ABSTRACT
This paper presents a multi-scale stereo correspondence algorithm that incorporates different modules corresponding to every stage. A correspondence method is proposed based on minimizing an energy function using simulated annealing. Multiresolution techniques are also introduced to improve both response time and the quality of results. A technique of calculating the distance between two colours is defined for comparing colour features in HLS model. This measure not only highlights the difference in the H component but also considers the differences in S and L values. A metrics for evaluating the quality of a disparity map is defined. Experimental results with real images are presented to illustrate the approach taken.

KEY WORDS
Multiresolution, energy function, colour model, disparity, interpolation

1 Introduction
Because of the position and function of eyes in the human body, our brains receive two very similar images of a scene, taken from two separate points, situated extremely close together on a horizontal line. Two objects, at different distances from the observer, present different relative positions in their retinal images. The brain can measure this displacement (retinal disparity) and use it for estimating depth [1]. The disparity of every point in the image forms the so-called disparity map.

Stereovision is the name given to the set of techniques used to extract 3D information from a scene of two or more images taken from different viewpoints. A stereo system must solve several problems:

- Camera calibration, that is, estimating the values of the intrinsic and extrinsic parameters of the camera model.
- Rectification of the epipolar geometry to simplify the search on a scan line.
- Correspondence between tokens of the images. This is the problem of determining which elements in the left image correspond to those in the right image. Our three-dimensional perception of the world is due to our ability to interpret the differences in retinal location between corresponding objects. This problem is considered the most difficult part of the stereo problem.
- Reconstruction of the 3D scene, that is, the calculation of the depth from the disparity.

1.1 The Correspondence Problem
This problem can be seen as a search problem: given an element in the left image, the corresponding element in the right image must be found. From a human point of view, the process of stereovision is so natural that its complexity is not appreciated until attempts are made to automate the process. Uniqueness, line smoothness and epipolar geometry are some of the physical constraints that are imposed in the stereo problem to restrict the space of solutions and to obtain a well-posed problem.

The correspondence algorithms are classified into two classes: correlation-based and feature-based methods [2]. Region-based correspondence is included within the family of feature-based methods. Generally, the higher the semantic level of the primitive, the more robust the matchings. Nevertheless, there are important disadvantages: the primitive extraction is more difficult and the disparity map is sparser. In paper [3] an interesting review of region-based correspondence is shown.

Some authors have taken into account the idea of incorporating multiresolution schemes to detect stereo correspondence with the objective of obtaining and estimating the depth of a scene [4]. Multi-scale or coarse-to-fine scheme is a method of efficiently and effectively representing data with the objective of reducing the computational complexity. A multi-scale approach has been followed in this paper.

Stochastic relaxation methods have received a lot of attention in the field of stereovision. These techniques try to obtain an optimal solution to the correspondence problem without falling in a local minimum.
In artificial vision the colour has been used in segmentation processes [5], image classification, image restoration of a database, etc. The use of colour has always been determined by its three-dimensional composition. Its representation has been studied in [6], emphasizing the subjects of construction of perceptual spaces that allow vision methods to be applied by computer.

In this work an approximation that integrates different modules involved in stereo (feature extraction, matching and interpolation) is formulated. An energy function is built for each module at different resolutions and it is minimized in an integrated way yielding to a dense disparity map.

In sections 2 and 3, a new technique for measuring distance between two colours is defined, an energy function is described to formulate the correspondence problem, the proposed algorithm to minimize the energy function and to obtain the disparity map is described and the applied multiresolution scheme is indicated. Finally, some experiments and results are shown.

2 Colour Model

Colour model can affect correspondence processes in an important way. RGB model distributes intensity in three channels, which causes a serious colour constancy problem, that is, colour variations due to either light orientation or to different light sources. Two close colours in the RGB space can result in very different colours from a perceptual point of view. A simple way of obtaining the colour constancy is changing the image to HLS space. It is a perceptual space, that is, colour characteristics match their perceptual or psychological representation.

2.1 HLS space

The HLS space is a colour model that formalizes the system developed by Munsell. Its design reflects the manner in which humans see colours, so it offers many advantages in image processing. This space is composed of three components: hue, lightness and saturation. Hue (H) represents the colour shade and it is expressed as an angle varying from 0 to \( \frac{\pi}{2} \) radians. Saturation (S) represents the similarity of a colour to a neutral grey, and lightness (L) denotes the degree of lightness or darkness. Figure 1 compares the representation of RGB and HLS models.

![Colour models](image)

**Fig. 1.** Colour models

Comparing colours in the HLS Space

Colour images imply solving the problem of distance between two colours. Every colour is defined by three values, one of these in polar coordinates. Using linear and angular values implies merging two types of data in a single value. We propose a measure that not only highlights the difference in chromaticity but also considers the differences in lightness and saturation. In (1) the angular distance to calculate the difference between the H components is presented. The distance used for the L and S components is obtained using (2).

\[
D_H(p,q) = \begin{cases} 
\|H_p - H_q\| & \text{if } |H_p - H_q| \leq 180 \\
180 - |H_p - H_q| & \text{if } |H_p - H_q| > 180.
\end{cases} \quad (1)
\]

\[
D_{LS}(p,q) = \sqrt{(L_p - L_q)^2 + (S_p - S_q)^2}. \quad (2)
\]

Where \((H_p, L_p, S_p)\) are the values of pixel \(p\) and \((H_q, L_q, S_q)\) are the values of pixel \(q\).

Some additional considerations must be observed: if the S values are low the chromaticity does not contribute important information, furthermore a high difference may not imply that the colours are visually different. So an S threshold below which the H value is not considered is defined.

As distances in L and S are Euclidian in a range of [0 to 1] and the distance in H is angular in a range of [0 to 360], a normalized constant \((c)\) is used to equilibrate the magnitudes.

When the values of the H components are high enough only the distance in H is considered, however, if they are below an H threshold the difference between L and S is incorporated into the calculations.

Table 1 shows the proposed way of calculating the distance between two colours, combining the distances between the components. The values used are: \(c=10\), S-threshold=0.3, H-threshold=25.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>If (S_p &lt; S\text{-threshold}) and (S_q &lt; S\text{-threshold})</td>
<td>(\text{Dist}(p,q) := c \cdot D_{LS}(p,q))</td>
</tr>
<tr>
<td>Else if (\text{dif}H &gt; \text{H-threshold})</td>
<td>(\text{Dist}(p,q) := \text{dif}H)</td>
</tr>
<tr>
<td>Else if (\text{dif}H &lt; \text{H-threshold})</td>
<td>(\text{Dist}(p,q) := c \cdot D_{LS}(p,q) \cdot \text{dif}H)</td>
</tr>
<tr>
<td>Else</td>
<td>(\text{Dist}(p,q) := \text{dif}H)</td>
</tr>
</tbody>
</table>

3 Correspondence Method

A correspondence method is proposed based on minimizing an energy function using simulated annealing. Multiresolution techniques are also introduced to improve the response time and the quality of the results. The algorithm
is as follows: firstly, original images are scaled, then a disparity map is obtained through minimizing an energy function, the resulting map is then interpolated to the next level of resolution. The process is repeated until the original scale level is reached.

3.1 Energy Function

The energy function minimized by the simulated annealing algorithm is composed of five terms, each one weighted by a control parameter \( \gamma_n \). Equation (3) defines this function.

\[
U(p) = \sum_{n=1}^{5} \gamma_n U_n(p).
\]  

(3)

The used notation is shown in Table 2.

| \( II \) | Left image |
| \( ID \) | Right image |
| \( (p_x, p_y) \) | Coordinates of pixel |
| \( N(p) \) | Neighbourhood of \( p \) |
| \( vL(p), vR(p) \) | Vertical edges of the images, it contains a value 1 if there is an edge between the pixel \( (p_x, p_y) \) and the pixel \( (p_x, p_y-1) \), and 0 otherwise |
| \( hL(p), hR(p) \) | Horizontal edges of the images, it contains a value 1 if there is an edge between the pixel \( (p_x, p_y) \) and the pixel \( (p_x-1, p_y) \) and 0 otherwise |
| \( disp \) | The disparity map |
| \( \delta(a, b) \) | Function returning the value 1 if \( a = b \) and 0 otherwise |
| \( Dist(p,q) \) | Distance between two colours |
| \( \Xi(p,N(p)) \) | Census transform of \( p \) |
| \( H(v1, v2) \) | Hamming distance between two vectors of bits |

The term \( U_1 \) is the colour-level matching at the selected pixel. Instead of comparing one pixel in the left image with just one pixel in the right image, a neighbourhood around pixel \( p \) is considered.

\[
U_1(p) = \sum_{N(p)} \text{Dist}(p, q)| (p \in II) \land (q \in ID) \land (p_x = q_x) \land (q_y = p_y + \text{disp}(p)).
\]  

(4)

In previous investigations with images without colour information, the square difference between intensity values [7] and the absolute value of the difference between intensity values [8] have been used. It has been noticed that use of the latter improves the results in the presence of atypical values (outliers).

The term \( U_2 \) is the transform Census matching cost [9]. The value of the transform is calculated comparing the colour value of a pixel with the value of the neighbourhood. The transform must give similar results near corresponding points between the two images.

\[
U_2(p) = \sum_{N(p)} |H(\Xi(p, N(p)), \Xi(q, N(q)))| 
\]

\[
(\text{p} \in II) \land (\text{q} \in ID) \land (p_x = q_x) \land (q_y = p_y + \text{disp}(p)).
\]  

(5)

The term \( U_3 \) is the edge-level matching cost. We use information from both vertical and horizontal edges. It can be supposed that if an edge exists in the left image, then there should also be another edge in the right image corresponding to the first one, but in a position displaced by the amount of pixels determined by the disparity.

\[
U_3(p) = \sum_{p \in II} |(1 - \delta(vL(p), vR(q))) + (1 - \delta(hL(p), hR(q)))|
\]

\[
(\text{p} \in II) \land (\text{q} \in ID) \land (p_x = q_x) \land (q_y = p_y + \text{disp}(p)).
\]  

(6)

The term \( U_4 \) is the smoothing constraint: it is assumed that disparity varies smoothly between edges. This term switches off smoothing whenever a line field is encountered in the image. Clearly, the disparity value at two locations cannot be similar when there is an edge between them.

\[
U_4(p) = (\text{disp}(p) - \text{disp}(q))^2 * (1 - vL(p)) + (\text{disp}(p) - \text{disp}(q))^2 * (1 - vR(p))
\]

\[
(p_x = q_x) \land (q_y = p_y + 1).
\]  

(7)

The final term incorporates the uniqueness constraint. It means that along any row \( i \), by calculating the disparity at the \( f \) and the \( q \) column then according to the uniqueness constraint, \( f + \text{disp}(i,j) \neq q + \text{disp}(i,q) \).

\[
U_5(p) = \sum_{q=0}^{\infty} \delta(p_x + \text{disp}(p), q_y + \text{disp}(q)) | (p_x = q_x).
\]  

(8)

3.2 Simulated Annealing

A well-known stochastic relaxation method, called simulated annealing (SA) has been used to obtain the global or nearly global optimum solution depending on the annealing schedule. The algorithm tries to minimize an energy function that incorporates a similitude error measure between corresponding points. In this paper, we adopt the Metropolis algorithm [10]. Each pixel is visited and the value of the disparity within a given range is updated. Considering the energy function previously defined, the algorithm is applied iteratively. The algorithm is described more precisely in [8].
3.3 Multiresolution Scheme

A multiresolution structure is proposed that will speed up the process that initialises the correspondence problem with a solution coming from a lower resolution. In this work sampling methods for scaling are applied: some pixels are selected from the original image. The original images are scaled down by the power of two. The process consists of selecting the first row and deleting the next 2n-1 rows, and so on. The process concerning the columns is analogous.

Interpolation

A multiresolution method requires an interpolation technique to allow the use of results from one level of the scale in the following one. In this work we use linear interpolation. Starting at the coarsest resolution and obtaining the initial disparity map at that level, the optimal solution is found quickly due to the limited number of elements in the disparity range. At the following resolution level a new disparity map is generated by interpolating values of pixels obtained from the previous map. Disparity obtained at the previous level is assigned to the corresponding points at the current level, so only some points will have a value. Next, gaps must be filled in using linear interpolation. This map is used as the initial configuration. The algorithm is applied to obtain a disparity map at that resolution. The process is then repeated until level 0 (original image) is reached.

The disparity estimation at each level is median filtered. The median filtering step is suitable to correct outlier disparity estimations that deviate from the correct expected estimation (a form of smoothness constraint on the estimations).

4 Experimental results

Here some results using real images are presented. It is important to remark that raw images were used, without any type of pre-process. The images were taken with a DCAM stereo camera at a resolution of 320 x 240 pixels. In this section we present two of the experiments. We have limited the search space to a reduced interval.

Quality of the disparity map

The capacity to validate the obtained results in an objective manner is a problem that must be resolved by means of metrics. This metrics must allow the measurement of the difference between two disparity maps: the obtained result and the ideal result. Some authors [11] use the real disparity map. Nevertheless, our aim is navigation so the most important information for us is the distance between objects.

To evaluate the quality of the results we proceed as follows. When the images are taken, the distance to the camera from each relevant object is measured. A segmentation algorithm is applied to the disparity map obtained and the distance to the objects calculated using the relation between disparity and depth shown in eq. (9), and so a comparison can be made between the obtained distance and the manual measure.

\[ Z = f \frac{T}{d}, \quad (9) \]

Where, \( Z \) represents the distance from the camera to the object, \( f \) the camera focal length, \( T \) the base line and \( d \) the disparity obtained from the correspondence algorithm.

The quality measure that we present, \( \Delta_{\text{disp}} \), gives the error made when the disparity map is calculated. Once the distance from the camera to the relevant objects in the scene is obtained, eq. (10) is calculated.

\[ \Delta_{\text{disp}} (Z, R) = \frac{\sum_{i=1}^{n} | Z_i - R_i |}{n}. \quad (10) \]

Where, \( Z_i \) represents the obtained distance for region \( i \) applying eq. (9), \( R \), the real distance from the object to the camera and \( n \) the number of objects under consideration in the scene.

Experiments

The method used in the experiments is described below. With real images the proposed method to estimate the disparity has been applied. Next, to show the improvement obtained using the colour, the images have been transformed to grey level and the disparity map is again obtained. In this case, the grey level has been used as a first term in the energy function. Finally, the quality of the disparity maps obtained using the metrics \( \Delta_{\text{disp}} (Z, R) \) is analysed.

To find the most appropriate weights of the parameters of the energy function, the method has been tested with different values: 0, 1, 20, 50 and 100. The best results are obtained with: \( \gamma_1=20, \gamma_2=50, \gamma_3=1, \gamma_4=50 \) and \( \gamma_5=1 \).

Figure 2 shows the first example where images are scaled by a factor of 23 before our algorithm is applied. The search space is 80 pixels. Figure 3 compares the resulting disparity maps without using colour information and the result calculated applying colour information. \( \Delta_{\text{disp}} (Z, R) \approx 2.89 \) for grey level and 0.06 for colour images.

Figure 4 shows another example of an indoor scene. Figure 5 presents a further comparison between the method without using colour information and the result calculated applying colour information. This example is particularly interesting because the object in the foreground obscures part of the more distant object. In the disparity map calculated without colour information, a region is noticed, whereas, when working with colour information two regions appear. The search space is 80 pixels. In the disparity map generated by the method without using colour information the nearer object is not located, for this reason \( \Delta_{\text{disp}} (Z, R) = \infty, \Delta_{\text{disp}} (Z, R) = 0.04 \) for colour images.
Although initially it may seem that colour information could slow down the process, this situation does not occur due to the fact that the algorithm needs less iterations to obtain the disparity map. The execution time for all the experiments is only 2.7 seconds (colour), 8.5 (grey) with the time measured using a Pentium 2.40 Gz processor. Moreover, the disparity map obtained with colour information is more suitable for our proposals as it is tested comparing the metrics values for grey and colour algorithms (2.89 and 0.06 respectively in experiment 1; and $\infty$ for grey and 0.04 for colour in experiment 2).

5 Conclusions

In this paper we have described a matching algorithm which is suitable for producing dense depth maps from stereo image pairs. The algorithm is multiresolution so less iterations are needed to obtain a solution. The following conclusions can be highlighted:

- A measure for comparing two colours in HLS space is developed, that highlights the difference in chromaticity but considers the differences in lightness and saturation too.
- The colour information speeds up the process in addition to improving the quality of the obtained results.
- The colour information improves the quality of the result: the obtained depth (distances to objects) is more precise.
- Colour features, non-parametric transforms, edges, smoothness and uniqueness for building a robust energy function are incorporated.
- The information obtained in previous executions of the algorithm is employed for enhancing the initial configuration for the next iteration.

The experiments in figures 2 to 5 show that colour improves the results.