We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Segmentation and Stereoscopic Correspondence in Images Obtained with Omnidirectional Projection for Forest Environments

P. Javier Herrera, Gonzalo Pajares, María Guijarro,
José J. Ruz and Jesús M. de la Cruz
Complutense University of Madrid
Spain

1. Introduction

Stereoscopic vision systems have been used manually for decades to capture three-dimensional information of the environment in different applications. With the growth experienced in recent years by the techniques of computer image processing, stereoscopic vision has been increasingly incorporating automated systems of different nature. The central problem in the automation of a stereoscopic vision system is the determination of the correspondence between pixels of the pair of stereoscopic images that come from the same point in three-dimensional scene.

The research undertaken in this work comprises the design of a global strategy to solve the stereoscopic correspondence problem for a specific kind of hemispherical image from forest environments. The images are obtained through an optical system based on the lens known as fisheye because this optic system can recover 3D information in a large field-of-view around the camera; in our system it is $183^\circ \times 360^\circ$. This is an important advantage because it allows one to image 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).

The focus is on obtaining this information from tree trunks using stereoscopic images. The technicians carry out forest inventories which include studies on wood volume and tree density as well as the evolution and growth of the trees with the measurements obtained. Because the trees appear completely imaged, the stereoscopic system allows the calculation of distances from the device to significant points into the trees in the 3D scene, including diameters along the stem, heights and crown dimensions to be measured, as well as determining the position of the trees. These data may be used to obtain precise taper equations, leaf area or volume estimations (Montes et al., 2009). As the distance from the device to each tree can be calculated, the density of trees within a determined area can be also surveyed and growing stock; tree density, basal area (the section of stems at 1.30 m height in a hectare) and other interesting variables may be estimated at forest stand level using statistical inference (Gregoire, 1998).

This work stems from the interest generated by the Spanish Forest Research Centre (CIFOR) part of the National Institute for Agriculture and Food Research and Technology (INIA) to automate the process of extracting information through the measurement mechanism with patent number MU-200501738.

The main contribution of this chapter is the proposal of a strategy that combines the two essential processes involved in artificial stereo vision: segmentation and correspondence of certain structures in the dual images of the stereoscopic pair. The strategy is designed according the type of images used and lighting conditions from forest environments. These refers to Scots pine forests (*Pinus sylvestris* L.) where images were obtained on sunny days and therefore they exhibit highly variable intensity levels due to the illuminated areas. Due to the characteristics of this environment - in terms of light and the nature of trees themselves and textures that surround them - the segmentation and correspondence processes are specifically designed according to this type of forest environment. This sets the trend for future research when analyzing other forest environments. The segmentation process is approached from the point of view of isolating the trunks by excluding the textures that surround them (pine needles, ground and sky). For this reason, we propose the use of the specific techniques of texture identification for the pine needles (Pajares & Cruz, 2007) and of classification for the rest (Pajares et al. 2009; Guijarro et al. 2008, 2009). The correspondence problem can be defined in terms of finding pairs of true matches, as explained below, pixels in two images that are generated by the same physical element in the space. These true matches generally satisfy some constraints (Scharstein & Szeliski, 2002): 1) *epipolar*, given a pixel in an image, the matched pixel in the second image must lie following the called epipolar line; 2) *similarity*, matched pixels must have similar properties or attributes; 3) *ordering*, the relative position between two pixels in an image is preserved in the other image for the corresponding matches; 4) *uniqueness*, each pixel in one image should be matched to a unique pixel in the other image, although a pixel could not be matched because of occlusions. The proposed matching process identifies the homogeneous pixels in separate stereo pair images, by means of the combination of similarity measurements calculated from a set of attributes extracted of each pixel.

The proposed strategy based on segmentation and correspondence processes can be favourably compared from the perspective of the automation of the process and we suggest it can be applied to any type of forest environment, with the appropriate adaptations inherent to the segmentation and correspondence processes in accordance with the nature of the forest environment analyzed.

This chapter is organized as follows. In section 2 we describe the procedures applied for the image segmentation oriented to the identification of textures. Section 3 describes the design of the matching process by applying the epipolar, similarity and uniqueness constraints. Section 4 presents the conclusions and future work.

2. Segmentation

In our approach, the interest is focused on the trunks of the trees because they contain the higher concentration of wood. These are our features of interest in which the matching process is focused. Figure 1 displays a representative hemispherical stereo pair captured with a fisheye lens from the forest. As one can see, there are three main groups of textures without 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 the textures out the 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, 2008) and b) a classification approach based on the combination of two single classifiers, they are the parametric Bayesian estimator and the Parzen's window (Duda et al., 2001). The first tries to isolate the leaves based on statistical measures and the second classifies the other two kinds of textures. The performance of combined classifiers has been reported as a promising approach against individual classifiers (Kuncheva, 2004; Guijarro et al., 2008, 2009; Pajares et al., 2009; Herrera et al., 2011a).

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. Therefore, there is not a unique type of texture (dark or illuminated trunks and even though alternatively in bits), as we can see in Figure 1. Observing the textures we can also see the following: *a*) the areas covered 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 of 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 shaded 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 as we said above; 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 position of the projection (west part in the images of Figure 1); the trunks receiving the direct projection display a high degree of illumination (east part in the images of Figure 1); there are a lot of trunks where the shades produce different areas.

Based on the above, 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, described in section 3.
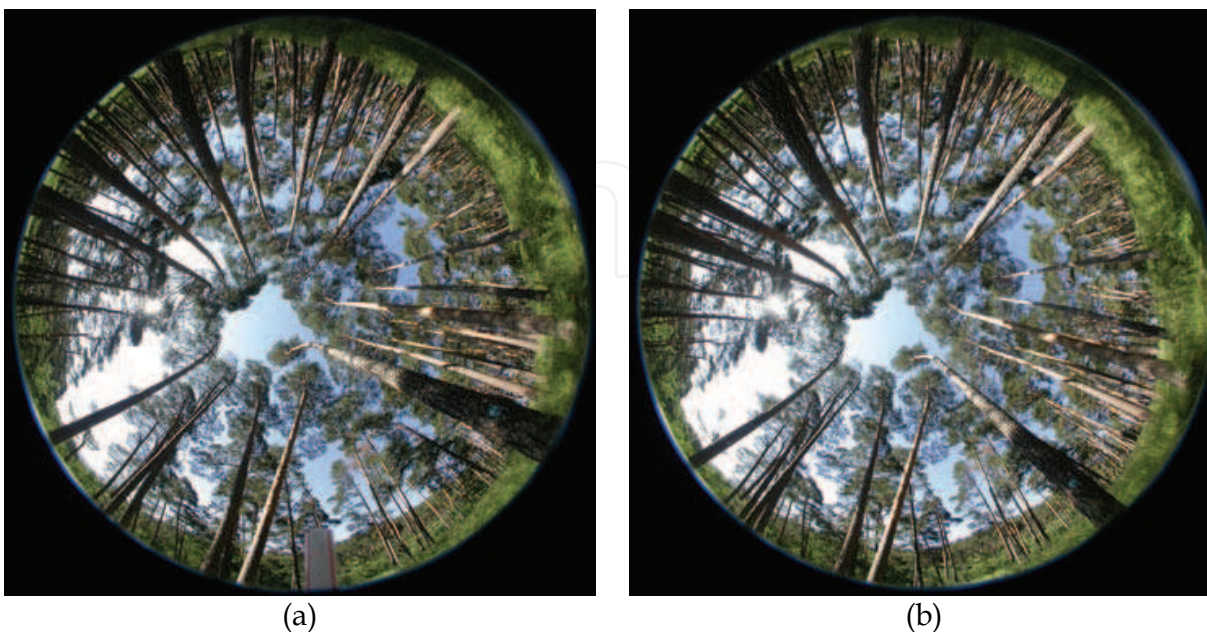

Fig. 1. A representative hemispherical stereo pair; (a) left image; (b) right image.

2.1 Identification of high contrasted textures

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 in our approach as in (Gonzalez & Woods, 2008; Herrera, 2010). The criterion for identifying a high textured area is established by considering that it should have a value for the intensity contrast coefficient R, normalized in the range [0, +1], greater than a threshold T_1, set to 0.8 in this chapter after experimentation. 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 (forests) display a lot of areas with different levels of contrast.

2.2 Identification of homogeneous textures: combining classifiers

Any classification process in general and in particular the identification of textures in natural images has associated two main phases: training and decision. We refer to the first phase as learning phase also, by identifying both concepts in the literature. By the nature of processing in the time sometimes appear as off-line and on-line processes respectively. This is due to the fact that the training phase is usually carried out during system downtime, being at this time when the parameters involved in the process are estimated or learned. However, the decision phase is performed for a fully operational system, using the parameters learned in the training phase.

Figure 2 shows an overview of a training-decision system particularized to the case of natural texture images. Both phases consist of both common and different processes. Indeed, the processes of image capture, segmentation and coding information are common, while learning and decision processes are different. We briefly describe each of them. Then in each method the appropriate differentiation is provided.
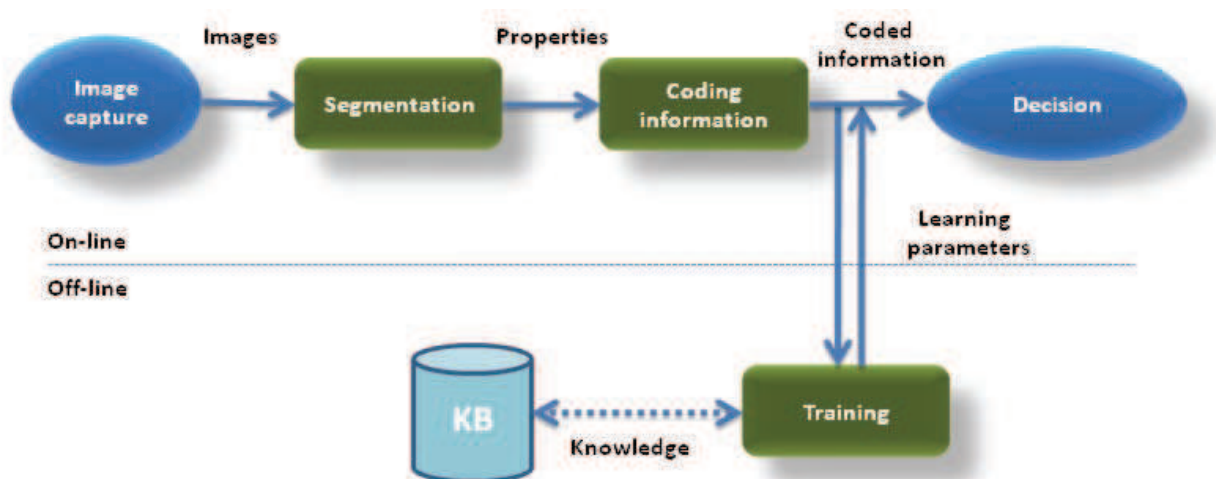

Fig. 2. General scheme of a training-decision process

This scheme is valid for both the individual nature and combined classifiers.

- *Image capture*: it consists in obtaining the images, either obtained from a databank or directly from the scene by the corresponding sensor.

- *Segmentation*: segmentation is the process involving the extraction of structures or features in the images. From the point of view of image treatment, a feature can be a region or an edge that belongs to any object. A feature can also be a pixel belonging to a border, a point of interest or simply a pixel of the image regardless of inside or outside any of the aforementioned structures. In the case of a region can be its area, perimeter, intensity average or any other property describing the region. The pixels are the features used in this work. In our case, the attributes or properties of the pixels will be their spectral components. Consequently, the segmentation process includes both, feature extraction and properties.
- *Coding information*: This phase includes the structuring of the information to be subsequently used by both methods learning and classification. Each feature taken during the previous phase are the samples represented by vectors, whose components are the properties of the feature under analysis. As mentioned previously, the features to consider are the pixels. Given a pixel in the spatial location (i, j), if it is labelled as k we have $k \equiv (i,j)$, being x_k the vector whose components are the representative spectral values of that pixel in the *RGB* colour model, i.e. $x_k = \left\{ x_{k_1}, x_{k_2}, x_{k_3} \right\} \equiv \left\{ R, G, B \right\} \in \Re^3$ and therefore, in this case, the vector belongs to the three-dimensional space. The samples are coded for both the training process and the decision process; then we will have training samples and classification samples according to the stage where they are processed.
- *Learning/Training*: with the available samples properly encoded, the training process is carried out according to the method selected. The learning resulting parameters are stored in the Knowledge Base (*KB*), Figure 2, for being used during the decision phase.
- *Identification/Decision*: at this stage we proceed to identify a new feature or sample, which has not yet been classified as belonging to one of the existing classes of interest. To do that the previously learned and stored parameters in KB are retrieved, thereafter through the corresponding decision function, inherent to each method, the class to which it belongs is identified. This process is also called recognition or classification. It is sometimes common that the classified samples can be incorporated back into the system, now as training samples to proceed to a new learning process and therefore to carry out a new updating of the parameters associated to each method, that are stored again in the KB. This is known as incremental learning.

As mentioned before, in our approach there are other two relevant textures that must be identified. They are specifically the sky and the grass. For a pixel belonging to one of such areas the R coefficient should be low because of its homogeneity. This is a previous criterion for identifying such areas, where the 'low' concept is mapped assuming that R should be less than the previous threshold T_1. Nevertheless, this is not sufficient because there are other different areas which are not sky or grass fulfilling this criterion. Therefore, we apply a classification technique based on the combination of the parametric Bayesian estimator (PB) and Parzen window (PZ) approaches. The choice of these classifiers is based on its proven effectiveness when applied individually in various fields of application, including image classification. According to (Kuncheva, 2004), if they are combined the results improve. Both PB and PZ consist of two phases: training and decision.

2.2.1 Training phase

We start with the observation of a set X of n training patterns, i.e. $X = \left\{ x_1, x_2, ..., x_n \right\} \in \Re^q$. 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 our experiments, we identify the cluster c_1 with the sky and the cluster c_2 with the grass. The 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 $q = 3$.

a.   *Parametric Bayesian Classifier (PB)*

This method has traditionally been identified within the unsupervised classification techniques (Escudero, 1977). Given a generic training sample $x \in \Re^q$, the goal is to estimate the membership probabilities to each class c_j, i.e. $P_1(c_j \mid x)$. This technique assumes that the density function of conditional probability for each class is known, resulting unknown the parameters. A widespread practice, adopted in our approach, is to assume that the shape of these functions follows the law of Gaussian or Normal distribution, according to the following expression,

$$p_1\left(x \mid m_j, C_j\right) = \frac{1}{\left(2\pi\right)^{q/2}\left|C_j\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(x - m_j\right)^T C_j^{-1}\left(x - m_j\right)\right\} \tag{1}$$

where m_j and C_j are, respectively, the mean and covariance matrix of class c_j, i.e. statistical or unknown parameters to be estimated, T denotes the transposed matrix and q express the dimensionality of the data by $x \in \Re^q$.

The hypotheses assumed by the unsupervised classification techniques are:

1.   There are c classes in the problem.
2.   The sample x comes from these c classes, although the specific class to which it belongs is unknown.
3.   The a priori probability that the sample belongs to class c_j, $P(c_j)$ is in principle unknown.
4.   The density function associated with each class has a known form, being unknown the parameters of that function.

With this approach it is feasible to implement the Bayes rule to obtain conditional probability that x_s belongs to class c_j, by the following expression (Huang & Hsu, 2002),

$$P_1\left(c_j \mid x_s\right) = \frac{p\left(x_s \mid m_j, C_j\right)P\left(c_j\right)}{\displaystyle\sum_{j=1}^{c} p\left(x_s \mid m_j, C_j\right)} \tag{2}$$

Knowing the shapes of probability density functions, the parametric Bayesian method seeks to estimate the best parameters for these functions.

b.   *Parzen window (PZ)*

In this process, as in the case of parametric Bayesian method, the goal remains the estimation of the membership probabilities of sample x to each class c_j, that is $P_2(c_j \mid x)$. Therefore, the problem arises from the same point of view, making the same first three hypotheses and replacing the fourth by a new more general: "the shape of the probability density function associated with each class is not known". This means that in this case there are no parameters to be estimated, except the probability density function (Parzen, 1962; Duda et al. 2001). The estimated density function turns out to be that provided by equation

(3), where $D(\cdot) = (x - x_k)^T C_j^{-1} (x - x_k) / 2h_j^2$, q represents the dimension of the samples in the space considered, T indicates the vector transpose operation.

$$p_2\left(x \mid c_j\right) = \frac{1}{n_j} \sum_{k=1}^{n_j} \left\{ \frac{\exp\left\{-D\left(x, x_k, h_j\right)\right\}}{(2\pi)^{q/2} h_j^{n_j} \left|C_j\right|^{1/2}} \right\} \tag{3}$$

According to equation (3), this classifier estimates the density function probability given the training samples associated with each class, requiring that the samples are distributed, i.e. the partition must be available. Also the covariance matrices associated with each of the classes are used. The full partition and covariance matrices are the parameters that this classifier stored in the KB during the training phase. In fact, the covariance matrices are the same as those obtained by PB.

During the decision phase, PZ extracts from KB both the covariance matrices C_j and the available training samples are distributed in their respective classes. With them the probability density function given in equation (3) is generated. Thus, from a new sample x_s conditional probabilities are obtained according to this equation, $P_2\left(x_s \mid c_j\right)$. The probability that the sample x_s belongs to the class w_j can be obtained by again applying Bayes rule,

$$P_2\left(c_j \mid x_s\right) = \frac{p_2\left(x_s \mid c_j\right) P\left(c_j\right)}{\sum_{j=1}^{c} p_2\left(x_s \mid c_j\right)} \tag{4}$$

2.2.2 Decision phase

After the training phase, a new unclassified sample $x_s \in \Re^q$ must be classified as belonging to a cluster c_j. Here, each sample, like each training sample, represents a pixel at the image with the R,G,B components. PB computes the probabilities that x_s belong to each cluster from equation (2) and PZ computes the probabilities that x_s belong to each cluster from equation (4). Both probabilities are the outputs of the individual classifiers ranging in [0,1]. They are combined by using the *mean rule* (Kuncheva, 2004). (Tax et al., 2000) compare performances of combined classifiers by averaging and multiplying. As reported there, combining classifiers which are trained in independent feature spaces result in improved performance for the product rule, while in completely dependent feature spaces the performance is the same for the product and the average. In our RGB feature space high correlation among the R, G and B spectral components exists (Littmann & Ritter, 1997; Cheng et al., 2001). High correlation means that if the intensity changes, all the three components will change accordingly. Therefore we chose the mean for the combination, which is computed as: $m_{sj} = \left(P_1\left(c_j \mid x_s\right) + P_2\left(c_j \mid x_s\right)\right) / 2$. The pixel represented by x_s is classified according to the following decision rule: $x_s \in c_j$ if $m_{sj} > m_{sh}$ and $m_{sj} > T_2$ otherwise the pixel remains unclassified. We have added, to the above rule, the second term with the logical *and* operator involving the threshold T_2 because we are only identifying pixels belonging to the sky or grass clusters. This means that the pixels belonging to textures different from the

previous ones remain unclassified, and they become candidates for the stereo matching process. The threshold T_2 has been set to 0.8 after experimentation. This is a relative high value, which identifies only pixels with a high membership degree in either c_1 or c_2. We have preferred to exclude only pixels which belong clearly to one of the above two textures.

Figure 3(b) displays the result of applying the segmentation process to the left image in Figure 3(a). 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 correspondence process. There are white trunks representing trees very far from the sensor. They are not considered because are out of our interest from the point of view of forest inventories.
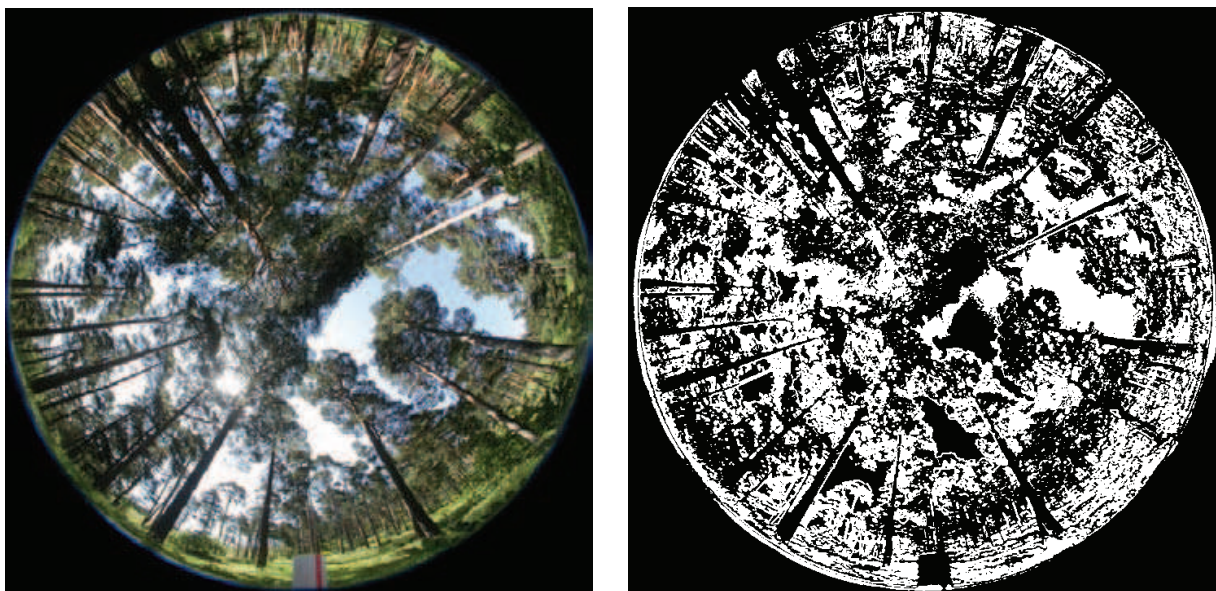
Fig. 3. (a) Original left image; (b) segmented left image where white areas are textures without interest (sky, grass and leaves) and the black ones the pixels to be matched.

It is difficult to validate the results obtained by the segmentation process, but we have verified that without this process, the error for stereovision matching strategies is increased by a quantity that represents on average about 9-10 percentage points. In addition to this quantitative improvement it is easy to deduce the existence of a qualitative improvement by the fact that some pixels belonging to textures not excluded by the absence of segmentation, they are incorrectly matched with pixels belonging to the trunks, this do not occur when these textures are excluded because they were not offered this possibility. This means that the segmentation is a fundamental process in our stereovision system and justifies its application.

3. Stereovision matching

Once the image segmentation process is finished, we have identified pixels belonging to three types of textures which are to be discarded during the next stereovision matching process, because they are without interest. Hence, we only apply the stereovision matching

process to the pixels that do not belong to any of the previous textures. As we explained before, due to the different locations of the tree's crowns there exists an important lighting variability between both images of the stereoscopic pair; this makes the matching process a difficult task.

As mentioned in section 1, in stereovision there are several constraints that can be applied. In our approach we apply three of them: epipolar, similarity and uniqueness. 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 first pixel. Through the combination of similarity measurements computed from a set of attributes extracted of each pixel (similarity constraint), we obtain the final decision about the best match among candidates by applying the uniqueness constraint. Epipolar constraint is explained in section 3.1 and similarity and uniqueness constraints in section 3.2.

3.1 Epipolar constraint: system geometry

Figure 4 displays the stereovision system geometry (Abraham & Förstner, 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 along the y-axes with respect to the two positions of the camera; r is the distance between the image point and the optical axis; R is the image radius, identical in both images.

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

$$r = \sqrt{x_{i1}^2 + y_{i1}^2} \; ; \; \alpha_1 = (r90^\circ)/R \; ; \; \beta = tg^{-1}\left(y_{i1}/x_{i1}\right) \tag{5}$$

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,

$$X_1 = d\cos\beta \; ; \; Y_1 = d\sin\beta \; ; \; Z_1 = \sqrt{X_1^2 + Y_1^2}\Big/\tan\alpha_1 \tag{6}$$

From (6) 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,

$$X_2 = X_1 \; ; \; Y_2 = Y_1 + y_{12} \; ; \; Z_2 = Z_1 \tag{7}$$

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

$$x_{i2} = \frac{2R\arctan\left(\sqrt{X_2^2 + Y_2^2}\Big/Z_2\right)}{\pi\sqrt{(Y_2/X_2)^2 + 1}} \; ; \; y_{i2} = \frac{2R\arctan\left(\sqrt{X_2^2 + Y_2^2}\Big/Z_2\right)}{\pi\sqrt{(X_2/Y_2)^2 + 1}} \tag{8}$$

Using only a camera or a camera position, we capture a unique image and the 3D points belonging to the line $\overline{O_1P}$ are all imaged in the unique point (x_{i1}, y_{i1}). So, the 3D coordinates cannot be obtained from a single image. 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 varying the parameter d in the 3-D space. So, given the imaged point (x_{i1}, y_{i1}) in image-1 (left) and following the epipolar line, we obtain a list of m potential corresponding candidates represented by (x_{i2}, y_{i2}) in image-2 (right).
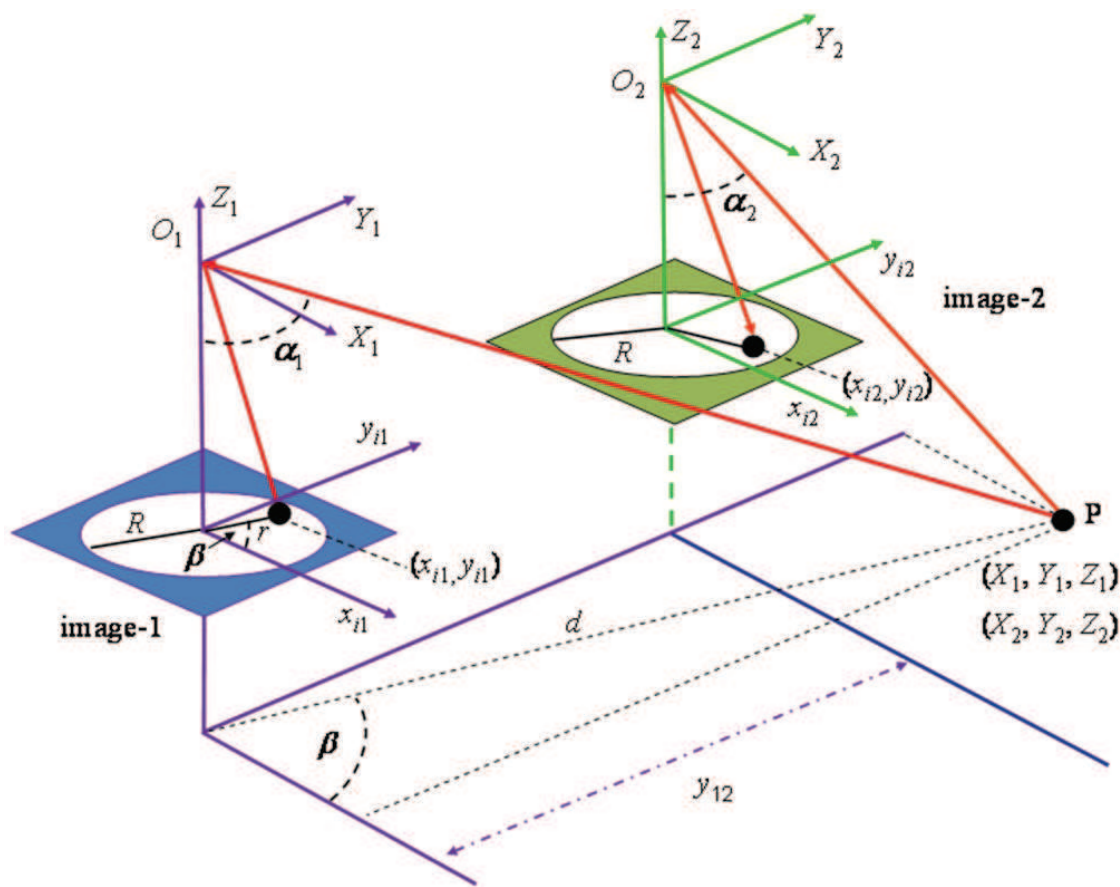

Fig. 4. Geometric projections and relations for the fisheye based stereo vision system.

3.2 Similarity and uniqueness constraints

Each pixel l in the left image is characterized by its attributes; one of such attributes is denoted as A_l. In the same way, each candidate i in the list of m candidates is described by identical attributes, A_i. So, we can compute differences between attributes of the same type A, obtaining a similarity measure for each attribute as,

$$s_{iA} = 1/\left(1 + |A_l - A_i|\right); \quad i = 1, \ldots, m \tag{9}$$

$s_{iA} \in [0,1]$, $s_{iA} = 0$ if the difference between attributes is large enough (minimum similarity), otherwise if they are equal ($s_{iA} = 1$, maximum similarity).

In this chapter, we use the following five attributes for describing each pixel (feature): a) Gabor filter; b) variance as a measure of the texture; c) RGB color; d) CIE lab color and e) gradient magnitude. The first two are area-based computed on a 3×3 neighborhood around each pixel (Pajares & Cruz, 2007). The latter three are considered as feature-based (Lew et

al., 1994). Gabor filter is basically a bi-dimensional Gaussian function centered at origin (0,0) with variance S modulated by a complex sinusoid with polar frequency (F,W) and phase P. The RGB color involves the three Red-Green-Blue spectral components and the absolute value in equation (9) is extended as: $|A_l - A_i| = \sum_H |H_l - H_i|$, H = R,G,B. In the same way, the CIE lab color involves the three l-a-b components and the absolute value in equation (9) is extended as: $|A_l - A_i| = \sum_H |H_l - H_i|$, $H = l,a,b$. Gradient magnitude is computed by applying the first derivative (Pajares & Cruz, 2007), over the intensity image after its transformation from the RGB plane to the HSI (hue, saturation, intensity) one.

Other attributes have been unsuccessfully used in previous works in the same forest environment, e.g. correlation, gradient direction and Laplacian (Herrera, 2010; Herrera et al., 2011a, 2011b). While gradient magnitude, RGB color and texture obtained the best individual results, respectively. For this reason they are used in this work.

Given a pixel in the left image and the set of m candidates in the right one, we compute the following similarity measures for each attribute A: s_{ia} (Gabor filter), s_{ib} (texture), s_{ic} (RGB color), s_{id} (CIE lab color) and s_{ie} (gradient magnitude). The identifiers in the sub-indices identify the attributes according to the above assignments.

Now we must match each pixel l in the left image with the best of the potential candidates (uniqueness constraint). This is based on a majority voting criterion (MVC). So, given l and its i candidates, we have available s_{ia}, s_{ib}, s_{ic}, s_{id} and s_{ie}, so that we can make individual decisions about the best candidate i based on maximum similarity measurements among the set of candidates. We determine the best match by choosing the candidates with the maximum similarity for each individual attribute and select the one which has been chosen according to the majority of attributes. Each one of the five attributes, used separately, allows determining a disparity map for comparison purposes.

The sexagesimal system is used in measuring angles. The practical unit of angular measure is the degree, of which there are 360 in a circle. 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.

Given a stereo pair of the twenty used for testing, for each pixel we obtain its disparity as follows. Considering the five attributes separately, used as criteria in the MVC, and applying a maximum similarity criterion according to equation (9) among the m candidates, we obtain a disparity map for each attribute. So, for comparative purposes we show for the area in Figure 5(b), the disparity maps obtained by Gabor Filter, texture, RGB and CIE lab colors, and gradient magnitude in Figures 5(c) to 5(g), respectively. By applying the MVC approach based on maximum similarity, we obtain the disparity map displayed in Figure 5(h). The color bar in Figure 5(i) shows the disparity level values according to the color for each disparity map.

An important observation comes from the main trunk in the Figure 5(b); indeed, in the corresponding disparity maps obtained by Gabor filter and texture, the disparity values range from 1.5 to 5.5, but in RGB and CIE lab colors, and gradient magnitude they range from 3.5 to 5.5. In the disparity map obtained by MVC strategy, the low level values have been removed, such that the disparities range from 4.5 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.

The best individual results, according to the five attributes, are obtained through the similarities provided by the gradient magnitude (s_{ie}). This implies that it is the most relevant attribute. Nevertheless, the main relevant results are obtained by the proposed MVC approach in terms of less percentage of error. This together with the qualitative improvement provided by this approach, as explained above, allows us to conclude that this is a suitable method for computing the disparity map in this kind of images.
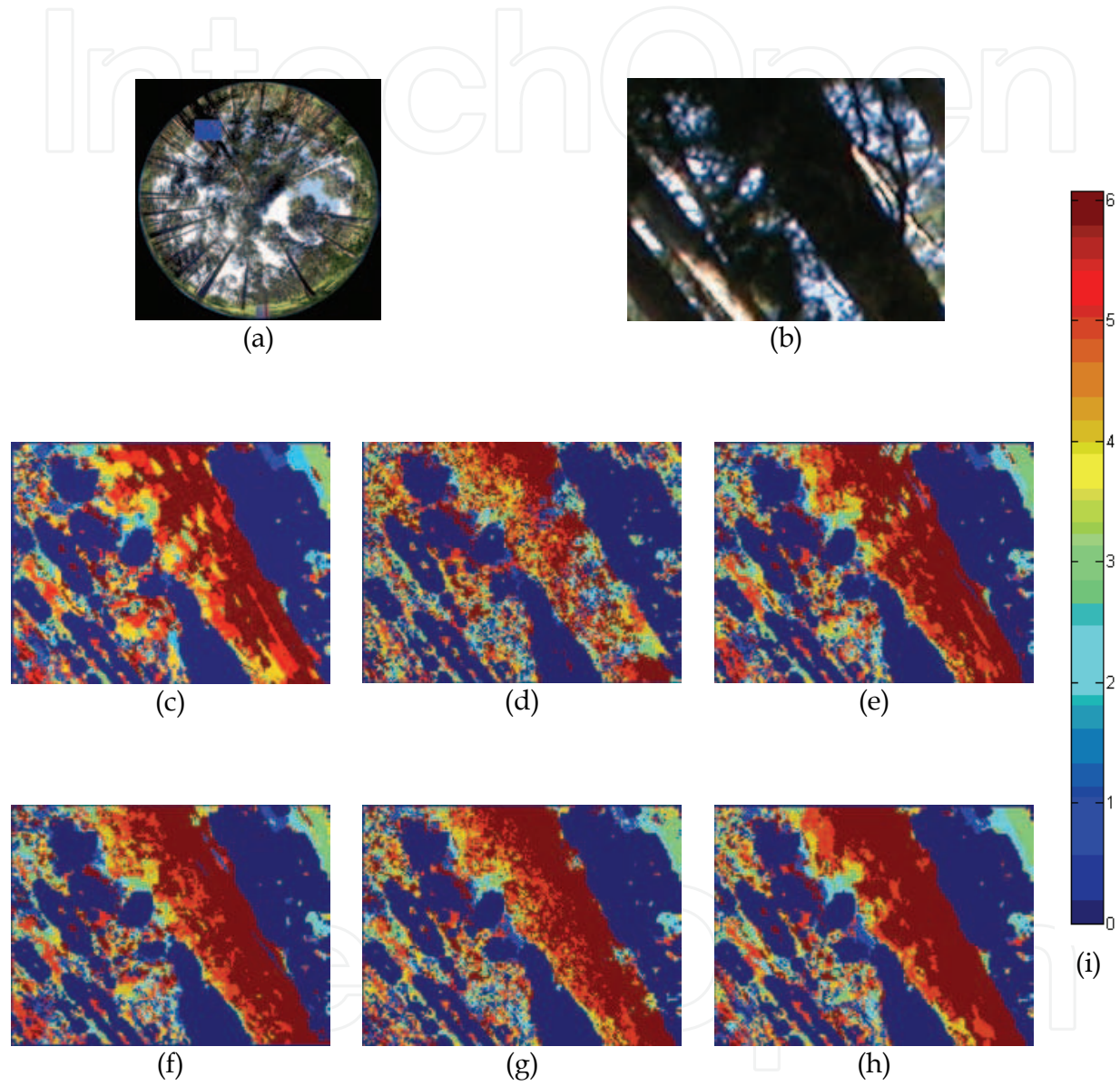
Fig. 5. (a) Hemispherical left image; (b) left expanded area corresponding to the blue signed area in (a); disparity maps obtained by (c) Gabor filter, (d) texture, (e) RGB color, (f) CIE lab color, (g) gradient magnitude, (h) MVC; (i) color bar shows the disparity level values according to the color for each disparity map.

Other combined decision making approaches have been successfully used in previous works in the same forest environment, where the final decision about the correct match,

among the candidates in the list, is made according to techniques used for combining classifiers conveniently adapted in our approach to be applied for the stereovision matching (Herrera et al., 2009a, 2009b, 2009c, 2011b; Herrera, 2010; Pajares et al., 2011). In (Herrera et al., 2011a) the similarity and uniqueness constraints are mapped through a decision making strategy based on a weighted fuzzy similarity approach.

4. Conclusions

This chapter presents segmentation and matching strategies for obtaining a disparity map from hemispherical stereo images captured with fisheye lenses. This is carried out through a segmentation process which uses the combination of the parametric Bayesian estimator and the Parzen's window classifiers and the variance as method for texture analysis. The goal of the image segmentation process is to classify and exclude the pixels belonging to one of the three kinds of textures without interest in the images: sky, grass in the soil and leaves. The combined classification strategy classifies sky and grass textures and the variance tries to isolate the leaves based on statistical measures. The exclusion of these textures is useful because the errors that they could introduce during the correspondence can be considerably reduced. While others individual classifiers might have been chosen as a different combined strategy, as the Fuzzy Clustering, the Generalized Lloyd algorithm and the Self-Organizing Maps (Pajares & Cruz, 2007), the combination of both in relation to the improvement of the results according to the set of images used shows its promising possibilities. All this does not preclude the future use of new classifiers and the combination of other strategies for the type of images analyzed.

Once the image segmentation process is finished, an initial disparity map is obtained by applying three stereovision matching constraints (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 a majority voting criterion, which is a decision strategy based on combining similarity measurements from five attributes extracted of each pixel. The proposed combined strategy outperforms the methods that use similarities separately. Based on this, some optimization approaches could be used, such as simulated annealing or Hopfield neural networks, where the smoothness constraint and the Gestalt's principles could be applied under an energy minimization based process.

The method proposed can be applied for similar forest environments where pixels are the key features to be matched. Applications using this sensor are based on identical geometry and image projection, although the matching strategy could be completely different. This occurs in (Herrera et al., 2009d), based on region segmentation where the images are very different and captured under different illumination conditions in Rebollo oak forests.

5. Acknowledgments

6. References

Abraham, S. & Förstner, W. (2005). Fish-eye-stereo calibration and epipolar rectification. *Photogrammetry and Remote Sensing*, Vol.59, pp. 278-288, ISSN 0924-2716

Cheng, H.D., Jiang, X.H., Sun, Y. & Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern Recognition*, Vol.34, No.12, pp. 2259–2281, ISSN 0031-3203

Duda, R.O., Hart, P.E. & Stork, D.S. (2001). *Pattern Classification* (2nd edition), Wiley & Sons, ISBN 978-04-710-5669-0, New York, USA

Escudero, L.F. (1977). *Reconocimiento de patrones*, Paraninfo, ISBN 978-84-283-0898-4, Madrid, Spain

Gonzalez, R.C. & Woods, R.E. (2008). *Digital Image Processing* (3rd edition), Prentice Hall, ISBN 978-01-316-8728-8, New Jersey, USA

Gregoire, T.G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, Vol.28, pp. 1429–1447, ISSN 0045-5067

Guijarro, M., Pajares, G. & Herrera, P.J. (2009). Image-Based Airborne Sensors: A Combined Approach for Spectral Signatures Classification through Deterministic Simulated Annealing. *Sensors*, Vol.9, No.9, pp. 7132-7149, ISSN 1424-8220

Guijarro, M., Pajares, G. & Herrera, P.J. (2008). On Combining Classifiers by Relaxation for Natural Textures in Images, In: *Lecture Notes in Artificial Intelligecnce 5271*, E. Corchado, A. Abraham & W. Pedrycz (Eds.), 345-352, Springer-Verlag, ISBN 978-3-540-87656-4, Berlin, Germany

Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J. & Cruz, J.M. (2011a). A Stereovision Matching Strategy for Images Captured with Fish-Eye Lenses in Forest Environments. *Sensors*, Vol.11, pp. 1756-1783, ISSN 1424-8220

Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J. & Cruz, J.M. (2011b). Combining Support Vector Machines and simulated annealing for stereovision matching with fish eye lenses in forest environments. *Expert Systems with Applications*, Vol.38, pp. 8622–8631, ISSN 0957-4174

Herrera, P.J. (2010). *Correspondencia estereoscópica en imágenes obtenidas con proyección omnidireccional para entornos forestales*. PhD Dissertation (in Spanish), Faculty of Computer Science, Complutense University of Madrid, 2010.

Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J. & Cruz, J.M. (2009a). Choquet Fuzzy Integral applied to stereovision matching for fish-eye lenses in forest analysis, In: *Advances in Soft Computing 61*, W. Yu & E.N. Sanchez (Eds.), 179–187, Springer-Verlag, ISBN 978-3-642-03156-4, Berlin, Germany

Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J. & Cruz, J.M. (2009b). Combination of attributes in stereovision matching for fish-eye lenses in forest analysis, In: *Lecture Notes in Computer Science 5807*, J. Blanc-Talon et al. (Eds.), 277-287, Springer-Verlag, ISBN 978-3-642-04697-1, Berlin, Germany

Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J. & Cruz, J.M. (2009c). Fuzzy Multi-Criteria Decision Making in Stereovision Matching for Fish-Eye Lenses in Forest Analysis, In: *Lecture Notes in Computer Science 5788*, H. Yin & E. Corchado (Eds.), 325-332, Springer-Verlag, ISBN 978-3-642-04394-9, Berlin, Germany

Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J., Cruz, J.M. & Montes, F. (2009d). A Featured-Based Strategy for Stereovision Matching in Sensors with Fish-Eye Lenses for Forest Environments. *Sensors*, Vol.9, pp. 9468-9492, ISSN 1424-8220

Huang, H.J. & Hsu, C.N. (2002). Bayesian classification for data from the same unknown class. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, Vol.32, No.2, pp. 137-145, ISSN 1083-4419.

Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms*, Wiley, ISBN 978-0-471-21078-8, New Jersey, USA

Lew, M.S., Huang, T.S. & Wong, K. (1994). Learning and feature selection in stereo matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16, pp. 869–881, ISSN 0162-8828.

Littmann, E. & Ritter, H. (1997). Adaptive color segmentation -A comparison of neural and statistical methods. *IEEE Transactions on Neural Networks*, Vol.8, No.1, pp. 175-185, ISSN 1045-9227.

Montes, F., Ledo, A., Rubio, A., Pita, P. & Cañellas, I. (2009). Use of estereoscopic hemispherical images for forest inventories, *Proceedings of the International Scientific Conference Forest, Wildlife and Wood Sciences for Society development*, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic.

Pajares, G, Herrera, P.J. & Cruz, J.M. (2011). Combining Stereovision Matching Constraints for Solving the Correspondence Problem, In: *Advances in Theory and Applications of Stereo Vision*, Asim Bhatti (Ed.), ISBN: 978-953-307-516-7, InTech, Available from:
http://www.intechopen.com/articles/show/title/combining-stereovision-matching-constraints-for-solving-the-correspondence-problem

Pajares, G., Guijarro, M., Herrera, P.J. & Ribeiro, A. (2009). Combining Classifiers through Fuzzy Cognitive Maps in natural images. *IET Computer Vision*, Vol.3, No.3, pp. 112-123, ISSN 1751-9632.

Pajares, G. & Cruz, J.M. (2007). *Visión por Computador: Imágenes digitales y aplicaciones* (2nd edition), RA-MA, ISBN 978-84-7897-831-1, Madrid, Spain

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, Vol.33, No.3, pp. 1065-1076, ISSN 0003-4851.

Scharstein, D. & Szeliski, R. (2002). A Taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, Vol.47, No.1-3, pp. 7–42, ISSN 0920-5691

Schwalbe, E. (2005). Geometric modelling and calibration of fisheye lens camera systems, *Proceedings of the ISPRS working group V/5 'Panoramic Photogrammetry Workshop', Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36, Part 5/W8, ISSN 1682-1750, Berlin, Germany, February 2005

Tax, D.M.J., Breukelen, van M., Duin, R.P.W. & Kittler, J. (2000). Combining multiple classifiers by averaging or by multiplying?. *Pattern Recognition*, Vol.33, pp. 1475-1485, ISSN 0031-3203

Advances in Stereo Vision
Edited by Prof. Jose R.A. Torreao

Stereopsis is a vision process whose geometrical foundation has been known for a long time, ever since the experiments by Wheatstone, in the 19th century. Nevertheless, its inner workings in biological organisms, as well as its emulation by computer systems, have proven elusive, and stereo vision remains a very active and challenging area of research nowadays. In this volume we have attempted to present a limited but relevant sample of the work being carried out in stereo vision, covering significant aspects both from the applied and from the theoretical standpoints.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Pedro Javier Herrera, Gonzalo Pajares, María Guijarro, José J. Ruz and Jesús M. De La Cruz (2011). Segmentation and Stereoscopic Correspondence in Images Obtained with Omnidirectional Projection for Forest Environments, Advances in Stereo Vision, Prof. Jose R.A. Torreao (Ed.), ISBN: 978-953-307-837-3, InTech, Available from: http://www.intechopen.com/books/advances-in-stereo-vision/segmentation-and-stereoscopic-correspondence-in-images-obtained-with-omnidirectional-projection-for-

INTECH
open science | open minds