Stereovision depth analysis by two-dimensional motion charge memories

José M. López-Valles\textsuperscript{a}, Miguel A. Fernández\textsuperscript{b}, and Antonio Fernández-Caballero\textsuperscript{b,*}

\textsuperscript{a} Departamento de Telecomunicación, Escuela Universitaria Politécnica de Cuenca, Universidad de Castilla-La Mancha, 13071 - Cuenca, Spain

\textsuperscript{b} Departamento de Informática, Escuela Politécnica Superior de Albacete, Universidad de Castilla-La Mancha, 02071 – Albacete, Spain

Abstract

Up to date several strategies of how to retrieve depth information from a sequence of images have been described. In this paper a method that turns around the symbiosis existing between stereovision and motion is introduced. The central idea behind our approach is to transpose the spatially-defined problem of disparity estimation into the spatial-temporal domain. Motion is analyzed in the original sequences by means of the so-called permanency effect and from the resulting two-dimensional motion charge maps the disparities are calculated. This is an important contribution to the traditional stereovision depth analysis, where disparity is gotten from the image luminescence. In our approach, disparity is studied from a motion-based persistency charge measure.

Keywords: Stereovision; Permanency memories; Disparity Analysis; Depth Analysis

\textsuperscript{*} Corresponding author. Tel.: +34-967-599200; fax: +34-967-599224.

E-mail address: caballer@info-ab.uclm.es (A. Fernández-Caballero).
1. Introduction

In general there are several strategies of how to retrieve depth information from a sequence of images, like depth from motion, depth from shading and depth from stereovision. In this paper we introduce a new method to retrieve depth based on motion and stereovision. In a conventional stereoscopic approach, usually two cameras are mounted with a horizontal distance between them. As a consequence, objects displaced in depth from the fixation point are projected onto image regions which are shifted with respect to the image center. Due to the geometry of the optic system, it is sufficient to restrict disparity analysis to the projection of corresponding linear segments in the left and right camera. In some approaches, the disparity is computed by searching the maximum of the cross-correlation between image windows along the epipolar lines of the left and right image (Haralick & Shapiro, 1992). This is called the epipolar constraint, which means that for any point in the left image, its matching point in the right image must lie on the corresponding epipolar line.

So far, many algorithms have been developed to analyze the depth in a scene. Brown et al. (2003) describe a good approximation to all of them in their survey article. In many previous works, a series of restrictions are used to approach the correspondence problem. The most usual restriction is the disparity restriction, which considers that is not probable that there exist objects very close to the camera. The scene uses to be limited to a medium distance. This way, too high disparities are eliminated (Sumi et al., 2003). Koenderink and van Doorn (1976) expressed the necessary theory in best initial works related to disparity restriction, and Wildes (1991) implemented some of the ideas of (Wilson & Knutsson, 1989). More recently, disparity in stereoscopy continues showing its great interest (e.g., (Muhlmann et al., 2002; Gutiérrez & Marroquin, 2004)).

According to the correspondence techniques used, we may classify methods into correlation-based, relaxation-based (Grimson, 1985), gradient-based (Choi et al., 2003), and feature-based (Venkateswar & Chellappa, 1995). The main correlation-based technique is the area correlation technique (e.g., (Zabih & Woodfill, 1994)). Area based approaches have the advantage of directly generating dense disparity maps. Matching elements for area-based methods are the individual pixels over which the matching cost
is evaluated; pixel-to-pixel correspondence is evaluated on image intensity function and similarity statistics. For example, a work (Binaghi et al., 2004) investigates the potential of neural adaptive learning to solve the correspondence problem within a two-frame adaptive area matching approach. The method is based on the use of the zero mean normalized cross-correlation coefficient integrated within a neural network model which uses a least-mean-square delta rule for training. Another approach (Di Stefano et al., 2004) proposes an area-based stereo algorithm suitable to real time applications, where the core of the algorithm relies on the uniqueness constraint and on a matching process that rejects previous matches as soon as more reliable ones are found.

In this paper an area-correlation-based method that turns around the symbiosis existing between stereovision and motion is introduced; motion minimizes correspondence ambiguities, and stereovision enhances motion information. This symbiosis has been studied to get a major performance in our artificial three-dimensional disparity depth analysis of moving non-rigid objects through stereovision. Most methods have as a common denominator that they work with static images and not with motion information, although some approaches have been introduced so far (Ho & Pong, 1996; Liu & Skerjane, 1993; Xu, 1995). More recently, in (Zhang et al., 2003) the traditional binocular stereo problem is extended into the spacetime domain, in which a pair of video streams is matched simultaneously instead of matching pairs of images frame by frame. By utilizing both spatial and temporal appearance variation, this modification reduces ambiguity and increases accuracy. Also recently, it has been shown that methods derived from the spacetime stereo framework can be used to recover depth in situations in which existing methods perform poorly (Davis et al., 2005). Also, a separate edge-preserving regularization scheme to calculate disparity fields for a stereoscopic image pair and a joint disparity and motion estimation algorithm for stereoscopic video sequences has been presented (Yang et al., 2005).

The motion analysis algorithm used in this work has already been tested in applications such as moving object shape recognition in noisy environments (Fernández-Caballero et al., 2003), moving objects classification by motion features such as velocity or acceleration (Fernández et al., 2003), and in applications related to
selective visual attention (Fernández-Caballero et al., 2004; López et al., 2006). Our proposal is to analyze motion in the original sequences by means of the so-called permanency effect (Fernández et al., 2003) and from the resulting charge maps to analyze the disparities. As a novelty, in this paper motion analysis performs separately on both stereovision sequences. Thus, the central idea behind our approach is to transpose the spatially-defined problem of disparity estimation into the temporal domain and compute the disparity simultaneously with the incoming data. This is an important contribution to the traditional disparity analysis, where disparity is gotten from the image luminescence. In our approach, disparity is studied from a timely consistent persistency charge measure.

The rest of the paper is structured as follows. Our Stereovision Depth Analysis method is described in Section 2. In Section 3, experimental results on a real video sequence are described. Finally Conclusions are given in Section 4.

2. Description of the Stereovision Depth Analysis

The computational structure which supports the Stereovision Depth Analysis can be seen in Fig. 1 and is described next. This structure is the result of the analysis of the stereovision’s geometric problem and the application of the pertinent restrictions, as well as the study of biological stereovision systems and the permanence mechanisms, which our research group is very familiar with. Next, the proposed structure is described in a general way, dedicating the rest of the section to the detailed description of each subtask in which the solution to the problem is broken down.
The input to our system is a pair of stereo image sequences. These sequences have been acquired by means of two cameras arranged in a parallel configuration. In a well-calibrated fronto-parallel camera arrangement the epipolar lines are horizontal and thereby identical to the camera scan-lines. Thus, they will capture two similar, although not exactly equal, scenes. In case the images have been acquired in a convergent configuration, horizontal epipolar lines can be obtained by image rectification techniques (Faugeras, 1993).

Firstly (see Fig. 1), frame by frame, the permanence or accumulative computation effect (Fernández et al., 2003; Fernández-Caballero et al., 2003; López et al., 2006) is applied for the purpose of Charge Memories Calculation to the stereo image pairs. The input is the stereo image pair in grey levels, \( GL_L(u_L,v_L,t) \) and \( GL_R(u_R,v_R,t) \), corresponding to a frame and the output to each frame is the state of a couple of two-dimensional charge memories; that is, a 2D Charge Memory for the left image, \( C2_L(u_L,v_L,t) \), and another one for the right image, \( C2_R(u_R,v_R,t) \). Then Charge Disparity Analysis carries out the matching process between both right and left
sequence charge memories for each frame. The output is a depth memory per frame, $D3(u,v,t)$. Well-known concepts in stereovision will be applied in this step: restrictions to the correspondences, complete primitives, etc., except that in our case applied to the charge memories obtained in the previous section instead of object-shape information, as usual.

2.1. Charge Memories Calculation

The purpose of Charge Memories Calculation is to represent two-dimensional motion for every input sequence in the permanence elements’ charge levels. Fig. 2 shows the whole process as well as the interrelations between the sub-processes.

![Fig. 2. “Charge Memory Calculation” process outline](image-url)
Two parallel processes are observed; each one belongs to an input sequence, right and left. From there, the Grey Level Bands Segmentation subsystem separates each of the frames into related regions for the purpose of later analyzing its movements. The motion detection system, by means of permanency, requires as its input, the current image segmented into grey level bands, as well as the previous image. The purpose of this is to analyze which memory elements have skipped between the bands, detecting movement in the corresponding pixels.

2.1.1. Grey Level Bands Segmentation

The Grey Level Bands Segmentation subtask transforms the input images captured in 256 grey levels into a lower number of levels ($nGLB$). In general, using 8 levels produces good results (as demonstrated in previous works of the same authors for monocular sequences (e.g. (Fernández-Caballero et al., 2003)). These $nGLB$-level images are called grey level band segmented images ($GLB_{L/R}$)\(^1\), and they are calculated as

\[
GLB_{L/R}(u_{L/R}, v_{L/R}, t) = \left[ \frac{GL_{L/R}(u_{L/R}, v_{L/R}, t) \cdot nGLB}{GL_{\max} - GL_{\min} + 1} \right] + 0.5
\]

where $GL_{\max}$ and $GL_{\min}$ are the maximum and minimum grey level values, respectively, for the input image. Typically, $GL_{\max} = 255$ and $GL_{\min} = 0$, in all our implementations. In equation (1), the expression $GL_{\max} - GL_{\min} + 1$ corresponds to the number of input levels, which will generally be 256.

The two reasons why we prefer to work with grey level bands are: (a) A traditional movement detection system is based on image difference. By matching one grey level category in one single band, into considering that there is movement when a variation in the grey-level band is detected, the noise level is reduced due to small brightness variations in a single object between two consecutive images. (b) There is also a

\(^1\) Notice that throughout the whole text “L/R” is equivalent to “L or R”.

7
decrease in the computational complexity, noting the great parallelism used in the algorithms of the proposed model. We go on to parallel computing in the order of number of grey level bands $n_{GLB}$, and not grey levels $n_{GL}$, with $n_{GL}>n_{GLB}$.

Up to this moment, we have only performed a simple scale change. A detailed analysis of the features and performances of this segmentation method is not the purpose of this paper; a good description may be found in (Fernández-Caballero et al., 2003). Notice that we are not yet deciding whether there is variation or not in the $(u,v)$ point’s grey level band. To sum up, the result of subtask Grey Level Bands Segmentation is, for each input image pixel, the grey level transformation in its corresponding grey level band. That is, the result will be a matrix of the same size as the input image, but its content will have values between 1 and $n_{GLB}$.

2.1.2. Motion Detection

Once the grey level bands for each pixel in both images have been established, the next step is motion characterization. For this, we first establish a two-dimensional motion memory for each sequence, which will be updated for each frame. This two-dimensional motion memory will have as many motion detection elements as pixels in the horizontal and vertical dimensions of the images.

The Movement Presence, $M_{2L/R}(u_{L/R},v_{L/R},t)$, is obtained through pixel by pixel comparison of two consecutive images segmented into grey level bands. If pixel $(u_{L/R},v_{L/R})$ at a moment of time $t$ belongs to the same grey level band than at moment of time $t-1$, than there has been no movement; whereas, if there has been a change in $GLB_{L/R}$, then it is assumed that there has been movement.

$$M_{2L/R}(u_{L/R},v_{L/R},t) = \begin{cases} 
0, & \text{if } GLB_{L/R}(u_{L/R},v_{L/R},t) = GLB_{L/R}(u_{L/R},v_{L/R},t-1) \\
1, & \text{if } GLB_{L/R}(u_{L/R},v_{L/R},t) \neq GLB_{L/R}(u_{L/R},v_{L/R},t-1) 
\end{cases} \quad (2)$$

These two-dimensional motion presence charge memories identify those input image pixels where a jump between grey level bands has occurred and thus, the image points for coordinates $(u_{L/R},v_{L/R})$ where there has been movement.
With the purpose of obtaining more accurate information about movement, it is convenient to detect, not only the points where there has been movement, but also the movement’s more recent history. For this, we use the permanence charge memories. By means of accumulative computation mechanisms on the negation of the Motion Presence characteristic, this subtask obtains the 2D Motion Charge Memory associated with the accumulation process, as shown in equation (3):

\[ C_{2L/R}^2(u_{L/R}, v_{L/R}, t) = \begin{cases} C_{2\text{max}}, & \text{if } M_{2L/R}^2(u_{L/R}, v_{L/R}, t) = 1 \\ \max [C_{2\text{min}}, C_{2L/R}^2(u_{L/R}, v_{L/R}, t - 1) - C_{2\text{dis}}] & \text{if } M_{2L/R}^2(u_{L/R}, v_{L/R}, t) = 0 \end{cases} \]

(3)

\( C_{2\text{min}} \) and \( C_{2\text{max}} \) are the minimum and maximum values, respectively, that the values stored in the two-dimensional charge memory (2D Motion Charge Memory, \( C_{2L/R}^2(u_{L/R}, v_{L/R}, t) \)) can reach. \( C_{2\text{dis}} \) is the memory discharge value. When giving values to variable \( C_{2\text{dis}} \), we must consider the sampling-speed ratio of the sequences captured. The idea behind this process is that if no movement exists in point \((u_{L/R}, v_{L/R})\), that is to say, when \( M_{2L/R}^2(u_{L/R}, v_{L/R}, t) = 0 \), the charge value \( C_{2L/R}^2(u_{L/R}, v_{L/R}, t) \) will decrease until it reaches \( C_{2\text{min}} \). If movement does exist \( (M_{2L/R}^2(u_{L/R}, v_{L/R}, t) = 1) \), the complete charge takes place, taking the \( C_{2\text{max}} \) value. Thus, a point in which there has been recent movement will have charge values between saturation \( (C_{2\text{max}}) \) and complete discharge \( (C_{2\text{min}}) \), but different from them. The values will be closer to saturation, the more recent the movement is; inversely, the values will be lower, the longer it has been since movement took place in this area of the image. This way, the charge value is proportional to the time that has elapsed since the last significant brightness variation for each pixel of the image.

Fig. 3 shows all these issues. Fig. 3a and 3b show two images of a monocular sequence. The advance of a car may be appreciated, as well as a more slight movement of a pedestrian. In Fig. 3c you may observe the effect of these moving objects on the permanence memory drawn as a two-dimensional image.
Fig. 3. Motion charge Memory: (a) one image of a sequence, (b) same perspective after some seconds, (c) motion trails as represented on the two-dimensional motion charge map

The difference between a quick object as the car, which is leaving a very long motion trail (from dark grey to white), and a pedestrian whose velocity is clearly slower and whose motion trail is nearly unappreciable with respect to the car’s one, is presented. Thus, permanency memories enable representing the motion history of the frames that form the image sequence, and it is possible to segment from the motion of the objects present in the scene. For a more extended description of the permanence effect, see (López et al., 2006).

2.2. Charge Disparity Analysis

The output of Charge Disparity Analysis is a three-dimensional depth memory, $D3(u,v,t)$, which shows the depth of the points in the scene where there has been movement. Using the charge memories as input has two important advantages. In first place, only information about motion is used, filtering out all static information from the scene, whether it is 2D or 3D. Since our objective is to obtain a three-dimensional memory of the scene’s motion, it is more an advantage than a disadvantage to have a filtered memory, as static elements only contribute to noise to this project. On the other hand, object motion leaves similar charge trails in both permanence charge memories. Thus, matching of the moving objects trails in the sequence will be simpler and more robust.
In order to explain our disparity analysis method, it is sufficient to analyze the process at the level of epipolar lines. The key idea is that a moving object causes two identical trails to appear in epipolar lines of the permanency stereo-memories. The only difference relies in their relative positions, affected by the disparity of the object at each moment. In Fig. 4, a simple example is offered, where the charge values in two corresponding superimposed epipolar lines of the memories are represented. In a parallel configuration as the one we have chosen, there will be no disparity in right and left image for objects that are in a great depth - imagine in the infinite. Nevertheless, when an object approaches to the central point of the base line, that is to say, between the two cameras, the object goes appearing more to the right on the left image and more to the left on the right image. This is precisely the disparity concept; the more close objects have a greater disparity than the more distant ones.

So, looking at Fig. 4 it is possible to analyze the motion of each one of the three objects present in the permanency memories (or charge memories) from their motion trails. You may observe that object ”a”, which has a long trail and has its maximum charge towards the left, is advancing to the left at a high speed. Object ”b”, with a shorter trail, is also advancing towards the same direction but at a slower velocity. Finally, object ”c”, whose trail is inverted in horizontal, is moving to the right at a medium velocity, as shown by its trail.

Fig. 4. Disparity by motion charge (permanency) memories
This simple example draws two main conclusions. Firstly, in order to consider two
motion trails to be correspondent, it must only be checked that both are equal enough in
length and in discharge direction in epipolar lines of the permanency stereo-memories.
Secondly, we may affirm that, in order to analyze disparities, one possibility is to
displace one epipolar line over the other one, until we get the exact point where both
lines are completely superimposed. In other words, an epipolar line has to be displaced
over the other until motion trails maximally coincide. Of course, the right epipolar line
can be displaced over the left or the left epipolar line over the right. When the motion
trails coincide (with a given error ratio or threshold), the displacement value applied to
the epipolar line is precisely the searched disparity value.

But, objects usually are not uniform and the layout of the permanence memories is
much more complicated. This is why an object seen as the set of parts that compose it
shows various disparities. Thus, it is necessary to analyze the correspondence from the
values of the various parts of the objects to obtain one valid overall disparity value. The
more efficient way to manage this is that each pixel obtains its disparity in such a way
that the maximum of its neighboring charge values confirm a consensus disparity.

All these considerations tell us that the disparity analysis at epipolar line level
consists in superimposing both epipolar lines with different relative displacements and
in analyzing the correspondences produced in the neighborhood of each charge unit.
The one displacement, which produces that a maximum number of surrounding
elements confirm its correspondence, is assigned the more trustful disparity value.
Precisely, in Fig. 5, the process diagram for this subtask is shown. In it, we can also see
the interrelation between the processes Charge Correspondence Analysis and Depth
Memory Calculation, as well as the inputs and outputs involved in each process. The
correspondence analysis output will be called 3D Charge Correspondence Memory,
$S_3(u,v,d,t)$, where $d$ stands for disparity. From this charge correspondence memory, the
maximum reliability depth for each coordinate $(u,v)$ will be decided. This will be done
by means of the 3D Depth Memory Calculation process. The output will be called 3D
Depth Memory, $D_3(u,v,t)$. 
According to the taxonomy proposed by Scharstein and Szelinsky (2002), a dense stereo matching process, as the one performed in our proposal, can be divided into three tasks: matching cost computation, aggregation of local evidence and computation of disparity values. Local methods usually compute final disparity adopting a local winner-take-all strategy which selects the pair with the best matching cost under assumption of uniqueness. This is also the overall approach taken in our work. Charge Correspondence Analysis performs the matching cost computation and the aggregation of local evidence, whereas Depth Memory Calculation computes the disparity values.

2.2.1. Charge Correspondence Analysis

The purpose of this subtask is to prepare the necessary information to decide on the disparity with the greatest reliability for each of the processing elements in the input charge memory for 3D Depth Memory Calculation. This task will take into account, mainly, the epipolar, the ordering and the disparity restrictions.
It is common knowledge that the most robust correspondence primitives are those with the highest contrast, such as contours or regions. In our case, we intend to carry out a correspondence analysis per regions (area-based disparity calculation), called *constant disparity* regions. Therefore, we must group together those neighboring charge elements whose corresponding elements have the same (or very similar) disparity. But before grouping the neighboring elements together, remember what we call corresponding charge elements. It is, basically, a question of finding which pixel has a similar history of movement in the opposite epipolar line and, consequently which processing element for the corresponding charge memory has an instant charge level stored with a similar value.

Now, the reliability criterion chosen will depend on the position of each processing element and it has to do with the size of each *constant disparity* region. This size is calculated in two phases: (a) First of all, we carry out a horizontal counting of all adjoining neighbors that belong to this region. (b) Afterwards, the vertical values found for all the adjoining vertical processing elements, which also belong to this region, are accumulated. In the following subsections, all related phases, that is to say, *Pixel-wise Charge Correspondence Analysis, Horizontal Charge Counting and Homogenizing*, and *Vertical Charge Accumulation and Homogenizing*, will be explained.

**a) Pixel-wise Correspondence Analysis**

The *Pixel-wise Charge Correspondence Analysis* on the charge elements of both corresponding 2D Motion Charge Memories, $C2_{L/R}(u_{L/R},v_{L/R},t)$ is carried out. By applying the epipolar restriction, each charge element from a charge memory is compared to those from the other charge memory on the same row, although displaced horizontally up to the maximum limit $d_{\text{max}}$, imposed by the disparity restriction. The calculation expression for the elements of this three-dimensional output matrix is as follows:
The three-dimensional matrix \( Sa3(u,v,d,t) \) indicates whether or not there is a specific correspondence (similarity in charge above threshold value \( 3\theta_a \)) between both charge memories for each coordinate \((u,v)\) and for each disparity value \(d\).

**b) Horizontal Charge Counting and Homogenizing**

The purpose in this second step is to establish a charge elements’ matrix the same size as the three-dimensional matrix \( Sa3 \) from the previous step, where each element would have the amount of horizontally corresponding adjacent input elements as the final result (those where \( Sa3(u,v,d,t) = 1 \)). This process is carried out in two periods: the first (Horizontal Charge Counting) running to the left counts the input elements set on 1 and stores the value in the corresponding output charge element \((Sb3(u,v,d,t))\).

\[
\text{for } u = 1 \text{ to image \_ width} \\
Sb3(u,v,d,t) = \begin{cases} 
Sb3(u-1,v,d,t) + 1, & \text{if } Sa3(u,v,d,t) = 1 \\
0, & \text{if } Sa3(u,v,d,t) = 0 
\end{cases}
\]

Once the horizontal counting to the right is done, a charge homogenizing is carried out (Horizontal Charge Homogenizing) so that all charge elements belonging to a horizontal constant disparity region acquire the same charge value. This value corresponds to the horizontal size of the constant disparity region, formed by all of them.

\[
\text{for } u = \text{image \_ width} - 1 \text{ downto } 1 \\
Sc3(u,v,d,t) = \begin{cases} 
\max[Sc3(u+1,v,d,t), Sc3(u,v,d,t)], & \text{if } Sb3(u,v,d,t) = 1 \\
0, & \text{if } Sb3(u,v,d,t) = 0 
\end{cases}
\]
c) Vertical Charge Accumulation and Homogenizing

On this third step, we aim to establish a new charge element matrix of the same size as the three-dimensional matrix $Sc3$ from the previous step, where each element has, as final result, the charge accumulation from the neighboring input elements, which are considered vertically correspondent. This process is done, as in the previous case, in two periods: a first running downwards (rising $v$ values) counts the input elements other than 0 and stores the accumulated value in the corresponding output charge element ($Sd3(u,v,d,t)$).

$$\text{for } v = 2 \text{ to image}_\text{height}$$

$Sd3(u,v,d,t) = \begin{cases} 
Sd3(u,v,d,t) + Sd3(u,v-1,d,t), & \text{if } Sc3(u,v,d,t) = 1 \\
0, & \text{if } Sc3(u,v,d,t) = 0 
\end{cases}$ \hspace{1cm} (7)

Once the counting towards positive $v$ values is done, charge homogenizing is carried out in such a way that all charge elements belonging to a constant disparity region have vertically the same charge value. This value will correspond to a region’s total size.

$$\text{for } v = \text{image}_\text{height} - 1 \text{ downto } 1$$

$Se3(u,v,d,t) = \begin{cases} 
\max \left[ Se3(u,v,d,t), Se3(u,v+1,d,t) \right], & \text{if } Sd3(u,v,d,t) = 1 \\
0, & \text{if } Sd3(u,v,d,t) = 0 
\end{cases}$ \hspace{1cm} (8)

2.2.2. 3D Depth Memory Calculation

Once we have calculated the region’s sizes which each charge memory correspondence belongs to, we need to associate, as maximum reliability disparity for each pixel, those values whose charge $S3(u,v,d,t)$ is maximal in $d$. With this, we are imposing the uniqueness restriction, since each processing element will only have a single disparity value as a final value.

The processing carried out to obtain the disparity associated to each charge element is shown in the following expression:
This operation calculates the value $i$ whose $S_3(u,v,i,t)$ is maximum in the third dimension. The ordering restriction is included in the method proposed for the charge disparity analysis subtask, since the specific correspondence verification and subsequent region configuration means maintaining the order of the correspondences found. Based on the charge disparity calculation done, we can estimate the moving elements’ depth. With this, we have obtained a stereoscopic motion memory in which each moving element from the scene appears associated to its depth.

3. Data and Results

In order to test our algorithms, the results of applying our algorithms are shown on a real stereo sequence, called "OutdoorZoom" (see Fig. 6), downloaded from “labvisione.deis.unibo.it/ smattoccia/stereo.htm”. The whole sequence is 30 seconds long and has been acquired at a rate of 10 images per second. This is precisely a stereo sequence created and used by the authors of (Di Stefano et al., 2004) within a research activity aimed at developing a 3D People Tracking application. The tracking approach is based on first merging disparity maps with the information provided by a grayscale change-detection algorithm and then building a suitable plane-view representation that enables to track moving objects in the 3D space.

![Fig. 6. Frame # 211 from sequence “OutdoorZoom”. (a) Left input image. (b) Right input image.](image)
Fig. 7. Results for "OutdoorZoom" scenario for several frames. (a) Right input image. (b) Right image segmented into grey level bands. (c) Right 2D Charge Memory. (d) 3D Depth Map.
The values of the main parameters used in our test series were:

\[ C_{\text{dis}} = 128; \quad nGLB = 8; \quad C_{\text{min}} = 0; \quad C_{\text{max}} = 255; \quad 0a3 = 1 \]

Fig. 7 shows the result for some of the more representative results of applying our algorithms to the "OutdoorZoom" scenario. In column (a) the input image is shown, in column (b) the segmentation in grey level bands may be appreciated, in column (c) motion information as represented in the right motion charge map is offered, and in column (d) the final output, that is to say, the scene depth as detected by the algorithms, is presented. You may observe that light colors in column (d) means that persons are closer to the cameras. Black means there is no motion detected. The main information is available in columns (c) and (d). We may observe some details, as, for example, the following ones:

- In frame 35, a person is entering the scene on the right side, very close to the cameras. This is why, in column (c), the final output, very light grey levels appear.
- This person progressively is moving away from the cameras, in such a way that on frame 50 it is represented by intermediate grey levels.
- In frame 65, the person is now far away from the cameras. Its shape appears in dark grey values.
- Let us now focus on frame 215. A person is walking down the steps and at the same time an object is appearing on the right side of the image. It may be appreciated at the output of the system that the object is a bit lighter than the person. Thus, the object has to be closer to the cameras than the walking person.
- From frame 215 to frame 228, the pedestrian is walking horizontally (to the left). Thus, we appreciate no difference in the grey levels present in these frames.
- In frame 245, the person turns around, but there is still no difference appreciated in its depth in the scene.
- Lastly, in frame 261, we may observe the person leaving the scene on the right
side, and at the output very light grey levels. This obviously means that the man is very close to the cameras.

The only possible comparison for the “OutdoorZoom” scenario is with the results of (Di Stefano et al., 2004). As the aim of both approaches is totally different - Di Stefano and collaborators work on luminescence parameters on the whole image, whereas we compute on the persistency of moving objects -, it is only possible to provide a qualitative impression of the reliability of our algorithms. According to this reasoning, we may state that our stereovision depth analysis is very accurate for the objective.

4. Conclusions

According to the conclusions provided by (Gutiérrez & Marroquín, 2004), the matching approach to solve the so called correspondence problem in stereovision has intrinsic limitations. Hence, the stereovision depth analysis problem is still a bug challenge in computer vision.

In this paper we have introduced a new method to retrieve disparity information based on motion and stereovision. A motion detection representation helps establishing further correspondences between different motion information. This representation bases in the permanency memories mechanism, where jumps of pixels between grey level bands are computed in a matrix of charge accumulators. Thus, for the purpose to analyze scene depth from stereo images, we have chosen the alternative not to use direct information from the image, but rather the one derived from motion analysis. This alternative provides an important advantage. Through motion information stored in two-dimensional motion charge maps it is easier to use correspondences than by grey level information of the frames. When observing motion features of a concrete object in both stereo sequences at the same time instant, we appreciate that these features are extremely similar. This is the reason why it is easy and robust to establish correspondences between the motion information of an object at the right image respect to the object at the left image. The number of ambiguity possibilities is lower, as we
have filtered a lot of information by eliminating all static elements in the scene. By reducing the number of elements (what does not move, does not exist), the matching process is also easier. Moreover, information about motion is associated to the history of the last few frames in the scene, not only to the present scene. Thus, the matching process takes place based on more accurate information (two similar elements with different recent motion history generate different trails).

Charge memory displacement allows us to transform disparity into time, time into charge, and finally, charge into depth. The solution proposed involves a type of process which tries to take advantage of the use of high-order primitives and pixels only. On the one hand, the elements placed in correspondences are regions obtained from moving objects' motion trails by means of permanence memory interpretation. On the other hand, the fact that each pixel can decide, through a local analysis and based on motion trail overlapping, which disparity is more reliable, creates a dense disparity memory, which is considered a great advantage in pixel-based correspondence systems.

Acknowledgements

This work is supported in part by the Spanish CICYT TIN2004-07661-C02-02 grant.

References


