Choquet Fuzzy Integral Applied to Stereovision Matching for Fish-Eye Lenses in Forest Analysis

P. Javier Herrera[1], Gonzalo Pajares[1], María Guijarro[2], José J. Ruz[3], and Jesús M. de la Cruz[3]

[1] Dpto. Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Universidad Complutense, 28040 Madrid, Spain
pjherrera@pdi.ucm.es, pajares@fdi.ucm.es
[2] Centro Superior de Estudios Felipe II, Ingeniería Técnica en informática de Sistemas 28300 Aranjuez, Madrid, Spain
mguijarro@cesfelipesegundo.com
[3] Dpto. Arquitectura Computadores y Automática, Facultad de Informática, Universidad Complutense, 28040 Madrid, Spain
{jjru, jmcruz}@dacya.ucm.es

Abstract. This paper describes a novel stereovision matching approach based on omni-directional images obtained with fish-eye lenses in forest environments. The goal is to obtain a disparity map as a previous step for determining the volume of wood in the imaged area. The interest is focused on the trunks of the trees, due to the irregular distribution of the trunks; the most suitable features are the pixels. A set of six attributes is used for establishing the matching between the pixels in both images of the stereo pair. The final decision about the matched pixel is taken based on the Choquet Fuzzy Integral paradigm, which is a technique well tested for combining classifiers. The use and adjusting of this decision approach to our specific stereo vision matching problem makes the main finding of the paper. The procedure is based on the application of three well known matching constraints. The proposed approach is compared favourably against the usage of simple features and other fuzzy strategy that combines the simple ones.

Keywords: Choquet Fuzzy Integral, Fish-eye stereo vision, Stereovision matching, omni-directional forest images.

1 Introduction

One important task in forests maintenance is to determine the volume of wood in an area for different purposes, including the control of growth of the trees. This task can be carried out by stereovision systems. Fish-eye lenses allow imaging a large sector of the surrounding space with omni-directional vision. This justifies its use.

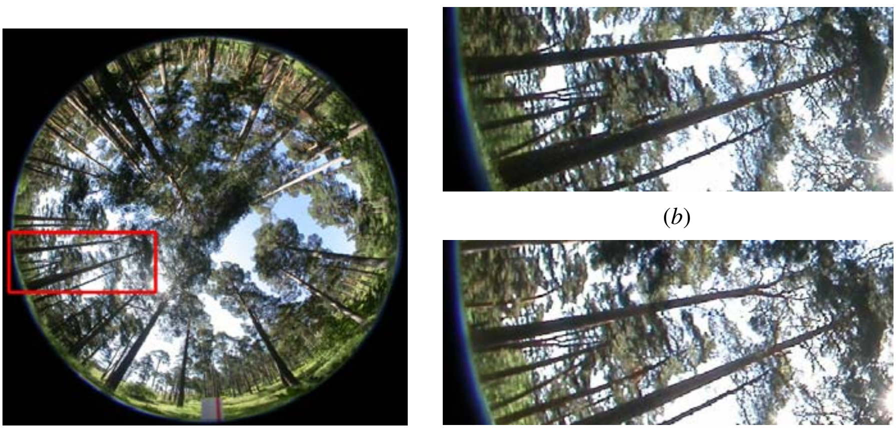
According to [1] we can view the classical problem of stereo analysis as consisting of the following steps: image acquisition, camera modelling, feature acquisition, image matching, depth determination and interpolation. The key step is that of image matching. This is the process of identifying the corresponding points in two images

W. Yu and E.N. Sanchez (Eds.): Advances in Computational Intell., AISC 61, pp. 179–187.
springerlink.com

that are cast by the same physical point in the 3-D space. This paper is devoted solely to the matching one. Two sorts of techniques have been used for matching: area-based and feature based [2].

Area-based stereo techniques [3] use correlation between brightness (intensities) patterns in the local neighbourhood of a pixel in one image with brightness patterns in the local neighbourhood of the other image. Also statistical textures can be considered under this category. Feature-based methods [4] use set of pixels with similar attributes, colour, gradient (module and direction) or Laplacian. These are the six attributes available to be used in our matching procedure.

Figure 1(*a*) displays one omni-directional image (let's say the left one) of the stereo pair captured with a fisheye lens. Figure 1(*b*) displays the signed and expanded area on Figure 1(*a*). In Figure 1(*c*) the corresponding area in the right image of the stereo pair is displayed. Due to the different locations of the tree's crowns there exists an important lighting variability between both areas; this makes the matching process a difficult task. This is applicable for the whole image.



(*a*) (*b*) (*c*)

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

The following three stereovision constraints can be applied for solving the matching problem. *Epipolar*: derived from the system geometry, given a pixel in one image its correspondence will be on the epipolar line. *Similarity*: matched pixels display similar attributes. *Uniqueness*: a pixel in the left image must be matched to a unique pixel in the right one.

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. We also apply the similarity constraint based on the six attributes, obtaining six similarity measures, which are conveniently combined. The final decision about the correct match, among the list of candidates, is made according to the support that each candidate receives by applying the Choquet Fuzzy Integral (CFI) paradigm. This unique selection implies the application of the uniqueness constraint. The matching through the CFI makes the main contribution of the paper. The proposed approach is compared favourably against the usage of individual area-based and feature-based correspondence techniques.

This work is organized as follows. Section 2 describes the design of the matching process; including a brief overview of the CFI paradigm. Section 3 describes the results obtained by using the combined CFI approach, and comparing these results with those obtained by applying each individual strategy. Section 4 presents the conclusions and future work.

2 Design of the Matching Process

2.1 Epipolar: System Geometry

Figure 2 displays the stereo vision system geometry [5]. 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 [6], the following geometrical relations can be established,

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

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,

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

From (2) 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:

$$X_2 = X_1;\quad Y_2 = Y_1 + y_{12};\quad Z_2 = Z_1 \tag{3}$$

Assuming no lenses radial distortion, we can find the imaged coordinates of the 3D point in image-2 as [6],

$$x_{i2} = \frac{2R\arctan\left(\sqrt{X_2^2 + Y_2^2}\Big/Z_2\right)}{\pi\sqrt{\left(Y_2/X_2\right)^2 + 1}}\ ;\ y_{i2} = \frac{2R\arctan\left(\sqrt{X_2^2 + Y_2^2}\Big/Z_2\right)}{\pi\sqrt{\left(X_2/Y_2\right)^2 + 1}} \tag{4}$$

Using only a camera, we capture a unique image and the 3D points belonging to the line $\overline{O_1P}$, are all imaged in 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).

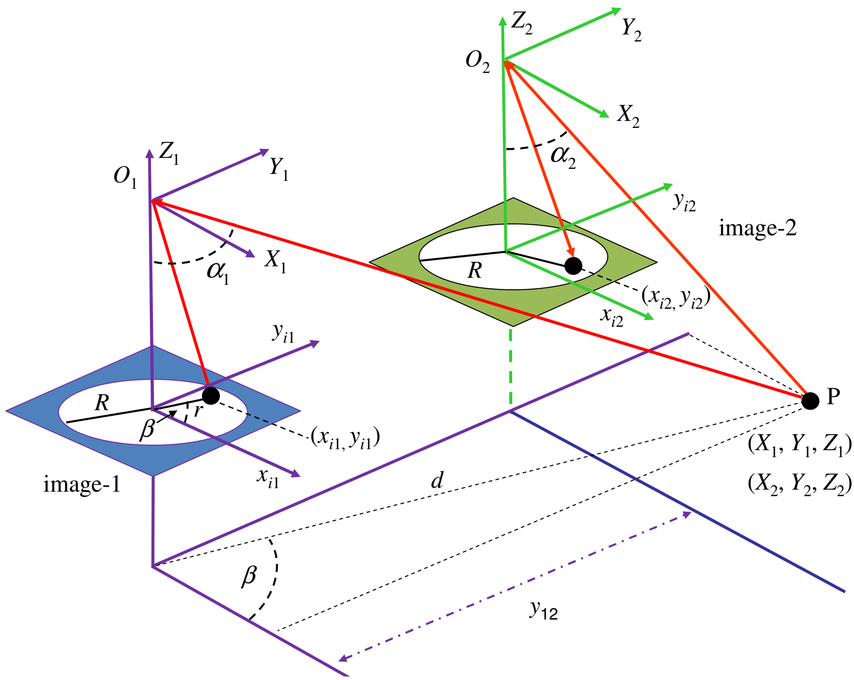


Fig. 2. Geometric projections and relations for the fish-eye based stereo vision system

2.2 Similarity: Attributes for Area and Feature-Based

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,

$$s_{iA} = \left(1 + \left|A_l - A_i\right|\right)^{-1}; \quad i = 1,\dots,m \tag{5}$$

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

As mentioned before, in this paper we use the following six attributes for describing each pixel (feature): *a*) correlation; *b*) texture; *c*) colour; *d*) gradient magnitude; *e*) gradient direction and *f*) Laplacian. Both first ones are area-based computed on a 3×3 neighbourhood around each pixel through the correlation coefficient [7] and standard deviation [8]. The four remaining ones are considered as feature-based [4]. The colour involves the three red-green-blue spectral components

(R,G,B) and the absolute value in the equation (5) is extended as: $\left|A_l - A_i\right| = \sum_H \left|H_l - H_i\right|,\ H = \mathrm{R,G,B}$.

Gradient (magnitude and direction) and Laplacian are computed by applying the first and second derivatives [8], over the intensity image after its transformation from the RGB plane to the HSI (hue, saturation, intensity) one. 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} (correlation), s_{ib} (colour), s_{ic} (texture), s_{id} (gradient magnitude), s_{ie} (gradient direction) and s_{if} (Laplacian). The identifiers in the sub indices identify the attributes according to the above assignments.

2.3 Uniqueness: Applying the Choquet Fuzzy Integral Paradigm

Now we must match each pixel l in the left image with the best of the potential candidates (uniqueness). This is a decision based on the CFI paradigm. This paradigm allows combining the individual similarities, which are computed through the equation (5).

The CFI requires the computation of the relevance assigned for each attribute from which we can compute the so-called fuzzy densities. This is solved by computing the $\lambda-$fuzzy measure using the data [9]. The calculation starts with selecting a set of six fuzzy values, g^a, g^b, g^c, g^d, g^e, g^f, each one representing the individual relevance of the associated attribute. The attributes are the six described above, i.e. $\Omega \equiv \{a,b,c,d,e,f\}$ associated to correlation, texture, color, gradient magnitude, gradient direction and Laplacian.

The value of λ needed for calculating the fuzzy densities g is obtained as the unique real root greater than -1 of the polynomial,

$$\lambda + 1 = \prod_{j\in\Omega}\left(1+\lambda g^j\right) \tag{6}$$

The individual relevancies for each attribute are computed from the data, as described later in the section 3*a*.

Once the g^a, … g^f are obtained and λ is found, the fuzzy integral works as follows:

1. For a given pixel l in the left image, we compute the similarities through the equation (5) between l and every candidate i, with $i = 1,\ldots,m$, obtaining a column vector as: $[s_{ia}, s_{ib}, s_{ic}, s_{id}, s_{ie}, s_{if}]^T$; without lost of generality assume that s_{ia} is the highest similarity value and s_{if} the lowest. This vector is arranged under this criterion, i.e. $s_{ia} > s_{ib} > s_{ic} > s_{id} > s_{ie} > s_{if}$.
2. Arrange the above fuzzy values correspondingly with the mentioned arrangement, i.e. g^a, g^b, g^c, g^d, g^e, g^f and set the first fuzzy density $g(a) = g^a$.
3. Compute the remaining fuzzy densities according to the recursive procedure given in the equation (7).

$$\begin{aligned} g(b) &= g^b + g(a) + \lambda g^b g(a) \\ g(c) &= g^c + g(b) + \lambda g^c g(b) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ g(f) &= g^f + g(e) + \lambda g^f g(e) \end{aligned} \tag{7}$$

4. Calculate for each candidate i, the support received to be matched with l as,

$$\mu_i(l) = s_{ia} + \sum_{h=b}^{f} \left[s_{i(h-1)} - s_{ih} \right] g(h-1) \tag{8}$$

5. The decision about the best match is made by selecting the maximum support $\mu_i(l)$ among all candidates.

3 Results

The system is based on the scheme of the figure 2, with a baseline of 1 meter. The cameras are equipped each one with Nikon FC-E8 fisheye lens, with an angle of 183º. The valid colour images in the circle contain 6586205 pixels.

The tests have been carried out with twelve pairs of stereo images. We use two of them for computing the relevance of each attribute, from which the fuzzy densities can be obtained. At a second stage, we apply the CFI approach pixel by pixel for the remainder ten stereo pairs.

Our interest consists of determining the disparity of the trees trunks located in an area of 25 m^2 around the stereo vision system.

The disparity is the absolute difference value in sexagesimal degrees, taking into account the imaged circle, between the pixel in the left image and its matched pixel in the right one.

We have available the information of disparities 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 ten stereo images used for testing, we compute the disparity error for the trunks and then average these errors among the ten pairs of stereo images.

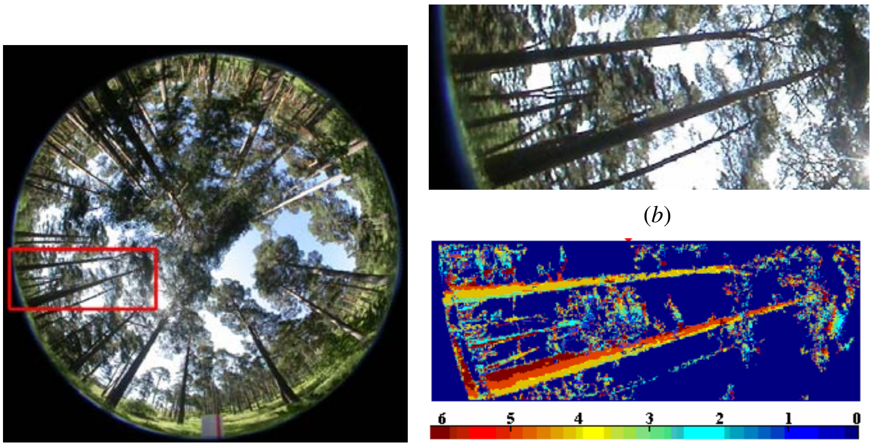
a) Computing the Relevance for Each Criterion

Given both available stereo images for this purpose, for each pixel in the left images, we compute the disparity with respect its matched pixel in the right ones, but considering each one of the six attributes separately through the equation (5). Each match is established according to the maximum similarity value computed for each attribute individually. But this does not imply that maximum similarity corresponds to a true match. Therefore, we need a mechanism for computing the relevance based on the rate of success or failure of each attribute. So, we compute the averaged percentage of error for both stereo images and for each attribute, based on the expert knowledge available about the disparities in the trunks. These probabilities are: p_a = 28 (correlation), p_b = 10 (colour), p_c = 14 (texture), p_d = 9 (gradient magnitude), p_e = 30 (gradient direction) and p_f = 27 (Laplacian). So, the individual relevancies are computed as $g^h = p_h \big/ \sum_k p_k$, h, k = a,b,c,d,e,f. Finally, these fuzzy values are exactly the following: $g^a = 0.150$, $g^b = 0.179$, $g^c = 0.187$, $g^d = 0.189$, $g^e = 0.145$ and $g^f = 0.152$. As one can see, the most relevant attribute is the gradient magnitude.

b) *CFI Performance*

As before, for each pixel in each one of the ten stereo images, available for testing, we obtain its disparity considering the six attributes separtley by applying the equation (5) and a maximum similarity criterion among the m candidates and also by applying the IF approach based on maximum supports, equation (8).

Figures 3(*a*) and 3(*b*) are the same that Figures 1(*a*) and 1(*b*) respectively. Figure 3(*c*) displays the disparity map obtained by the CFI approach in the area. The colour bar shows the disparity level values according to the colour.



(*a*) (*b*) (*c*)

Fig. 3. (*a*) Left image; (*b*) expanded area; (*c*) disparity map obtained by the CFI approach

Table 1 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 CFI approach is also displayed. For comparative pouposes we have tested the performance of our approach against the decision making approach proposed by Yager [10] based on fuzzy sets aggregation. The combination is made two to two similarity measures accordign to the following expression,

$$\mu_i(l) = 1 - min\left\{1, \left(\left(1 - s_{ih}\right)^p + \left(1 - s_{ik}\right)^p\right)^{\frac{1}{p}}\right\} \quad p \geq 1 \tag{9}$$

where h and k denote two similarity measures. Then, by applying the associative property of this aggregation operator we compute a final support for the six similaty values. The parameter p is estimated from the two stereo pairs used for computing the relevances of each attribute. Indeed, we vary p from 1 to 4, which is the range generally used, and compute the percentage of error, obtaining the best results with p set to 2.0.

Table 1. Averaged percentage of errors and standard deviations obtained through maximum similarity criteria for each attribute separately and the CFI decision making approach

Averaged percentage of error and standard deviations															
s_a		s_b		s_c		s_d		s_e		s_f		YAG		CFI	
%	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	13	1.9	**11**	**1.3**

From results in table 1, one can see that both combined strategies, Yager and CFI outperforms the individual similarity based approaches. This means that the combination of similarities between attributes improve the results obtained by using similarities separately. The CFI approach obtains better results than the Yager's one.

The best individual similarity results are obtained through the similarities provided by the gradient magnitude attribute (s_d). This implies that it is the most relevant attribute. This agrees with its relevance obtained previosly, as it has turned out to be the most relevant attribute.

4 Concluding Remarks

In this paper we have proposed a method for stereovision matching, in omni-directional images, in a system equipped with fish-eye lenses. The method applies three well-known constraints (*epipolar*, *similarity* and *uniqueness*) by combining area-based and feature-based matching strategies, which are classical constraints used in conventional stereovision systems.

For each pixel in the left image, a list of possible candidates in the right image is obtained for determining its correspondence.

The similarity between attributes establishes measures for the matching between the pixel and its candidates. Each candidate receives a support, which establishes the degree of similarity, consequently of correspondence between it and the pixel in the left image.

Under the CFI paradigm, we combine the similarities between six attributes and make a decision for choosing the unique candidate, if any, for the given pixel in the left image. The proposed combined strategy outperforms the methods that use similarities separately and it is compared favorably.

Although the results achieved can be considered satisfactory, they could be improved by applying additional constraints such as *smoothness* or *ordering*, which have been used for matching in conventional stereovision systems.

Another issue still open in future works is that concerning with the correspondence between pixels out of the trunks. This will allows discarding all pixels belonging to the background and also those pixels which belong to the leaves or the sky.

Moreover, the disparity map can be still refined by applying smoothing techniques, perhaps optimization ones, such as simulated annealing or Hopfield neural networks, where unmatched pixels in the trunks can be removed for obtaining a disparity map without gaps or another undesired artifacts.

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 first author. Also to Dra. Isabel Cañellas and Fernando Montes from the Forest Research Centre (CIFOR, INIA) for his support and the material supplied.

References

1. Barnard, S., Fishler, M.: Computational Stereo. ACM Computing Surveys 14, 553–572 (1982)
2. Cochran, S.D., Medioni, G.: 3-D Surface Description from binocular stereo. IEEE Trans. Pattern Anal. Machine Intell. 14(10), 981–994 (1992)
3. Tang, L., Wu, C., Chen, Z.: Image dense matching based on region growth with adaptive window. Pattern Recognit. Letters 23, 1169–1178 (2002)
4. Lew, M.S., Huang, T.S., Wong, K.: Learning and feature selection in stereo matching. IEEE Trans. Pattern Anal. Machine Intell. 16, 869–881 (1994)
5. Abraham, S., Förstner, W.: Fish-eye-stero calibration and epipolar rectification. Photogrammetry and Remote Sensing 59, 278–288 (2005)
6. Schwalbe, E.: Geometric Modelling and Calibration of Fisheye Lens Camera Systems. In: Proc. 2nd Panoramic Photogrammetry Workshop, Int. Archives of Photogrammetry and Remote Sensing, vol. 36, Part 5/W8 (2005)
7. Barnea, D.I., Silverman, H.F.: A Class of Algorithms for Fast Digital Image Registration. IEEE Trans. Computers 21, 179–186 (1972)
8. Pajares, G., de la Cruz, J.M.: Visión por Computador: Imágenes digitales y aplicaciones, RA-MA (2008)
9. Kuncheva, L.: Combining Pattern Classifiers: Methods and Algorithms. Wiley, Chichester (2004)
10. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. System Man and Cybernetics 18(1) (1988)