Expert Systems with Applications 38 (2011) 8622–8631

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Combining Support Vector Machines and simulated annealing for stereovision matching with fish eye lenses in forest environments

P. Javier Herrera [a], Gonzalo Pajares [b,*], María Guijarro [b], José J. Ruz [a], Jesús M. de la Cruz [a]

[a] Dpto. Arquitectura Computadores y Automática, Facultad de Informática, Universidad Complutense, 28040 Madrid, Spain
[b] Dpto. Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Universidad Complutense, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Fish-eye stereovision matching
Support Vector Machines
Simulated annealing
Texture classification
Omni-directional forest images

ABSTRACT

We present a novel strategy for computing disparity maps from omni-directional stereo images obtained with fish-eye lenses in forest environments. At a first segmentation stage, the method identifies textures of interest to be either matched or discarded. Two of them are identified by applying the powerful Support Vector Machines approach. At a second stage, a stereovision matching process is designed based on the application of four stereovision matching constraints: epipolarity, similarity, uniqueness and smoothness. The epipolarity guides the process. The similarity and uniqueness are mapped once again through the Support Vector Machines, but under a different way to the previous case; after this an initial disparity map is obtained. This map is later filtered by applying the Discrete Simulated Annealing framework where the smoothness constraint is conveniently mapped. The combination of the segmentation and stereovision matching approaches makes the main contribution. The method is compared against the usage of simple features and combined similarity matching strategies.

1. Introduction

1.1. Problem description

One important task in forests analysis is to determine distances from the stereo system to specific points on the trees. This is intended for forest inventories where densities, volumes, heights and other variables of the trees can be obtained. The increasing computer vision technologies are demanding solutions for making the above task automatic up to where it is possible. One of such technologies is concerned with a stereovision system. Because of the large areas to be processed in forest environments, a system based on fish-eye lenses allows imaging a large sector of the surrounding space with omni-directional vision. Fish eye optics systems can recover 3D information in a large field-of-view around the cameras; our system is $183^\circ \times 360^\circ$. This is an important advantage because it allows imaging the trees in the 3D scene close to the system from the base to the top, unlike in systems equipped with conventional lenses where close objects are partially mapped (Abraham & Förstner, 2005). This is the reason by which these systems are suitable for the proposed task.

According to Barnard and Fishler (1982) or Cochran and Medioni (1992), we can view the classical problem of stereo analysis as consisting of the following steps: image acquisition, camera modelling, feature extraction, image matching and depth determination. The key step is that of image matching. This is the process of identifying the corresponding points in two images that are cast by the same physical point in the 3-D space. This paper is devoted solely to the matching one, which is exactly where the major efforts have been applied by the computer vision community.

In our approach, the interest is focused on the trunks of the trees because they contain the higher concentration of wood and define univocally a tree i.e. the main inventory tasks are done based on them. These are our features of interest in which the matching process is focused. Fig. 1(a) displays a representative omni-directional image of the stereo pair (let's say the left one) captured with a fish-eye lens of the forest. As one can see there are three main groups of textures out of interest, such as grass in the soil, sky in the gaps and leaves of the trees. Hence, the first step consists on the identification of these textures out of interest to be excluded during the matching process. This is carried out through a segmentation process which uses both: (a) methods for texture analysis (Gonzalez & Woods, 2007) and (b) a classification approach based on the well-known Support Vector Machines (SVM) strategy applied to classification problems (Cherkassky & Mulier, 1998; Duda, Hart, & Stork, 2000; Vapnik, 2000). The first tries to isolate the leaves on the trees based on statistical measures and the second classifies the other two kinds of textures.

* Corresponding author. Tel.: +34 1 3947546; fax: +34 1 3947547.
E-mail addresses: pjherrera@pdi.ucm.es (P. Javier Herrera), pajares@fdi.ucm.es (G. Pajares), mguijarro@fdi.ucm.es (M. Guijarro), jjruz@dacya.ucm.es (J.J. Ruz), jmcruz@dacya.ucm.es (J.M. de la Cruz).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.01.066

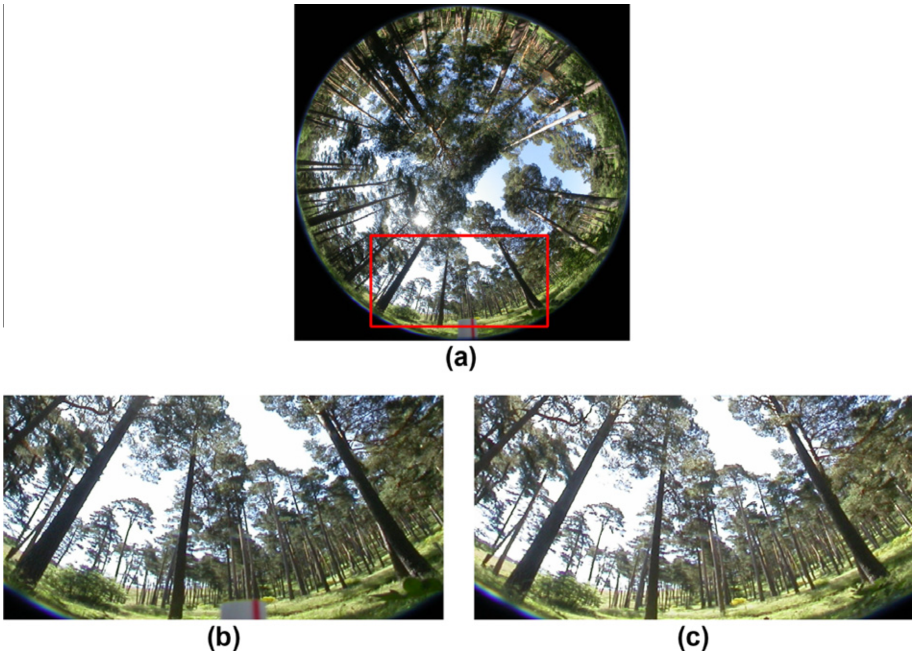

Fig. 1. (a) Omni-directional left image; (b) left expanded area; (c) corresponding right expanded area.

One might wonder why not to identify the textures belonging to the trunks. The response is simple. This kind of textures displays a high variability of tonalities depending on the orientation of the trunks with respect the sun, as detailed later in Section 2. Therefore, there is not a unique type of texture, depending on if the visible face of the trunk is illuminated or is on a shaded area, but even more there are trunks where shaded and illuminated patches appear in the same trunk, as we can see in Fig. 1(a).

Once the textures to be excluded have been identified, now the goal is to match trunks between the two images of the stereo pair. Fig. 1(b) displays the signed and expanded area on Fig. 1(a). This expansion is intended for making more explicit the details. In Fig. 1(c) the corresponding area in the right image of the stereo pair is displayed.

Based on the above reasoning, because of the irregular forms and distribution of the trunks, the most suitable features to be matched are pixels. With such purpose we exclude the pixels identified as belonging to one of the three kinds of textures out of interest mentioned above. The remaining pixels are the candidates to be matched, where those belonging to the trunks must be found.

Moreover, as the images are captured with two cameras, separated a certain distance (base-line), the tree's crowns are located at different positions with respect each camera and the incident rays of the sun produce important lighting variability between the pixels locations and surrounding areas in both images for the same structure in the scene; this adds a difficult to the matching process. This observation is valid for the whole set of images analyzed.

In stereovision matching there are a set of constraints that are generally applied for solving the matching problem, such as Barnard and Fishler (1982) and Cochran and Medioni (1992): epipolar, similarity, uniqueness or smoothness.

Epipolar: derived from the system geometry, given a pixel in one image its correspondence in the other image will be on the unique line where the 3D spatial points belonging to a special line are imaged. *Similarity*: matched pixels have similar attributes or properties. *Uniqueness*: a pixel in the left image must be matched to a unique pixel in the right one, except for occlusions. *Smoothness*: disparity values in a given neighbourhood change smoothly, except at a few discontinuities belonging to the edges, mainly in the borders of the trunks.

Also, two sorts of techniques have been broadly used for matching (Cochran & Medioni, 1992): area-based and feature-based. Area-based stereo techniques involve brightness (intensity) patterns in the local neighbourhood of a pixel in one image and the brightness patterns in the local neighbourhood of the corresponding pixel in the other image. Two kinds of approaches fall into this category. The first is concerned with the correlation coefficient and the second with statistical measures, generally used for identifying textures (Tang, Wu, & Chen, 2002). Feature-based methods (Lew, Huang, & Wong, 1994) compute some attributes for the pixels under correspondence; they can be simple attributes, such as the colour of the pixels or properties obtained by applying some operator such as the gradient (module and direction) or Laplacian. They were satisfactorily used in Lew et al. (1994), although some of them, such as the Laplacian, could become noise sensitive in some contexts. Really, these operators take into account the pixels and its neighbours; hence, from this point of view they could be considered as area-based. The colour is the unique attribute where the neighbourhood is not involved.

1.2. Motivational research

In Pajares and Cruz (2004) the combination of SVM and the optimization Deterministic Simulated Annealing (DSA) was exploited for stereovision matching with satisfactory results. This strategy was applied for images under perspective projection instead of fish-eye lenses and used edge segments as features because the scenarios were indoor environments. A network of nodes was built were each node was representing a pair of edge segments as potential matches. The states of the nodes are updated during the DSA process based on the mapping of the stereovision matching constraints. This is a global approach belonging to the category of methods that incorporate explicit smoothness assumption and determine all disparities simultaneously by applying an energy minimization process. Other methods, considered as global approaches, are those based on graph cuts (Bleyer & Gelautz, 2005b), belief propagation (Felzenszwalb & Huttenlocher, 2004) or Hopfield Neural Networks (Pajares, Cruz, & Aranda, 1998) among others.

As reported in Klaus, Sormann, and Karner (2006) some advances and good performances in stereovision matching have been obtained by applying consecutive processes under different layers (Bleyer & Gelautz, 2005a). First, regions of homogeneous colour are extracted. Second, a local matching method is used to determine disparities. Third, the disparities are refined by applying global matching strategies, i.e. after a first initial disparity map has been obtained. Hence, we can see that strategies applying filtering to the disparity map, previously computed, are suitable.

1.3. Contribution and organization of this paper

Our stereo images display two categories of textures. Textures of interest representing the trunks because they contain the major concentration of wood and textures out of interest (sky, grass in the soil and leaves of the trees).

Therefore, at a first stage we apply a segmentation strategy for identifying the textures out of interest which are to be discarded during the matching process. The trunks are textures with a degree of difficult to be identified, as detailed in Section 2, hence the pixels belonging to this kind of textures are the feature selected to be matched.

At a second stage, given a pixel in the left image, we apply the epipolar constraint for determining a list of candidates, which are potential matches, in the right image. Each candidate becomes an alternative for the pixel in the left image. For each pair of pixels, we apply the similarity constraint based on the six attributes mentioned above: (a) correlation coefficient (Tang et al., 2002), (b) variance as a measure of the texture (Pajares & Cruz, 2004; Pajares et al., 1998), (c) colour for each pixel (Klaus et al., 2006), (d) gradient magnitude (Bleyer & Gelautz, 2005a; Klaus et al., 2006; Lew et al., 1994), (e) gradient angle (Klaus et al., 2006; Lew et al., 1994) and (f) Laplacian (Bleyer & Gelautz, 2005a; Lew et al., 1994). The six attributes are used as components of a similarity vector which is used in the decision function of the SVM approach. Each one of the six attributes used separately allows determining a disparity map for comparison purposes. The final decision about the correct match, among the candidates in the list, is made according to the support that each candidate receives through the SVM mechanism. The disparity value at each pixel location is the absolute difference value in sexagesimal degrees between the angle for the pixel in the left image and the angle of its matched pixel in the right one. Each pixel is given in polar coordinates with respect the centre of the image. This mechanism is detailed in Section 3.2.

At a third stage the goal is to improve the disparity map up to where it is possible. Erroneous disparity values must be removed and the disparities associated to pixels belonging to the trunks must be smoothed. These two sub-goals, can be achieved by applying the stereovision smoothness constraint, where it considers not only the isolated disparity values at each pixel location but the pixels in the neighbourhood. For such purpose we have selected the DSA because it is an energy optimization approach that can avoid local minima. Indeed, according to Geman and Geman (1984) and reproduced in Haykin (1994), when the temperature involved in the simulated annealing process satisfies some constraints (explained in Section 3.4.2) the system converges to the minimum global energy which is controlled by the annealing scheduling.

The main contribution of this paper is the combination of a segmentation process for identifying three kinds of textures and a stereovision matching process, where the SVM paradigm classifies textures and also allows the mapping of the similarity and uniqueness constraints obtaining an initial disparity map. This map is later filtered for its improvement by applying the smoothness stereovision matching constraint through the DSA paradigm. The proposed approach is compared favourably against the usage of individual area-based and feature-based matching techniques.

This work is organized as follows. In Section 2 we describe the procedures applied for image segmentation oriented to the identification of textures. Section 3 is split in two parts; the first describes the design of the matching process by applying the epipolar, similarity and uniqueness constraints; including the overview of the fuzzy SVM paradigm. The second part describes the DSA paradigm and the method for applying the smoothness constraint. Section 4 displays the results obtained by using the proposed approach, and compares them with those obtained by considering the individual similarities and also by applying only the local SVM strategy. Section 5 presents the conclusions and future work.

2. Image segmentation

The images analyzed belong to different pinewoods, Fig. 1 displays a representative stereo image of the set of twenty stereo images analyzed in this work, see Section 4. As mentioned before, the goal of the image segmentation process is to exclude the pixels belonging to one of the three kinds of textures out of interest: sky, grass in the soil and leaves.

The exclusion of these textures is useful because the errors that they could introduce during the correspondence can be considerably reduced. This justifies the application of the proposed segmentation process.

Observing the textures we can see the following: (a) the areas occupied with leaves display high intensity variability in a pixel and the surrounding pixels in its neighbourhood; therefore methods based on detecting this behaviour could be suitable; (b) on the contrary, the sky displays homogeneous areas, where a pixel is surrounded with neighbouring pixels with similar intensity values, where the dominant spectral visible component is blue; (c) the grass in the soil also tend to fall on the category of homogeneous textures although with some variability coming from shades; in both, shading and sunny areas the pixels belonging to the grass have the green spectral component as the dominant one; (d) the textures coming from the trunks are the most difficult; indeed due to the sun position, the angle of the incident rays from the sun produce strong shades in the part of the trunks in the opposite projection position (west part in the image of Fig. 1(a)); the trunks receiving the direct projection display a high degree of illumination (east part in the image of Fig. 1(a)); there are some trunks containing merged patches of shaded and illuminated areas.

Based on the above, the identification of the trunks based on texture analysis is a difficult task, because they display irregular distributions and appearances depending on their position and orientation with respect the illumination coming from the sun. For identifying the textures coming from leaves, we use texture analysis techniques based on statistical measures that can cope with the high intensity variability. This is explained in Section 2.1. Because of the homogeneity of grass and sky textures we can use methods based on learning approaches as explained in Section 2.2. Finally, the textures coming from the trunks are not specifically identified during the segmentation phase and they are processed during the stereovision matching process, which is described in Section 3.

2.1. Identification of high contrasted textures

A variety of techniques have been used for texture identification (Trias-Sanz, Stamon, & Louchet, 2008). Most techniques rely on comparing values of what are known as second-order statistics (Gonzalez & Woods, 2007). These methods calculate measures of image texture such as the degree of contrast, coarseness, entropy or regularity; and also periodicity, directionality and randomness (Liu & Picard, 1996). Alternative methods of texture analysis for image retrieval include the use of Gabor filters localized in space and frequency, which can be used to retrieve frequential properties of a texture (Wan, Canagarajah, & Achim, 2007); wavelets which identify the textures based on the image decomposition on different sub-bands according to the orientation (Wang & Boesch, 2007); fractals used as measures of complexity for identifying repetitive patterns (Tao, Lam, & Tang, 2000); Fourier based for computing the orientation and spatial period for textures with at least two prominent directions (Lillo, Motta, & Storer, 2007).

The textures produced by the leaves of the trees under analysis do not display spatial distributions of frequencies nor textured patterns; they are rather high contrasted areas without any spatial orientation. Hence, we have verified that the most appropriate texture descriptors are those capturing the high contrast, i.e. statistical second-order moments.

One of the simplest is the variance. It is a measure of intensity contrast defined as follows (Gonzalez & Woods, 2007). Given a pixel i, we consider a window centred at i of size $n \times n$. Let z be a random variable denoting intensity levels in the window and let $p(z_j)$, $j = 0, 1, 2, \ldots, L-1$, be the corresponding histogram, where L is the number of distinct intensity levels. The variance is defined as,

$$\sigma^2(z) = \sum_{j=0}^{L-1} (z_j - \overline{m})^2 p(z_j) \tag{1}$$

where $\overline{m}$ is the average intensity value of z, i.e. the average intensity level computed as follows:

$$\overline{m} = \sum_{i=0}^{L-1} z_i p(z_i) \tag{2}$$

From (1) and according to Cochran and Medioni (1992), an intensity contrast coefficient, normalized in the range $[0, +1)$ can be defined as,

$$Z = 1 - \frac{1}{1 + \sigma^2(z)} \tag{3}$$

As one can see, Z is 0 for areas of constant intensity, where the variance is zero, and approaches +1 for large values of $\sigma^2(z)$, i.e. high contrasted areas. As we will see later, the original images used in our experiments are represented in the *RGB* (Red, Green and Blue) colour space, and for identifying the high textured areas, they are mapped to the *HSI* (Hue, Saturation and Intensity) space (Gonzalez & Woods, 2007); the intensity image I is used for computing the variance and the intensity values z_i ranging in $[0, 255]$ with $L = 256$. Finally, the criterion for identifying a high textured area is established by considering that it should have a value for the Z coefficient greater than a threshold T_1, set to 0.8 after experimentation in this paper. This value is established taking into account that only the areas with large values should be considered, otherwise a high number of pixels could be identified as belonging to these kinds of textures because the images coming from outdoor environments, in our case forests, display many areas with different levels of contrast.

2.2. Support Vector Machines: identifying relevant smooth textures

As mentioned before, in our approach there are two relevant textures that must be identified. They are specifically the sky and the grass. A pixel belonging to one of such textures displays a low value for Z because of its homogeneity. This is a previous criterion for identifying such areas, where the low concept is mapped assuming that Z should be less than the previous threshold T_1. Nevertheless, this does not suffice because there are other different areas which are not sky or grass fulfilling this criterion. Therefore, because we must to identify two classes, the SVM approach is a well texted classifier working appropriately in bi-class problems (Cherkassky & Mulier, 1998; Duda et al., 2000; Vapnik, 2000). The following two phases are involved in the SVM process: training and decision.

2.2.1. Training phase

We start with the observation of a set X of n training patterns, i.e. $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in \Re^d$. Each sample is to be assigned to a given cluster c_j, where the number of possible clusters is c, i.e. $j = 1, 2, \ldots, c$. In our approach the number of clusters is two corresponding to grass and sky textures, i.e. $c = 2$. For simplicity, in this paper, we identify the cluster c_1 with the sky and the cluster c_2 with the grass. The $\boldsymbol{x}_i$ patterns represent pixels in the *RGB* colour space. Their components are the R, G, B spectral values. This means, that the dimension of the space $\Re$ is $d = 3$. Once we have identified the samples as belonging to one of the two clusters, the set X is split into two subsets $X^1 = \left\{\boldsymbol{x}_1^1, \boldsymbol{x}_2^1, \ldots, \boldsymbol{x}_{n_1}^1\right\}$ and $X^2 = \{\boldsymbol{x}_1^2, \boldsymbol{x}_2^2, \ldots, \boldsymbol{x}_{n_2}^2\}$, where X^1, X^2 are the samples belonging to c_1 and c_2, respectively; with $X = X^1 \cup X^2$; the number of samples belonging to each subset is n_1 and n_2, respectively, i.e. $n = n_1 + n_2$.

2.2.1.1. Support Vector Machines (SVM).

The goal of the SVM approach is to estimate a decision function as follows (Cherkassky & Mulier, 1998; Duda et al., 2000; Vapnik, 2000),

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} \alpha_j y_j H(\boldsymbol{x}_j, \boldsymbol{x}) \tag{4}$$

H is chosen as the Radial Basis kernel given by: $H(\boldsymbol{x}, \boldsymbol{y}) = \exp\{-\|\boldsymbol{x} - \boldsymbol{y}\|^2/\sigma^2\}$ with $\sigma^2 = 3.0$. Others kernels have been tested in our approach without apparent improvement.

The parameters α_j, $j = 1, \ldots, n$, in Eq. (4) are the solution for the following quadratic optimization problem: maximize the functional

$$Q(\alpha) = \sum_{j=1}^{n} \alpha_j - \frac{1}{2} \sum_{j,k=1}^{n} \alpha_j \alpha_k y_j y_k H(\boldsymbol{x}_j, \boldsymbol{x}_k) \tag{5}$$

$$\text{subject to}: \sum_{j=1}^{n} y_j \alpha_j = 0; \quad 0 \leqslant \alpha_j \leqslant C/n, \quad j = 1, \ldots, n \tag{6}$$

Avoiding the superscripts, for simplicity, in the data points: $x_j, x_k \in X$. If $x_j \in X^1$ then $y_j = +1$ otherwise $y_j = -1$. This is applicable for each member in X. C is a regularization parameter set to 2000 as suggested in Cherkassky and Mulier (1998).

The data points $\boldsymbol{x}_i$ associated with the nonzero α_i are called *support vectors*. Once the support vectors have been determined, the SVM decision function has the form

$$f(\boldsymbol{x}) = \sum_{\text{support vectors}} \alpha_j y_j H(\boldsymbol{x}_j, \boldsymbol{x}) \tag{7}$$

2.2.2. Decision phase

After the training phase, a new unclassified sample $\boldsymbol{x}_s \in \Re^d$ must be classified as belonging to a cluster c_1 or c_2. Here, each sample, like the training samples, represents a pixel at the image with its R, G, B components.

Given, the attribute vector $\boldsymbol{x}_s$ for the sample i, it is possible to compute $f(\boldsymbol{x}_i)$ through (7), obtaining a scalar output value ranging in the interval $[-1, +1]$ whose magnitude can be interpreted as a measure of belief or certainty about its membership grade to the classes c_1, c_2. From the definition of y_j, if $\boldsymbol{x}_i \in X^1$ then $y_i = +1$ but if $\boldsymbol{x}_i \in X^2$ then $y_j = -1$. This means that the polarity of $f(\boldsymbol{x}_i)$ determines

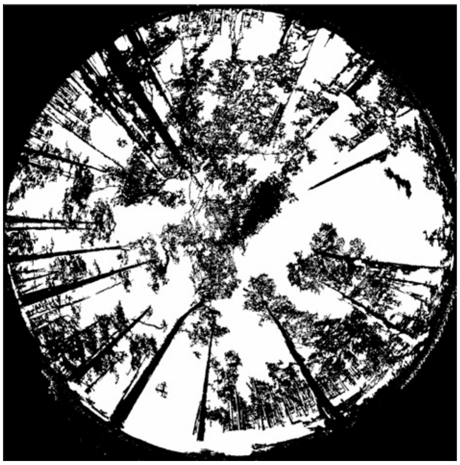

Fig. 2. Segmented image, where white areas are textures out of interest (sky, grass and leaves) and the black ones the pixels to be matched.

this membership degree, i.e. positive/negative values allows to assign $\boldsymbol{x}_i$ to c_1/c_2, respectively during this decision phase.

Fig. 2 displays the result of applying the segmentation process to the image in Fig. 1. The white areas are identified either as textures belonging to sky and grass or leaves of the trees. On the contrary, the black zones, inside the circle defining the image, are the pixels to be matched. As one can see the majority of the trunks are black, they really represent the pixels of interest to be matched through the corresponding matching process. There are white trunks representing trees very far from the sensor. They are not considered because are out of our interest, as explained in Section 4.

3. Stereovision matching process

Once the image segmentation process is finished, we have pixels identified as belonging to three types of textures which are to be discarded during the next stereovision matching process, because they are out of interest. Hence, we only apply the matching process to the pixels which remain unclassified.

As mentioned during the introduction, in stereovision there are several constraints that can be applied. In our approach we apply four of them: epipolar, similarity, uniqueness and smoothness. The epipolar allows restricting the search space for correspondence. The similarity and uniqueness, which are once again based on the SVM approach allows computing an initial disparity map, which is refined through the DSA approach based on the smoothness one. The three first ones constraints are addressed in Sections 3.1 and 3.2. This initial disparity map is described in Section 3.3. Finally, the smoothness constraint, mapped under the DSA is explained in Section 3.4.

3.1. Epipolar: system geometry

Fig. 3 displays the stereo vision system geometry (Abraham & Förstner, 2005; Schwalbe, 2005). The 3D object point P with world coordinates with respect to the systems (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) is imaged as (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}) in image-1 and image-2, respectively in coordinates of the image system; α_1 and α_2 are the angles of incidence of the rays from P; y_{12} is the baseline measuring the distance between the optical axes in both cameras along the y-axes; r is the distance between image point and optical axis; R is the image radius, identical in both images.

According to Schwalbe (2005), the following geometrical relations can be established,

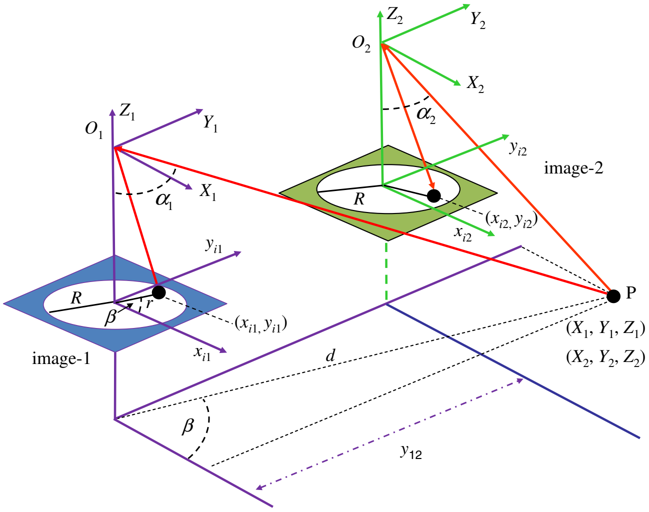

Fig. 3. Geometric projections and relations for the fish-eye based stereo vision system.

$$r = \sqrt{x_{i1}^2 + y_{i1}^2}; \quad \alpha_1 = (r\pi)/R; \quad \beta = tg^{-1}(y_{i1}/x_{i1}) \tag{8}$$

Now the problem is that the 3D world coordinates (X_1, Y_1, Z_1) are unknown. They can be estimated by varying the distance d as follows:

$$X_1 = d\cos\beta; \quad Y_1 = d\sin\beta; \quad Z_1 = \sqrt{X_1^2 + Y_1^2}/\tan\alpha_1 \tag{9}$$

From (7) we transform the world coordinates in the system $O_1X_1Y_1Z_1$ to the world coordinates in the system $O_2X_2Y_2Z_2$ taking into account the baseline as follows:

$$X_2 = X_1; \quad Y_2 = Y_1 + y_{12}; \quad Z_2 = Z_1 \tag{10}$$

Assuming no lenses radial distortion, we can find the imaged coordinates of the 3D point in image-2 as (Schwalbe, 2005),

$$x_{i2} = \frac{2R\arctan\left(\sqrt{X_2^2 + Y_2^2}/Z_2\right)}{\pi\sqrt{(Y_2/X_2)^2 + 1}}, \quad y_{i2} = \frac{2R\arctan\left(\sqrt{X_2^2 + Y_2^2}/Z_2\right)}{\pi\sqrt{(X_2/Y_2)^2 + 1}} \tag{11}$$

Using only a camera, we capture a unique image and the 3D points belonging to the line $\overline{O_1P}$, are all imaged on the unique point represented as (x_{i1}, y_{i1}). So, the 3D coordinates with a unique camera cannot be obtained. When we try to match the imaged point (x_{i1}, y_{i1}) into the image-2 we follow the epipolar line, i.e. the projection of $\overline{O_1P}$ over the image-2. This is equivalent to vary the parameter d in the 3-D space. So, given the imaged point (x_{i1}, y_{i1}) in the image-1 (left) and following the epipolar line, we obtain a list of m potential corresponding candidates represented by (x_{i2}, y_{i2}) in the image-2 (right).

The best match is associated to a distance d for the 3D point in the scene, which is computed from the stereo vision system. Hence, for each d we obtain a specific (x_{i2}, y_{i2}), so that when it is matched with (x_{i1}, y_{i1}) d is the distance for the point P from the system. Our matching strategy identifies correspondences between two pixels (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}). Based on this correspondence we start from Eq. (8) giving values to the variable d until the values of (x_{i2}, y_{i2}) obtained through Eq. (11) are equal or as close as possible to the ones obtained by the stereovision matching process. So, we can compute distances from the system to the 3D point P, Fig. 3. So, we can compute distances from the system to points in the base and the top of a tree if they are identified in subsequent processes after the correspondence, with these distances and using the angles of projection α_1 obtained with Eq. (8) for these points, we can compute the height of the tree by applying the trigonometric rules such as the cosine theorem. The above reasoning is also applicable for computing other variables for forest inventory described during the introduction.

3.2. Similarity and uniqueness: attributes for area and feature-based

Each pixel l in the left image is characterized by its attributes, where each attribute is denoted as a_l. In the same way, each possible candidate i in the list of m candidates is described by identical attributes, a_i. So, we obtain two vectors $x_l, x_i \in \Re^d$, $d = 6$, the six components are the values describing each pixel (feature): (1) correlation; (2) texture; (3) colour (the three R, G and B spectral components); (4) gradient magnitude; (5) gradient direction and (6) Laplacian. Both first ones are catalogued as area-based, computed on a 3×3 neighbourhood around each pixel through the correlation coefficient (Tang et al., 2002) and standard deviation (Lew et al., 1994), respectively. The four remaining ones are considered as feature-based (Lew et al., 1994).

Gradient (magnitude and direction) and Laplacian are computed by applying the first (Sobel's operator) and second derivatives, respectively (Gonzalez & Woods, 2007), over the intensity image after its transformation from the *RGB* colour space to the *HSI* (hue, saturation, intensity) one.

A six-dimensional difference measurement vector $\boldsymbol{x}$ can be obtained from the above $\boldsymbol{x}_l$ and $\boldsymbol{x}_i$ vectors, $\boldsymbol{x} = \boldsymbol{x}_i - \boldsymbol{x}_j$.

The stereovision matching problem is viewed as a two classification problem, where a pair of pixels is classified as a true or false match (true and false classes). Now, following the same reasoning that the one in Subsection 2.2.1 but applied to the correspondence problem, we have available the two subsets X^1 and X^2 of true and false matches, respectively; if $\boldsymbol{x}_j \in X^1$ then $y_j = +1$ otherwise $y_j = -1$. We need a training phase where a decision function for matching is obtained through Eq. (7) with their associated support vectors. As before, after experimentation the best kernel was the Radial Basis. In our matching classification problem the $f(\boldsymbol{x})$ polarity, sign of $f(\boldsymbol{x})$, determines the class membership for a pair of pixels l and i with a difference vector $\boldsymbol{x}$. We interpret the magnitude of $f(\boldsymbol{x})$ as a measure of certainty during the decision phase for the matching between l and i. Given a pixel l and the list of m potential candidates i, we compute *similarity* measurements, ranging in $[-1, +1]$ through the warping function (12) that modifies the sigmoid function in Mousavi and Schalkoff (1994). In order to avoid severe bias in the distances for the training data, the parameter γ is estimated experimentally and set to 0.2 in our experiments

$$s_{li}(\boldsymbol{x}) = \frac{2}{1 + \exp(-\gamma f(\boldsymbol{x}))} - 1 \tag{12}$$

The final decision about the best match between l and the list of m candidates represented by i is made based on the maximum similarity value (*uniqueness* constraint), i.e. the pixel i is the correct match of l iff $s_{li}(\boldsymbol{x}) > s_{lk}(\boldsymbol{x})$ with $k \neq i$ and k is one of the remainder $m - 1$ candidates.

3.3. Disparity map computation

Taking as reference the left image of the stereo pair, for each pixel $l \equiv (x_l, y_l)$ in this image we have its corresponding match in the right one $i \equiv (x_i, y_i)$. Therefore, we know their corresponding locations in Cartesian coordinates, which are transformed to polar coordinates considering the centre of the image as the origin of the polar reference system; so both pixels l and i have polar angles θ_l and θ_i, respectively. We build a map with the same locations that the original left image, i.e. $q = M \times N$, where each location represents a pixel. Given the pixel location $l \equiv (x_l, y_l)$ it is loaded with the value $\Delta\theta_l = |\theta_l - \theta_i|$ which represents the disparity value for the pixel l, once it has been matched with its best candidate i. This process is carried out for all locations corresponding to unclassified pixels during the segmentation process, Section 2. We assign a null disparity value for those locations corresponding to pixels classified as belonging to sky, grass or leaves. The values in the disparity map range in the interval $[0, \theta_{\max}]$, where $\theta_{\max}$ is fixed to 6.0 in our approach because it is the maximum disparity value observed in all available stereo images. This is the initial disparity map which is used as input for the DSA approach.

3.4. Smoothness: Deterministic Simulated Annealing (DSA)

Once the disparity map is obtained according to the above process, we try its improvement based on the DSA paradigm. In Section 3.4.1 we give details about the topology of a DSA and its working process. In Section 3.4.2 we apply this paradigm for improving the incoming disparity map by applying the smoothness constraint.

3.4.1. Topology and basic concepts

An important issue addressed in neural computation for image applications is referred to how sensory elements in a scene perceive the objects, i.e. how the scene analysis problem is addressed. To deal with real-world scenes some criterion for grouping elements in the scene is required. In the work of Wang (2005) a list of major grouping principles is exhaustively studied. They are inspired in the Gestalt's principles (Koffka, 1935). In our approach we apply the following three principles: *proximity*, labelled pixels that lie close in space tend to group; *similarity*, labelled pixels with similar values tend to group; *connectedness*, labelled pixels that lie inside the same connected region tend to group. These principles allow defining a spatial neighbourhood. Now the problem is to build some structure that can cope with the above. Several approaches can be used; we have chosen the DSA because it is an optimization one based on energy minimization, i.e. the convergence can be controlled by the energy. In DSA a network of nodes is built and the above principles can be applied by considering the influences exerted by the nodes k in a neighbourhood over a node i and mapped as consistencies, as explained later.

From the disparity map available at this moment, we build a network of nodes, where the topology of this network is established by the spatial distribution of the disparity map. Each node in the network is located at the same position that the elements in the map, i.e. at the same position that the corresponding pixel in the left image with the associated disparity value. Hence, the number of nodes in the network is $q = M \times N$. The node i in the network is initialized with the disparity value obtained from the disparity map at the same location, i.e. $\Delta\theta_i$, but instead of using the range $[0, \theta_{\max}]$ we map linearly the disparity values for ranging in $[-1, +1]$; for simplicity $\Delta\theta_i$ is renamed as D_i.

The network states (activation levels) are the normalized disparity values associated to the nodes. Through the DSA, these network states are reinforced or punished iteratively based on the influences exerted by their neighbours. The goal is to smooth the disparity map based on more stable state values.

3.4.2. Applying the DSA

Suppose the network with the q nodes. The simulated annealing optimization problem is: modify the analogue values D_i so as to minimize the energy (Duda et al., 2000; Haykin, 1994),

$$E(t) = -\frac{1}{2}\sum_{i=1}^{q}\sum_{k=1}^{q} r_{ik}(t)D_i(t)D_k(t) \tag{13}$$

where $r_{ik}(t)$ is the symmetric weight interconnecting two nodes i and k in the network at the iteration t and can be positive or negative ranging in $[-1, +1]$; $D_k(t)$ is the state of the neighbouring node k. Each $r_{ik}(t)$ determines the influence that the node k exerts on i trying to modify the state $D_i(t)$. According to Duda et al. (2000) the self-feedback weights must be null (i.e. $r_{ii} = 0$). The DSA approach tries to achieve the most network stable configuration based on the energy minimization.

The term $r_{ik}(t)$ is a consistency coefficient which computes the consistency between the states of the nodes in a given neighbourhood, defined as the m-connected spatial region, $\boldsymbol{N}_i^m$, where m is set to 8 in this paper and allows the implementation of the proximity and connectedness Gestalt's principles (Koffka, 1935; Wang, 2005). This coefficient is computed at the iteration t as follows:

$$r_{ik}(t) = \begin{cases} 1 - |D_i(t) - D_k(t)| & k \in \boldsymbol{N}_i^m,\ i \neq k \\ 0 & k \notin \boldsymbol{N}_i^m,\ i = k \end{cases} \tag{14}$$

From (14) we can see that $r_{ik}(t)$ ranges in $[-1, +1]$. The influence exerted by the node k over the node i will be positive (reward) or negative (punishment). Hence, a positive data consistency will contribute towards the network stability. Table 1 shows the behaviour of the energy against consistencies and state values. As one can see, the energy decreases as the data and the state values are both simultaneously consistent (rows 1 and 4 in the left part of Table 1); otherwise under any inconsistency the energy increases.

The simulated annealing process was originally developed in Kirkpatrick, Gelatt, and Vecchi (1983) or Kirkpatrick (1984) under a stochastic approach. In this paper we have implemented the deterministic one described in Duda et al. (2000) and Hajek (1988) because, as reported in Duda et al. (2000), the stochastic is slow due to its discrete nature as compared to the analogue nature of the deterministic. Following the notation in Duda et al. (2000), let $u_i(t) = \sum_k r_{ik}(t)D_k(t)$ be the force exerted on node i by the other nodes $k \in N_i^m$ at the iteration t; then the new state $D_i(t+1)$ is obtained by adding the fraction $f(\cdot,\cdot)$ to the previous one as follows:

$$\begin{aligned} D_i(t+1) &= \frac{1}{2}[f(u_i(t), T(t)) + D_i(t)] \\ &= \frac{1}{2}[\tanh(u_i(t)/T(t)) + D_i(t)] \end{aligned} \tag{15}$$

where, as always, t represents the iteration index. The fraction $f(\cdot,\cdot)$ depends upon $u_i(t)$ and the temperature T at the iteration t.

Eq. (15) differs from the updating process in Duda et al. (2000) because we have added the term $D_i(t)$ to the fraction $f(\cdot,\cdot)$. This modification represents the contribution of the self-support from node i to its updating process. This implies that the updated value for each node i is obtained by taking into account its own previous state value and also the previous state values of its neighbours. The introduction of the self support tries to minimize the impact of an excessive neighbouring influence. Hence, the updating process tries to achieve a trade-off between its own influence and the influence exerted by the nodes k by averaging both values.

Table 1
Behaviour of the energy term against the consistency coefficient and state values.

$r_{ik}(t)$	$D_i(t)$	$D_k(t)$	$E(t)$	$r_{ik}(t)$	$D_i(t)$	$D_k(t)$	$E(t)$
+	+	+	–	–	+	+	+
+	+	–	+	–	+	–	+
+	–	+	+	–	–	+	+
+	–	–	–	–	–	–	+

One can see from Eq. (14) that if a node i is surrounded by nodes with similar state values, $r_{ik}(t)$ should be high. This implies that the $D_i(t)$ value should be reinforced through the Eq. (15) and the energy given by the Eq. (13) is minimum and vice versa. Moreover, at high T, the value of $f(\cdot,\cdot)$ is lower for a given value of the forces $u_i(t)$. Details about the behaviour of T are given in Duda et al. (2000). We have verified that the fraction $u_i(t)/T(t)$ must be small as compared to $D_i(t)$ in order to avoid that the updating is controlled only by $u_i(t)$. Under the above considerations and based on Kirkpatrick et al. (1983), Kirkpatrick (1984) or Geman and Geman (1984), the following annealing schedule suffices to obtain a global minimum: $T(t) = T_0/\log(t+1)$, with T_0 being a sufficiently high initial temperature. T_0 is computed as follows (Laarhoven & Aarts, 1989): (1) we select four pairs of stereo images containing the pixels to be matched and compute the energy in (13) for each pair of stereo images selected after the initialization of the networks; (2) we choose an initial temperature that permits about 80% of all transitions to be accepted (i.e. transitions that decrease the energy function), and the temperature value is changed until this percentage is achieved; (3) we compute the M transitions ΔE_k and we look for a value for T for which $\frac{1}{M}\sum_{k=1}^{M}\exp(-\Delta E_k/T) = 0.8$, after rejecting the higher order terms of the Taylor expansion of the exponential, $T = 8\langle \Delta E_k \rangle$, where $\langle\cdot\rangle$ is the mean value. In our experiments, we have obtained $\langle \Delta E_k \rangle = 1.51$, giving $T_0 = 12.08$ (with a similar order of magnitude as that reported in Geman and Geman (1984). We have also verified that a value of $t_{max} = 20$ suffices, although the expected condition $T(t) = 0$, $t \to +\infty$ in the original algorithm is not fully fulfilled. The assertion that it suffices is based on the fact that this limit was never reached in our experiments as shown later in the Section 3, hence this value does not affect the results. The DSA process is synthesized as follows Duda et al. (2000):

```
1. Initialization: load each node with $D_i(t = 0)$ according to the
   equation; set $\varepsilon = 0.01$ (constant to accelerate the
   convergence); $t_{max} = 20$. Define $nc$ as the number of nodes
   that change their state values at each iteration.
2. DSA process: $t = 0$
while $t < t_{max}$ or $nc \neq 0$
  $t = t + 1$; $nc = 0$;
  for each node $i$
    update $D_i(t)$ according to the Eq. (15) from Eq. (14)
    if $|D_i(t) - D_i(t-1)| > \varepsilon$
    then
      $nc = nc + 1$; else $nc = nc$
    end if
  end for
end while
3. Outputs: the states $D_i(t)$ for all nodes updated.
```

4. Results

The system geometry is based on the scheme of the Fig. 3, with a baseline of 1 m. The cameras are both equipped with a Nikon FC-E8 fisheye lens, with an angle of 183° × 360°, as mentioned in the introduction. The valid colour images in the circle contain 6,586,205 pixels.

The tests have been carried out with twenty pairs of stereo images. We use four pairs of them for the training phase involved in the SVM approach (Section 2.2.1); in both, for textures classification (Section 2.2.2) and matching (Section 3.2).

At a second stage, for the remainder sixteen stereo pairs we obtain the initial disparity map for each stereo pair by applying the SVM approach pixel by pixel (Section 3.3). Then, each initial disparity map is smoothed through the DSA method (Section 3.4).

The tests consist on the computation of the errors obtained in the disparity maps. For such purpose we have available the ground truth disparity maps for the trunks of each stereo pair, provided by the end users. Thus, for each pixel in a trunk we know its correct disparity value according to this expert knowledge; which allows us to compute the percentage of error. For each one of the sixteen pairs of stereo images used for testing, we compute the disparity error for the pixels belonging to the trunks and then average these errors among the sixteen pairs of stereo images. The protocol consists on the analysis of sample plots of the forest with radius ranging from 5 to 25 m, located in the forest stand at distances ranging from 100 to 1000 m from each other. With such purpose, the stereovision sensor is located at the centre of the plot. The images contain trees belonging to the sample plots and also trees out of the sample plots. Only the first ones are of interest. The centres of the plots are known 3D geographical positions previously obtained via GPS. Moreover, as mentioned during the introduction, the sensor is positioned under the identifiable geographical direction normally the left camera oriented towards the North and the right one toward the South and both with the base-line of 1 m. This allows that different measurements spaced in the time, probably years, are obtained under the same criteria. This allows comparing the values of the variables measured in different times and deriving annual increments.

4.1. Results for the training phase during the image segmentation process

Our strategy involves two training processes for estimating the decision function in Eq. (7). The first is concerned to the segmentation process during the classification of sky and grass textures. So, from the four pairs of stereo images available for this, we select manually the samples belonging two these two textures, obtaining a set of 2560 training samples belonging to both types of classes. As one can see from the image in Fig. 1, the grass texture displays several intensities values depending on if the pixels are in a shaded or a sunny area. Therefore, to avoid problems with the absolute values of the R, G, B spectral components, we normalize their values as percentages. So, given a sample $x = (R, G, B)$ it is normalized as $x = (R/U, G/U, B/U)$ with $U = R + G + B$. The number of support vectors found was 73. This represents an important decrease with respect the number of training samples.

Fig. 2 displays the results obtained after segmentation for the image in Fig. 1(a), where high contrasted areas are identified through the coefficient Z, Eq. (3). Sky and grass textures are identified through the decision function estimated in (7) during the decision phase, Section 2.2.2.

In summary, as one can see in Fig. 2, the white pixels have been identified as belonging to one of the three textures of interest, which are discarded during the matching process, making it easier.

The second training process is concerned with the mapping of the similarity constraint, Section 3.2, where also a decision function as given in Eq. (7) must be estimated. We also select manually 1100 pairs of pixels which are false matches and some others of true matches. We compute the difference measurement vector $\boldsymbol{x}$ for each pair according to the discussion in Section 3.2. Finally, the number of support vectors obtained was 43.

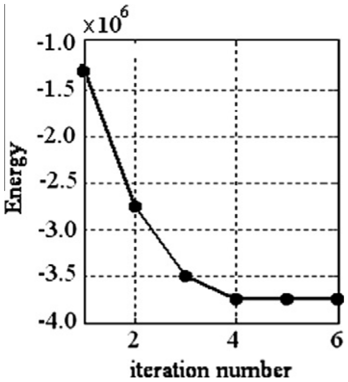

Fig. 5. Energy variation against the number of iterations during the DSA optimization process.

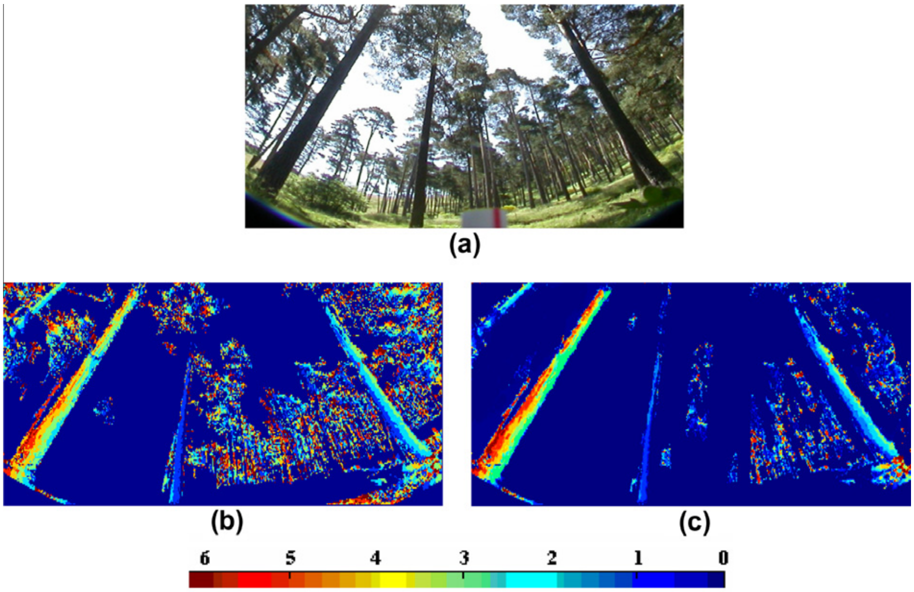

Fig. 4. (a) Expanded area corresponding to the signed area in the image of Fig. 1(a); (b) disparity map obtained by the SVM approach; (c) disparity map obtained by the DSA approach.

Table 2
Averaged percentage of errors and standard deviations obtained through maximum similarity criteria for each attribute separately and also for the SVM decision making approach and the DSA paradigm. Bold values remark the best performance of DSA..

s_a		s_b		s_c		s_d		s_e		s_f		SVM		DSA	
%	σ	%	σ	%	σ	%	σ	%	σ	%	σ	%	σ	%	σ
30	2.9	16	1.3	18	1.7	14	1.1	35	3.6	32	3.1	8	0.8	**5**	**0.5**

4.2. SVM and DSA matching performances

Given a stereo pair of the sixteen used for testing, for each pixel we obtain its disparity as follows. For facility, we reproduce in Fig. 4(a) the expanded area in Fig. 1(b).

Considering the six attributes separately, and building the unidimensional difference vectors $\boldsymbol{x}_l = \{a_l\}$ and $\boldsymbol{x}_i = \{a_i\}$, Section 3.2, and applying a maximum similarity criterion based on the absolute difference value for each attribute, i.e. $|a_l - a_i|$ among the m candidates, we obtain a disparity map derived from each attribute.

By applying the SVM approach, we compute the similarity between two pixels l and i through Eq. (12), obtaining the initial disparity map displayed in Fig. 4(b) for the area in Fig. 4(a). This initial map is filtered (smoothed) through the DSA procedure. After four iterations of the DSA we obtain the disparity map displayed in Fig. 4(c). The colour bar shows the disparity level values according to the colour for each disparity map. We have verified that more iterations do not improve the map. This is explained because as displayed in Fig. 5, the energy reaches a stable value at the iteration 4 and then remains stable for the other iterations. This is the general behaviour for the remainder stereo images. The average number of iterations for the sixteen stereo pairs is 3.5.

As one can see by observing the disparity map in Fig. 4(c), many isolated disparity values out and inside the trunks in Fig. 4(b) have been changed towards the values given by their neighbours. This leads to the desired smoothing in both the trunks and outside them. Another important observation comes from the main trunk in the left part of the expanded area; indeed, in the initial map, Fig. 4(b), the disparity values range from 1.5 to 5.5, but in the filtered map, Fig. 4(c), the low level values have been removed, now the disparities range from 3.5 to 5.5. Although there are still several disparity levels, this is correct because the trunk is very thick and it is placed near the sensor. This assertion is verified by the expert human criterion.

Table 2 displays the averaged percentage of errors and standard deviations based on the similarity for the six attributes when used separately, identified under the follows columns: ($s_a, s_b, s_c, s_d, s_e, s_f$). The averaged percentage of error obtained with the SVM and the DSA approaches are also displayed.

From results in Table 2 one can see that the strategies that SVM outperforms the individual similarity based approaches. This means that the combination of similarities between attributes improve the results obtained by using similarities separately.

The best individual results, according to the six attributes, are obtained through the similarities provided by the gradient magnitude attribute (s_d). This implies that it is the most relevant attribute.

Nevertheless, the main relevant results are obtained by the proposed DSA approach in terms of less percentage of error. This together with the qualitative improvement provided by this approach, as explained above, allow us to conclude that it is a suitable method for computing the disparity map in this kind of images.

We have verified that without the segmentation process the error for all strategies is increased about a quantity that represents on average about 21 percentual points. This means that the segmentation process is very important.

5. Concluding remarks

In this paper we have proposed a new strategy for obtaining a disparity map from omni-directional stereo images captured with fish-eye lenses. A first segmentation process identifies three types of textures, where the pixels classified as belonging to one of them are excluded for matching. This improves the final results. The stereovision matching process is based on the application of four stereovision matching constraints.

An initial disparity map is obtained by applying three of them (*epipolar*, *similarity* and *uniqueness*). For each pixel in the left image, a list of possible candidates in the right one is obtained for determining its correspondence. This is carried out through the SVM approach, which is a decision making strategy already used in previous stereovision matching approaches even though in a different environment (Pajares & Cruz, 2004).

The initial disparity map is improved by applying the *smoothness* stereovision matching constraint, inspired on the Gestalt's principles. This is carried out through the network built under the DSA paradigm, which can cope with the relations established between a pixel and its neighbours.

The proposed combined SVM strategy outperforms the methods that use similarities separately. The DSA outperforms the SVM, thanks to the optimization process. This means that it is a suitable strategy for filtering disparity maps.

Acknowledgements

The authors wish to acknowledge to the Council of Education of the Autonomous Community of Madrid and the Social European Fund for the research contract with the author P. Javier Herrera. Also to Drs. Fernando Montes and Isabel Cañellas from the Forest Research Centre (CIFOR, INIA) for his support and the imaged material supplied.

References

Abraham, S., & Förstner, W. (2005). Fish-eye-stereo calibration and epipolar rectification. *Photogrammetry and Remote Sensing, 59*, 278–288.

Barnard, S., & Fishler, M. (1982). Computational stereo. *ACM Computing Surveys, 14*, 553–572.

Bleyer, M., & Gelautz, M. (2005b). Graph-based surface reconstruction from stereo pairs using image segmentation. In *SPIE* (Vol. 5665, pp. 288–299).

Bleyer, M., & Gelautz, M. (2005a). A layered stereo matching algorithm using image segmentation and global visibility constraints. *ISPRS Journal of Photogrammetry and Remote Sensing, 59*(3), 128–150.

Cherkassky, V., & Mulier, F. (1998). *Learning from data: Concepts, theory and methods*. New York: Wiley.

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

Duda, R. O., Hart, P. E., & Stork, D. S. (2000). *Pattern classification*. Wiley.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient belief propagation for early vision. *International Journal of Computer Vision, 70*(1), 261–268.

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

Gonzalez, R. C., & Woods, R. E. (2007). *Digital image processing*. Prentice Hall.

Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematical Operation Research, 13*, 311–329.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York: Macmillan College Publishing Co..

Kirkpatrick, S. (1984). Optimization by simulated annealing: quantitative studies. *Journal of Statistical Physics, 34*, 975–984.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.

Klaus, A., Sormann, M., & Karner, K. (2006). Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Proceedings of the18th International Conference on Pattern Recognition (ICPR'06), Washington, USA* (pp. 15–18).

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt.

Laarhoven, P. M. J., & Aarts, E. H. L. (1989). *Simulated annealing: Theory and applications*. Holland: Kluwer Academic.

Lew, M. S., Huang, T. S., & Wong, K. (1994). Learning and feature selection in stereo matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 16*, 869–881.

Lillo, A., Motta, G., & Storer, J. A. (2007). Supervised segmentation based on texture signatures extracted in the Frequency domain. In J. Martí, J. M. Benedí, A. M. Mendoça, & J. Serrat (Eds.), *Pattern Recognition and Image Analysis. Lecture Notes in Computer Science* (Vol. 4477, pp. 89–96). Berlin: Springer-Verlag (Part I).

Liu, F., & Picard, R. W. (1996). Periodicity, directionality and randomness: Wold features for image modelling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(7), 722–733.

Mousavi, M. S., & Schalkoff, R. J. (1994). ANN implementation of stereo vision using a multi-layer feedback architecture. *IEEE Transactions on Systems, Man, and Cybernetics, 24*(8), 1220–1238.

Pajares, G., & Cruz, J. M. (2004). On combining support vector machines and simulated annealing in stereovision matching. *IEEE Transactions on System, Man, and Cybernetics, Part B, 34*(4), 1646–1657.

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

Schwalbe, E. (2005). Geometric modelling and calibration of fisheye lens camera systems. In *Proceedings of the 2nd panoramic photogrammetry workshop, international archives of photogrammetry and remote sensing* (Vol. 36, Part 5/W8).

Tang, L., Wu, C., & Chen, Z. (2002). Image dense matching based on region growth with adaptive window. *Pattern Recognition Letters, 23*, 1169–1178.

Tao, Y., Lam, E. C. M., & Tang, Y. Y. (2000). Extraction of fractal feature for pattern recognition. In *Proceedings of the international conference on pattern recognition, IAPR, Barcelona, Spain* (Vol. 2. pp. 527–530).

Trias-Sanz, R., Stamon, G., & Louchet, J. (2008). Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 63*, 156–168.

Vapnik, V. N. (2000). *The nature of statistical learning theory*. New York: Springer-Verlag.

Wan, T., Canagarajah, N., & Achim, A. (2007). Multiscale color-texture image segmentation with adaptive region merging. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP08)* (Vol. 1, pp. I-1213–I-1216).

Wang, D. (2005). The time dimension for scene analysis. *IEEE Transactions on Neural Networks, 16*(6), 1401–1426.

Wang, Z., & Boesch, R. (2007). Color- and texture-based image segmentation for improved forest delineation. *IEEE Transaction on Geoscience Remote Sensing, 45*(10), 3055–3062.