcM chooses cluster centers arbitrarily [20, 54].

(b) *Step* 1: The cluster associated to true matches is of unique interest to us. Hence, to compute the membership grades through Eq. (4), we take into account the Mahalanobis distance as a neighbourhood measure between the samples $\boldsymbol{x}_i$ and the cluster center $\boldsymbol{z}$. The original FcM considers several clusters of interest and computes each membership grade w_{ij} of pattern i with cluster j, (see Eqs. (2) and (3) for notation) as follows and according to [20]

$$w_{ij}(k) = \frac{1}{\sum_{l=1}^{c} (d_{ij}(k)/d_{il}(k))^{2/(m-1)}}. \quad (7)$$

(c) *Step* 2: We only compute the cluster center of interest through Eq. (5), and the original FcM computes several cluster centers according to Eq. (8), where the w_{ij} are taken into account,

$$\boldsymbol{z}_j(k+1) = \frac{\sum_{i=1}^{n} w_{ij}^m(k)\boldsymbol{x}_i}{\sum_{i=1}^{n} w_{ij}^m(k)} \quad \forall j. \quad (8)$$

(d) *Step* 3: The number of iterations is introduced as an additional stopping condition because the convergence of the original FcM algorithm is not completely guaranteed (see e.g. [3, 45]). This fact is supported by the experimental computed results and it is justified by a satisfactory effectiveness from the proposed application view point.

(e) *Step* 4: We introduce this additional step in our method in relation to the original FcM. The covariance matrix in Eq. (6) is estimated according to a maximum likelihood method (see [8, 9, 11, 39]). This is possible because it is assumed that an underlying probability density function is associated with the true matches in the cluster, and, without loss of generality, a Gaussian one. Eq. (6) is a learning law where the learning rate is replaced by the corresponding membership grades (neighbourhood function) following a self-organizing neural network criterion [16, 25]. So, the contribution of a sample with a high membership grade for updating $\boldsymbol{C}$ is very important and vice versa. The computation of $\boldsymbol{C}$ allows us to use the Mahalanobis distance (instead of the Euclidean distance), so that the dispersion of the samples in the cluster is taken into account.

4. The current stereo matching process

When a new and current pair of features is processed by the current stereo matching system with its associated $\boldsymbol{x}$ attribute difference vector, the squared Mahalanobis distance between $\boldsymbol{x}$ and the last learned center vector $\boldsymbol{z}$ is computed (where $\boldsymbol{C}$ is also the last computed covariance matrix). In that way, the dispersion of the samples in the cluster of true matches is once again taken into account. Hence, a minimum criterion distance is used to classify the current pair of features as a true or false correspondence.

Note that classical stereo matching techniques also use a minimum criterion distance where the Euclidean distance appears to be used more often, computing the corresponding distance between the $\boldsymbol{x}$ attribute difference vector and the null vector. This fact implies that no previous knowledge of the environment is taken into account in contrast to our proposed method where

a certain knowledge of the environment is obtained during a previous training process.

5. Experimental validation, comparative analysis and performance evaluation

In order to assess the validity and performance of the proposed method, experimental studies are designed after which, based on similarity criteria, a comparative analysis is performed with classical local stereo matching techniques [22, 36]. It will be seen that, when there is a learning process involved, the computed results are better than those in which no learning is considered, as could be expected. The images are 512×512 pixels in size with 256 gray levels.

5.1. Design of a test strategy

The objective is to prove the validity and generalization of the method by varying environmental conditions in two ways: by using new images with different features (different objects) and by changing the illumination. With this aim in mind, 8 pairs of stereo-images captured with natural illumination, are used as initial samples. Figs. 2–4 show three representative left images. Furthermore, three sets of stereo-pairs, which are different from each other and from the samples, are used and will constitute the inputs for the test: SP1, SP2 and SP3 with 6, 6 and 10 stereo-images, respectively, represented by the stereo-pairs given in Figs. 5–7. The set SP1 is captured with natural illumination (the same as the initial stereo-image samples) and the sets SP2 and SP3 with artificial illumination.

The test process tries to prove the effectiveness of learning as training increases. The test consists of four steps. An initial training is carried out (Step 0) with the initial samples. The three sets of stereo-pairs, SP1, SP2 and SP3 are then matched at different stages (Steps 1–3) and the true matches, selected by the system, are used in new and subsequent training processes. The set of stereo-pairs SP3 is matched twice (Steps 1 and 3) in order to prove how the degree of learning affects the corresponding results.

Step 0: The system is trained using the samples under the supervision of the human expert using the 8 pairs of initial stereo-images samples and setting

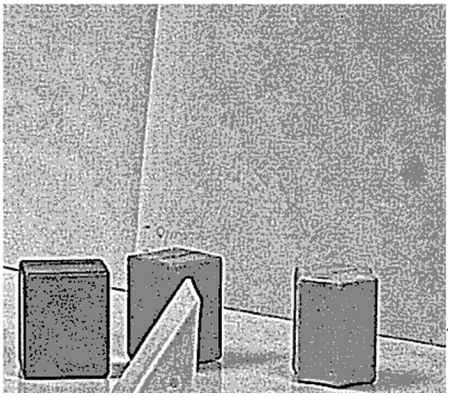

Fig. 2. Left original training image (blocks).

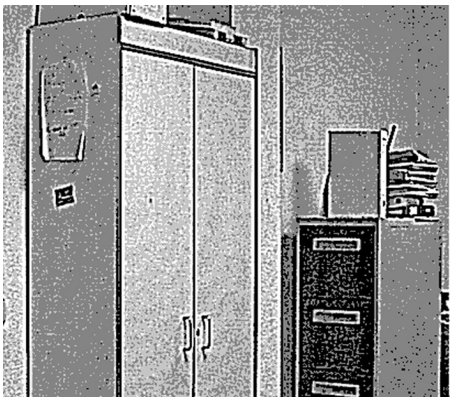

Fig. 3. Left original training image (furnitures).

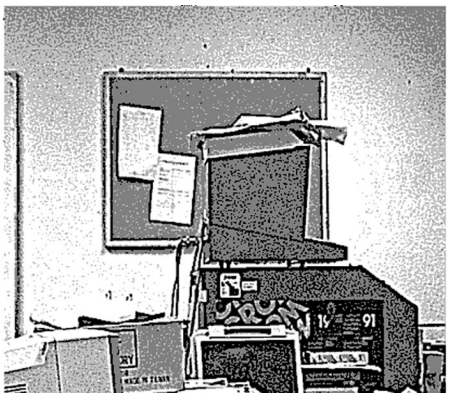

Fig. 4. Left original training image (computers).

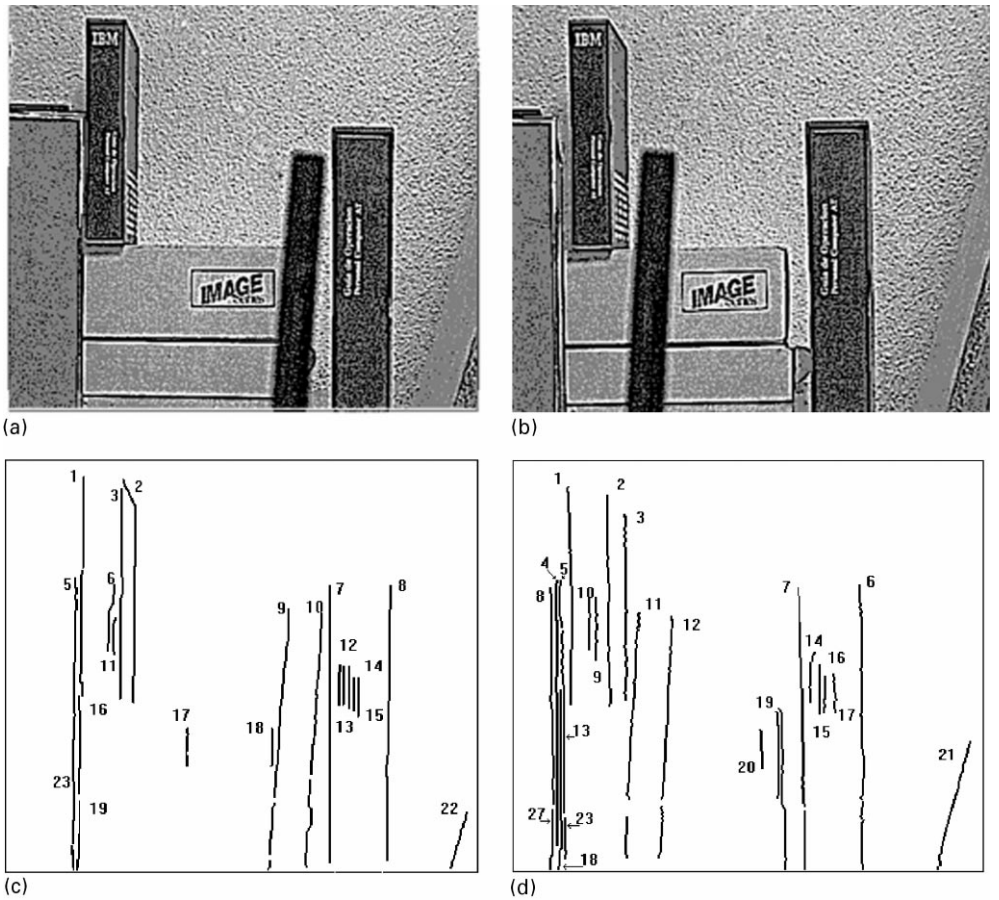

Fig. 5. (a) SP1: original left stereo image, (b) SP1: original right stereo image, (c) SP1: labeled segments left image and (d) SP1: labeled segments right image.

the parameters vector (z, C) to $(\mathbf{0}, I)$, after which the following updated parameter vector is computed:

$$z = \{0.226, -0.051, 0.461, 0.578\};$$

$$C = \begin{bmatrix} 0.983 & -0.017 & 0.012 & 0.009 \\ -0.017 & 0.955 & -0.010 & 0.003 \\ 0.012 & -0.010 & 1.147 & -0.014 \\ 0.009 & -0.003 & -0.014 & 1.199 \end{bmatrix}.$$

The changes in the covariance matrix C throughout the 4 steps are not statistically significant, in which case it suffices to give C for Step 0.

Step 1: The system processes two sets of stereo-pairs SP1 and SP3. Only the samples classified as true matches from set SP1 are used for a new training process because set SP3 will again be processed later in Step 3, so that no interference derived from its own processing arise. The updated mean representative vector is the following:

$$z = \{0.318, -0.093, 0.678, 0.697\}.$$

Step 2: The system processes the set SP2. The processing conditions are similar to those of set SP3 in Step 1, however, here the true matches are incorporated into the training process so that the new recalculated mean representative vector is,

$$z = \{0.405, -0.122, 0.698, 0.911\}.$$

Step 3: The system once again processes the set of stereo-pairs SP3. It is already partially familiar with

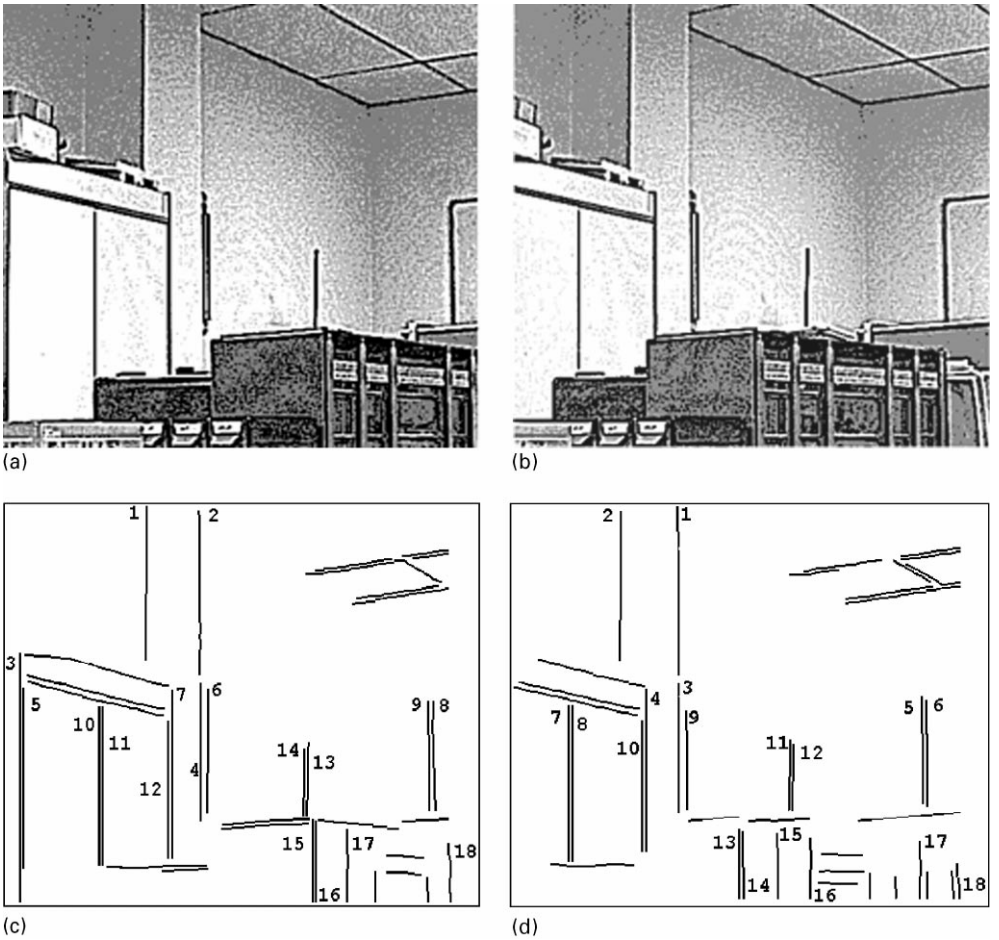

Fig. 6. (a) SP2: original left stereo image, (b) SP2: original right stereo image, (c) SP2: labeled segments left image and (d) SP2: labeled segments right image.

the environment of this set of stereo-pairs because of the incorporation of matches from stereo-pair SP2 with similar illumination. The mean representative vector is finally,

$$z = \{0.495, -0.143, 0.865, 0.988\}.$$

To compare the effectiveness of the methods, the Kim and Aggarwal [22] (KA) and Medioni and Nevatia [36] (MN) local matching techniques are selected. In KA, given two potential pixels for matching, a probability is computed through two weighting functions. One is based on the directional difference according to 16 fixed patterns, and the other is based on the difference in the gradients of gray-level intensity. That is to say, the method employs two attributes and computes two differences for each pair of pixels, with which the aforementioned probability is obtained.

$$p_{\mathrm{lr}} = \sum_{j=1}^{n} w_j \frac{1}{1 + |av_{\mathrm{l}j} - av_{\mathrm{r}j}|},$$

where

$$\sum_{j=1}^{n} w_j = 1; \quad 0 \leqslant w_j \leqslant 1, \tag{9}$$

where av denotes the corresponding attribute value for each feature, w_j is the associated weight for each attribute and n is the number of attributes used. In the KA method, n is 2 and the weights are fixed at

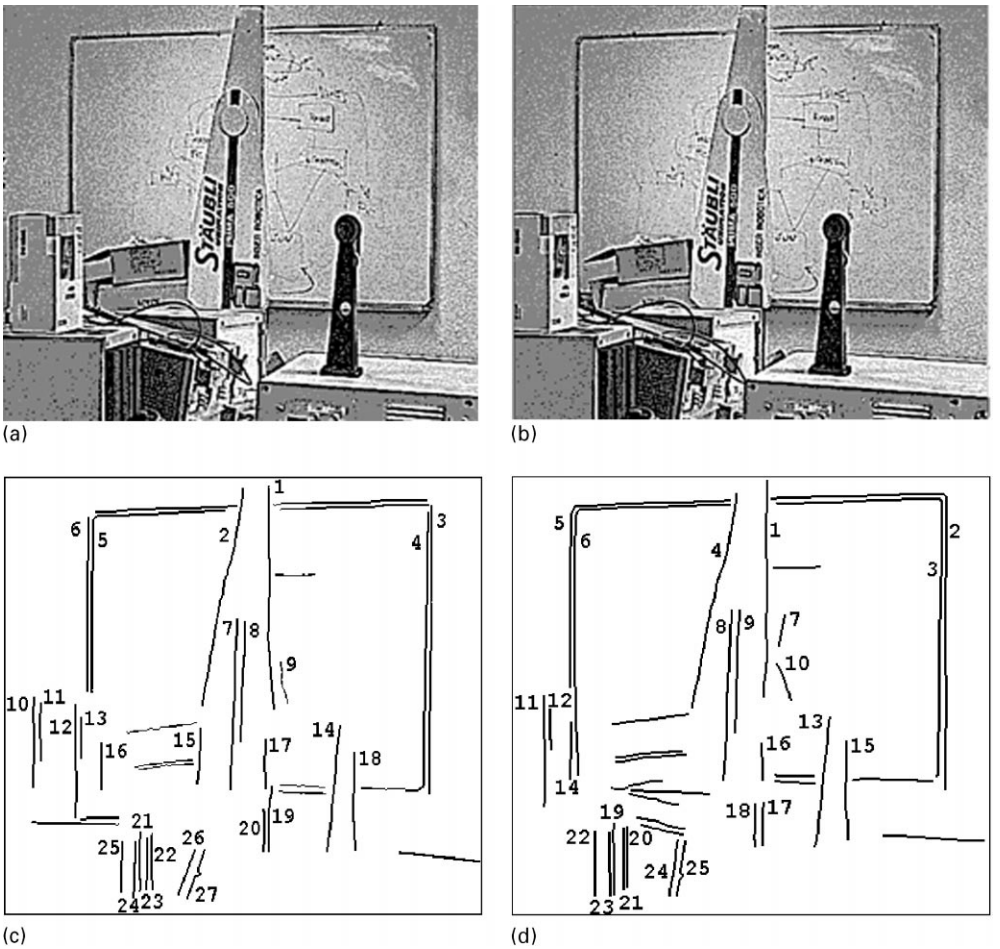

Fig. 7. (a) SP3: original left stereo image, (b) SP3: original right stereo image, (c) SP3: labeled segments left image, and (d) SP3: labeled segments right image.

a value of 0.5. In the MN method, the local stereo-correspondence is established between edge segments by defining a boolean function indicating whether two segments are potential matches if they overlap and have similar contrast and orientation, where with an amount of arrangement these last two could both be identified with our magnitude and direction gradient differences.

The KA and MN methods measure differences between attribute values and for comparison purposes they can be replaced by the Euclidean distance, as it computes the same measurement. Therefore, the comparison can be established with the Mahalanobis distance proposed in our method.

5.2. Comparative analysis

Table 1 records all computed results for the stereo-pair representative of set SP1. There are four columns: the first one indicates the assigned order number (*on*) for each pair of features appearing in the second column (*pair*), where the $*$ symbol denotes a true match as tested under a human expert criterion; in the third column the Mahalanobis distance, $d_{\mathrm{M}}(\boldsymbol{x},\boldsymbol{z})$ is computed, as required by our proposed learning method (environment known); in the fourth column the Euclidean distance is computed, $d_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{0})$ representing to the herein called classical methods (KA, MN) without learning (environment unknown).

Of all the possible combinations of pairs of matches formed by segments of left and right images, only 39 of them are considered, as the remainder do not meet the initial restriction which states that the value of the difference in the direction of the gradient must be less than $\pm 45°$ and the overlap rate greater than 75%. These matches are directly classified as False by the system and designated as missing. The choice of such thresholds is supported by the following technical points of view:

(a) Each image plane has defined a local co-ordinate system where the x-axes coincide, the y-axes are parallel and the optical axes and epipolar lines are also parallel.

(b) We are using only near-vertical edges for matching (the near-horizontal edges do not contribute to the computation of the disparity, which is the next step in any stereovision-matching system). From the above geometric constraints, the direction for two edges representing the same three-dimensional edge should be identical and the y-axis co-ordinate for the end-points in both edges should also be identical.

(c) During pre-processing and edge-detection, noise or errors could appear. For this reason, the thresholds are introduced so that a margin of error is tolerated. The corresponding values were arrived at through experimentation.

Of the 39 pairs considered, there are unambiguous and ambiguous ones, depending on whether a given left image segment corresponds to one and only one, or several right segments, respectively. In any case, the decision about the correct match is made by choosing the result of smaller value for each one of the methods (in the unambiguous case, there is only one) as long as it does not surpass the previous fixed threshold, set to 10 in this paper.

According to results from Table 1, the following conclusions may be inferred:

(a) The proposed method, hereinafter referred to as L (Learning) displays a greater number of True matches ($*$) than the classical local techniques, KA or MN, hereinafter designated as NL (No Learning), coinciding with relative minimum distance values.

(b) The range of values for L is greater than that for NL, making for better decisions when ambiguities arise. Table 2(a) shows the average (μ) and the standard deviation (σ) for both methods L and NL, in the difference between the value of the true match, following the criteria of the human expert, and that closest to it. Therefore, a lower failure probability is obtained with L. This is a direct consequence of the fact that

Table 1
Matching results from the stereo-pair representative of the set SP1, *on*: order number for the 39 pairs of considered features; *pair*: pairs of labeled features (l, r) from left and right images, where $*$ symbol means a true match: $d_M(\boldsymbol{x}, \boldsymbol{z}), d_E(\boldsymbol{x}, \boldsymbol{0})$: computed results for the Mahalanobis distance (learning) and Euclidean distance (without learning), respectively

on	pair	$d_M(\boldsymbol{x}, \boldsymbol{z})$	$d_E(\boldsymbol{x}, \boldsymbol{0})$
1	*(1, 1)	12.02	9.81
2	(2, 2)	7.21	7.92
3	*(2, 3)	6.54	8.10
4	(2, 6)	10.70	18.14
5	(2, 8)	61.20	23.21
6	(2, 14)	9.86	15.17
7	*(3, 2)	4.31	8.64
8	(3, 3)	7.22	7.21
9	(3, 8)	79.82	60.17
10	(3, 9)	8.62	7.23
11	(3, 10)	8.91	7.61
12	(3, 14)	7.25	8.92
13	*(7, 7)	3.24	8.22
14	(7, 19)	6.39	8.32
15	(8, 2)	4.11	17.22
16	(8, 3)	5.35	16.32
17	*(8, 6)	1.26	8.51
18	(8, 8)	25.21	17.21
19	(8, 9)	5.22	10.12
20	(8, 12)	10.12	8.14
21	*(9, 11)	2.01	5.82
22	(10, 6)	4.10	9.05
23	*(10, 12)	2.61	9.15
24	(11, 2)	4.61	7.55
25	(11, 3)	5.92	7.41
26	*(11, 9)	1.11	7.36
27	(11, 10)	4.61	8.65
28	*(13, 14)	3.42	5.45
29	(13, 16)	5.21	5.32
30	(14, 14)	3.21	7.01
31	*(14, 16)	3.91	7.21
32	(15, 15)	5.62	4.90
33	*(15, 17)	2.45	2.89
34	(16, 6)	16.23	12.12
35	(16, 12)	19.22	11.10
36	(16, 20)	9.15	10.21
37	(17, 18)	21.13	15.18
38	*(18, 20)	3.15	4.11
39	*(23, 27)	1.21	7.15

Table 2
Results for stereo-pair representative of the set SP1: (a) average (μ) and standard deviation (σ) for L and NL of the difference between the value of the true match, following the criteria of the human expert, and that closest to it for ambiguities and (b) final decisions from stereo-matching process

(a)	L	NL
μ	−2.18	−0.16
σ	1.48	0.72

(b)	L	NL
Success	12	9
Failures	3	6

the training images and the stereo images of set SP1 are all captured with natural illumination.

(c) The Success and Failure results of about 15 decisions made are shown in Table 2(b).

(d) The results obtained for the stereo-pair representative of set SP2 in Step 2 are not shown explicitly, because they are similar to the results of the stereo-pair representative of set SP3 in the Step 1 (see Table 3), both under the same conditions of illumination (artificial). The system makes 19 decisions with 16 successes and 3 failures with the stereo-pair representative of set SP2.

(e) Table 3 shows results for the stereo-pair representative of set SP3 with the same symbols and criteria as those explained for Table 1, also computing the Mahalanobis and Euclidean distances; although two values, identified with the numbers 1 and 3, are obtained for the Mahalanobis distance according to the processing for such stereo-pair in Steps 1 and 3, respectively.

According to results from Table 3, the following conclusions may be inferred:

(f) Table 4(a) shows the average (μ) and the standard deviation (σ) with the same criteria as Table 2(a), for the case of ambiguity. The average (μ) for L1 and L3 (results from our proposed learning method in Steps 1 and 3) displays values smaller than for NL (classical method without learning) making for better decisions when ambiguities arise. In general, the L1 results are worse that those of L3 since a greater training has been performed. For computing the L1 results, no training samples with artificial illumination were used, while for computing the L3 results, 11.9% of training samples proceeding from images with artificial illumination (the same as the stereo-pair set SP3) were used.

Table 3
Matching results from stereo-pair representative of SP3; *on*: order number for the 35 pairs of features; *pair*: pairs of labeled features (l, r) from left and right images respectively, where $*$ symbol means a true match; $d_M(\boldsymbol{x}, \boldsymbol{z})$, $d_E(\boldsymbol{x}, \boldsymbol{0})$: computed results for the Mahalanobis distance (learning) and Euclidean distance (without learning), respectively, where (1) and (3) means results computed according with test strategies in Steps 1 and 3, respectively

on	pair	$d_E(\boldsymbol{x}, \boldsymbol{0})$	$d_{M1}(\boldsymbol{x}, \boldsymbol{z})$	$d_{M3}(\boldsymbol{x}, \boldsymbol{z})$
1	$*(1,1)$	2.50	2.10	1.87
2	$*(2,4)$	3.08	3.07	2.96
3	$*(3,2)$	1.58	1.42	1.10
4	$(3,6)$	3.40	3.38	3.50
5	$*(4,3)$	2.04	2.05	1.52
6	$(4,5)$	11.14	12.02	13.70
7	$(5,1)$	80.12	80.50	81.01
8	$(5,2)$	6.76	7.50	7.73
9	$*(5,6)$	3.32	3.29	3.25
10	$(6,3)$	12.70	12.91	13.00
11	$*(6,5)$	4.01	3.83	3.12
12	$*(7,8)$	2.30	2.33	2.15
13	$(8,9)$	38.26	38.20	38.20
14	$(10,3)$	13.47	13.48	13.55
15	$(11,2)$	79.63	79.40	79.38
16	$(11,6)$	80.06	80.91	84.03
17	$*(12,11)$	0.51	0.49	0.42
18	$(12,15)$	7.45	8.29	8.28
19	$*(13,12)$	28.55	27.64	25.02
20	$*(14,13)$	3.50	3.45	3.33
21	$*(17,16)$	8.86	8.63	8.52
22	$(18,11)$	6.11	6.10	6.07
23	$*(18,15)$	7.89	6.98	5.15
24	$*(21,19)$	4.32	4.37	3.72
25	$(22,19)$	2.06	2.19	2.20
26	$*(22,20)$	2.55	2.11	2.00
27	$*(23,21)$	4.08	4.21	3.92
28	$(23,23)$	12.65	13.41	13.55
29	$*(24,22)$	15.78	13.26	9.13
30	$(25,21)$	1.19	2.15	2.80
31	$*(25,23)$	2.03	1.81	1.72
32	$(26,21)$	3.32	4.15	5.47
33	$(26,23)$	5.44	5.20	4.52
34	$*(26,24)$	2.35	2.16	2.11
35	$*(27,25)$	3.91	3.66	3.02

(g) The Success and Failure results of about 22 decisions taken are shown in Table 4(b).

(h) Considering all tested stereo-pairs, we conclude that the average percentages of successes for the NL, L1 and L3 strategies are 77.6%, 86.7% and 95.8%, respectively.

Table 4
Results for stereo-pair representative of the set SP3: (a) average (μ) and standard deviation (σ) for L and NL of the difference between the value of the true match, following the criteria of the human expert, and that closest to it for ambiguities. Indices 1 and 3 refer to results in the respective steps. (b) Final decisions from stereo-matching process

(a)	L3	L1	NL
μ	−5.10	−4.37	−3.67
σ	4.39	4.25	4.31

(b)	L3	L1	NL
Success	21	19	17
Failures	1	3	5

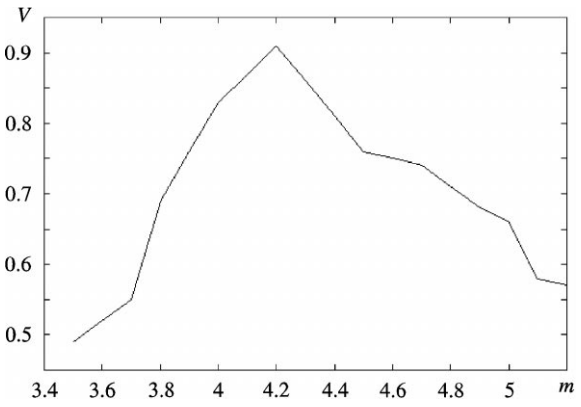

Fig. 8. Validity measure V, for the different values of m along the interval $[3.5, 5.2]$.

In the above experiments the required thresholds to control the stopping condition have been fixed as follows: $\varepsilon = 0.01$, $K = 10$, they were arrived at through experimentation and following a heuristic methodology. We have also tested a large range of values for the *exponential weight m* in the interval (1, 54], but we have verified that the best results, in terms of successes, are obtained exactly when m takes values inside of the interval $[3.5, 5.2]$. Indeed, taking values of m out of this interval the percentage of successes for the L1 and L3 strategies is always under 80% and 88% respectively, whereas the percentage of successes inside of the mentioned interval for the L1 and L3 strategies always surpasses these values. Then, to choose the exact value for m in the interval $[3.5, 5.2]$ we use a function which computes a number to measure the quality of the clustering associated with the true matches. A measure of this type is used in general by [50, 51] and it is called a *validity functional*. The quality of the clustering is indicated by how closely the samples $\boldsymbol{x}_i$ are associated to the cluster center $\boldsymbol{z}$, therefore, a measure of quality can be obtained from the membership grades which measure the level of association or classification (Eq. (4)). If the value of a membership grade for a given sample is large (low value for the Mahalanobis distance) then that sample is identified as being a true match (note that the true matches are represented by the cluster center). So, it is desirable to summarize the information contained in the memberships by a single number which indicates how well the sample is classified as a true match. This is done by defining the following validity expression,

$$V = \frac{1}{n}\sum_{i=1}^{n} \mu_i, \tag{10}$$

where n is the number of samples classified as true matches and μ_i is the corresponding membership value for the sample i. In our experiments n was 837 during the four steps of the test. From Eq. (10), the validity measure is the membership average value for all samples representing true matches. Fig. 8 shows the validity measure V, for the different values of m along the interval $[3.5, 5.2]$. We can see that the maximum value of V is obtained when m is 4.2. We choose this value for m in our experiments. We have also verified that this value is valid for the four steps in our test strategy.

6. Concluding remarks

We have verified that, due to the different physical cameras and that the left and right images are obtained at different angles, the features in both images may display different values. The true matches obtained by our stereo matching system cluster around a mean attribute difference center vector $\boldsymbol{z}$ in a cloud. This vector differs, at least lightly, from the ideal attribute difference vector $\boldsymbol{z} = \boldsymbol{0}$. In order to deal with the cluster of true matches and to discover which is its best representative mean difference vector $\boldsymbol{z}$, a learning strategy is derived from the FcM *algorithms* where the dispersion of the samples in the cluster is taken into account through the squared Mahalanobis distance. The above is the main contribution of this paper to the stereo matching methods.

When such a mean attribute difference vector $\boldsymbol{z}$ is learned and made available, the squared Mahalanobis distance between such vector and the associated $\boldsymbol{x}$ attribute difference vector for a current stereo pair of features, given as input, is computed. Once again the dispersion of the samples is considered through the squared Mahalanobis distance. This computed distance allows us to select the incoming current pair of features as a true or false correspondence according to a minimum criterion distance, so that if its value is smaller than a fixed threshold, set to 10 in this paper, the stereo pair is classified as a true correspondence and vice versa. If it is greater than such threshold, the stereo pair is classified as a false correspondence.

A comparative analysis is performed against classical stereo matching methods where no learning is considered. Computed values show that although the proposed local method of learning results in some incorrect matches, it is better than the classical local stereo matching methods where no learning is involved. We have also verified that the representative difference vector for the true matches moves away from the null vector as the training goes up (greater knowledge of the environment). Such behaviour is neither substantially affected by the nature of the different objects, nor by illumination conditions, and explains the better results when the knowledge of the environment increases. This fact leads us to consider that the intrinsic factors are decisive in the system behaviour.

The mismatches could be solved by considering global matching constraints or taking a maximum value of the accepted disparity. The latter has not been applied deliberately since we have objects in the stereo-images with a high disparity range (see Fig. 5(a) and (b)) that could violate such a constraint.

The method requires an exponential weight m and two parameters to control the stopping condition (ε and K). The parameter K is introduced because the convergence of the original FcM algorithm is not guaranteed. A decision-making threshold needs to be fixed for the distances involved. Although that is a limitation, we have verified that they can be fixed in a satisfactory manner after experimentation.

Finally, the method we propose has the natural limitations derived from the training process requiring a data base support and a certain number of training samples, where it is a difficult task to fix the ideal number and it is still undefined for our system.

Acknowledgements

The authors wish to acknowledge Prof. Dr. S. Dormido, Head of Department of Informática y Automática, CC Físicas, UNED, Madrid, for his support and encouragement. Part of this work has been performed under project CICYT TAP94-0832-C02-01. The constructive recommendations provided by the reviewers are also gratefully acknowledged.

References

[1] N. Ayache, B. Faverjon, Efficient registration of stereo images by matching graph descriptions of edge segments, Int. J. Computer Vision 1 (1987) 107–131.

[2] H.H. Baker, Building and using scene representations in image understanding, AGARD-LS-185 Machine Perception (1992) 3.1–3.11.

[3] J.C. Bezdek, R.J. Hathaway, M.J. Sabin, W.T. Tucker, Convergence theory for fuzzy c-means: counterexamples and repairs, IEEE Trans. System Man Cybernet. 17(5) (1987) 873–877.

[4] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

[5] A.W. Biermann, K.C. Gilbert, A.F. Fahmy, B. Koster, On the errors that learning machines will make, Int. J. Intelligent Systems 9 (1994) 269–302.

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

[7] D. Chaudhuri, C.A. Murthy, B.B. Chaudhuri, A modified metric to compute distance, Pattern Recognition 25 (7) (1992) 667–677.

[8] J.M. Cruz, G. Pajares, J. Aranda, A neural network approach to the stereovision correspondence problem by unsupervised learning, Neural Networks 8 (6) (1995) 805–813.

[9] J.M. Cruz, G. Pajares, J. Aranda, J.L.F. Vindel, Stereo matching technique based on the perceptron criterion function, Pattern Recognition Lett. 16 (1995) 933–944.

[10] A.R. Dhond, J.K. Aggarwal, Structure from stereo – a review, IEEE Trans. System Man Cybernet. 19 (1989) 1489–1510.

[11] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1989.

[12] E. Fernandez, Relaxation labeling algorithm based on learning automata, in: N. Perez de la Blanca, A. Sanfeliu, E. Vidal (Eds.), Pattern Recognition and Image Analysis, World Scientific, Singapore, 1992, pp. 3–20.

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

[14] K.S. Fu, R.C. González, C.S.G. Lee, Robótica: Control, detección, visión e inteligencia, McGraw-Hill, Madrid, 1988.
[15] W.E.L. Grimson, Computational experiments with a feature-based stereo algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 7 (1985) 17–34.
[16] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994.
[17] W. Hoff, N. Ahuja, Surface from stereo: integrating feature matching, disparity estimation, and contour detection, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 121–136.
[18] A. Huertas, G. Medioni, Detection of intensity changes with subpixel accuracy using Laplacian–Gaussian masks, IEEE Trans. Pattern Anal. Mach. Intell. 8 (5) (1986) 651–664.
[19] M.S. Kamel, S.Z. Selim, A relaxation approach to the fuzzy clustering problem, Fuzzy Sets and Systems 61 (1994) 177–188.
[20] M.S. Kamel, S.Z. Selim, New algorithms for solving the fuzzy clustering problem, Pattern Recognition 27 (3) (1994) 421–428.
[21] A. Khotanzad, A. Bokil, Y.W. Lee, Stereopsis by constraint learning feed-forward neural networks, IEEE Trans. Neural Networks 4 (1993) 332–342.
[22] D.H. Kim, J.K. Aggarwal, Positioning three-dimensional objects using stereo images, IEEE J. Robotics Automat. 3 (1987) 361–373.
[23] D.H. Kim, W.Y. Choi, R.H. Park, Stereo matching technique based on the theory of possibility, Pattern Recognition Lett. 13 (1993) 735–744.
[24] D.H. Kim, R.H. Park, Analysis of quantization error in line-based stereo matching, Pattern Recognition 7 (1994) 913–924.
[25] B. Kosko, Neural Networks and Fuzzy Systems, Prentice-Hall, Englewood Cliffs, NJ, 1992.
[26] E.P. Krotkov, Active Computer Vision by Cooperative Focus and Stereo, Springer, New York, 1989.
[27] S.H. Lee, J.J. Leou, A dynamic programming approach to line segment matching in stereo vision, Pattern Recognition 27 (1994) 961–986.
[28] J.G. Leu, H.L. Yau, Detecting the dislocations in metal crystals from microscopic images, Pattern Recognition 24 (1) (1991) 41–56.
[29] M.S. Lew, T.S. Huang, K. Wong, Learning and feature selection in stereo matching, IEEE Trans. Pattern Anal. Mach. Intell. 16 (9) (1994) 869–881.
[30] V.R. Lutsiv, T.A. Novikova, On the use of a neurocomputer for stereoimage processing, Pattern Recognition Image Anal. 2 (1992) 441–444.
[31] D. Maravall, E. Fernandez, Contribution to the matching problem in stereovision, Proc. 11th IAPR: Internat. Conf. on Pattern Recognition, 1992, pp. 411–414.
[32] D. Maravall, Reconocimiento de Formas y Vision Artificial, RA-MA, Madrid, 1993.
[33] D. Marr, Vision, Freeman, San Francisco, 1982.
[34] D. Marr, T. Poggio, A computational theory of human stereovision, Proc. Roy. Soc. London B 207 (1979) 301–328.
[35] D. Marr, T. Poggio, Cooperative computation of stereo disparity, Science 194 (1976) 283–287.
[36] G. Medioni, R. Nevatia, Segment based stereo matching, Comput. Vision Graphics Image Process. 31 (1985) 2–18.
[37] R. Nevatia, K.R. Babu, Linear feature extraction and description, Comput. Vision Graphics Image Process. 13 (1980) 257–269.
[38] T. Ozanian, Approaches for Stereo matching – a review, Modeling Identification Control 16 (2) (1995) 65–94.
[39] G. Pajares, Estrategia de Solucion al Problema de la Correspondencia en Vision Estereoscópica por la Jerarquía Metodológica y la Integración de Criterios, Thesis, Madrid Dpto. Informática y Automática, Facultad Ciencias, 1995.
[40] G. Pajares, R. Pereira, J. Rives, J.A. Diaz, Correspondencia Difusa en visión Estereoscópica, in: S. Barro, A. Sobrino (Eds.), III Congreso Español Sobre Tecnologías y Lógica Fuzzy, Santiago de Compostela, 1993, pp. 303–310.
[41] S.B. Pollard, J.E.W. Mayhew, J.P. Frisby, PMF: a stereo correspondence algorithm using a disparity gradient limit, Perception 14 (1981) 449–470.
[42] K. Prazdny, Detection of binocular disparities, Biol. Cybernet. 52 (1985) 93–99.
[43] F. Rhodes, On the metrics of Chauduri, Murthy and Chauduri, Pattern Recognition 28 (5) (1995) 745–752.
[44] P. Rubio, RP: un algoritmo eficiente para la búsqueda de correspondencias en visión estereoscópica, Informática y Automática 26 (1993) 5–15.
[45] S.Z. Selim, M.S. Kamel, On the mathematical and numerical properties of the Fuzzy c-Means Algorithm, Fuzzy Sets and Systems 49 (1992) 181–191.
[46] Y. Shirai, Three-Dimensional Computer Vision, Springer, Berlin, 1983.
[47] S. Tanaka, A.C. Kak, A rule-based approach to binocular stereopsis, in: R.C. Jain, A.K. Jain (Eds.), Analysis and Interpretation of Range Images, Springer, Berlin, 1990, pp. 33–139.
[48] E.C.K. Tsao, W.C. Lin, C.T. Chen, Constraint satisfaction neural networks for image recognition, Pattern Recognition 26 (1993) 553–567.
[49] H.F. Wang, C. Wang, G.Y. Wu, Bi-criteria Fuzzy c-Means analysis, Fuzzy Sets and Systems 64 (1994) 311–319.
[50] M.P. Windham, Cluster validity for fuzzy clustering algorithms, Fuzzy Sets and Systems 5 (1981) 177–185.
[51] M.P. Windham, Cluster validity for the Fuzzy c-Means clustering algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 4 (1982) 357–363.
[52] D.M. Wuescher, K.L. Boyer, Robust contour decomposition using a constraint curvature criterion, IEEE Trans. Pattern Anal. Machine Intell. 13 (1) (1991) 41–51.
[53] Y. Zhou, R. Chellappa, Artificial Neural Networks for Computer Vision, Springer, New York, 1992.
[54] H.J. Zimmermann, Fuzzy Set Theory and its Applications, Kluwer Academic Publishers, Norwell, 1991.