

Stereovision Disparity Analysis by Two-Dimensional Motion Charge Map Inspired in Neurobiology

José M. López-Valles¹, Miguel A. Fernández², Antonio Fernández-Caballero²,
and Francisco J. Gómez²

¹ Universidad de Castilla-La Mancha,
Escuela Universitaria Politécnica de Cuenca, 13071 - Cuenca, Spain

`JoseMaria.Lopez@uclm.es`

² Universidad de Castilla-La Mancha,
Escuela Politécnica Superior de Albacete, 02071 - Albacete, Spain

`{miki,caballer,fgomez}@info-ab.uclm.es`

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 is inspired in Neurobiology and that turns around the symbiosis existing between stereovision and motion is introduced. A motion representation in form of a two-dimensional motion charge map, based in the so-called permanency memories mechanism is presented. For each pair of frame of a video stereovision sequence, the method displaces the left permanency stereo-memory on the epipolar restriction basis over the right one, in order to analyze the disparities of the motion trails calculated.

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. So far, many algorithms have been developed to analyze the depth in a scene. Brown et al. [1] 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. According to the correspondence techniques used, we may classify methods into correlation-based [2], relaxation-based [3], gradient-based [4], and feature-based [5].

In this paper a method that is inspired in Neurobiology and that turns around the symbiosis existing between stereovision and motion is introduced; motion minimizes correspondence ambiguities, and stereovision enhances motion information. This symbiosis, evident in biological systems, has been studied to get a

major performance in our artificial three-dimensional disparity 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 [6],[7],[8]. In this paper, we have chosen as an alternative not to use direct information from the image, but rather the one derived from motion analysis. The system proposed uses as input the motion information of the objects present in the stereo-scene, and uses this information to perform a depth analysis of the scene, through the use of a two-dimensional motion charge map.

2 Neurobiological Inspiration

When looking for inspiration in Biology for stereoscopic disparity analysis, some questions hit our attention. As explained next we have based our method in motion perception -due to micro-saccadic eye movements and ego-motion of the targets-, luminescence perception -perception of brightness rather than colour perception-, and binocular perception, present in most superior primates.

2.1 Motion Perception

The first important question is that all living beings with the capacity of seeing, only perceive objects that move relatively with respect to their retinas. This characteristic, which is evident in most mammals, is present also in superior primates and humans; but this is not as evident in this case. This is due to an illusion that we are fixing our look in a static object. Nonetheless, the perception of static objects respect to the retina is only possible thanks to the micro-saccadic movements of our eyes [9].

On the other hand, these micro-saccadic eye movements only affect in a sufficient manner the fovea region. That is why it is usual that we do not detect objects in the periphery of the visual field, if the objects have no ego-motion. The perception of a moving object in the periphery produces a reflex movement, so that the object is instantaneously centred in the visual field to be observed in a correct and detailed way [10],[11]. Motion is really a top cue in our proposal.

2.2 Luminescence Perception

Motion and stereovision perception are closely related as largely demonstrated in isoluminance experiments [12]. Isoluminant stimuli are stimuli whose luminance does not change over time; only their wavelength (colour) changes. The magnocellular pathway, which relates motion perception with the depth, is practically insensible to colours and only distinguishes among stimuli whose brightness levels are different. Experiments performed with isoluminant stimuli show that the isoluminance is a difficulty for the perception of motion and of the depth of the scene [13],[14]. In our approach motion from brightness difference is calculated, and colour is not used at all.

2.3 Binocular Perception

In relation to stereoscopic visual perception, the ocular dominance columns and the near cells and far cells are a fundamental reference [15]. But it seems that these are not the only responsible for the complete three-dimensional perception [16]. Indeed, in humans and superior primates there is no total three-dimensional perception. There is only a little margin centred on the fixation point. Outside of this margin there is double vision. Both eyes travel along the tracked object, converging and diverging in order to fuse into a single image the couple of instantaneous perspectives gotten from the object. In our proposal, for each pair of frame of a video stereovision sequence, the method displaces the left image on the epipolar restriction basis over the right one.

3 Disparity Analysis from Motion Charge Map

Starting from these neurobiological evidences, our system for the analysis of the depth of a scene integrates stereovision and motion. Our proposal is to analyze motion in the original sequences by means of the so-called permanency effect [17], and from the resulting charge maps to analyze the disparities. 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 persistency charge measure.

3.1 Motion Charge Map

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. 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 can be achieved realizing that 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 [18].

The motion analysis algorithm used in this work has already been tested in applications such as moving object shape recognition in noisy environments [19],[20], moving objects classification by motion features such as velocity or acceleration [17], and in applications related to selective visual attention [21]. Motion analysis performs separately on both stereovision sequences in two phases. The first analysis phase is based in grouping neighbouring pixels that have similar grey levels in closed and connected regions in an image frame. The method used is segmentation in grey level bands. This method consists in reducing the resolution of illumination levels of the image, obtaining this way a lower number of image regions, which potentially belong to a single object in motion. Let

$B(x, y, t)$ be the grey level band associated to pixel (x, y) at time instant t , $GL(x, y, t)$ the grey level, n the number of grey level bands, and N the number of grey levels, then:

$$B(x, y, t) = \lfloor \frac{B(x, y, t-1) \cdot n}{N} + 0.5 \rfloor \quad (1)$$

A detailed analysis of the features and performances of this segmentation method is described in [22]. Obviously, segmentation in grey level bands performs in parallel on each couple of images of the stereo sequence.

Once the objects present in the scene are approximated in a broad way, the second phase has to detect possible motions of the segmented regions. Again, motion information of both video sequences that form the stereo pair is extracted. Motion detection is obtained from image pixels change in luminosity as the video sequence goes on through time. Motion in an image segmented in grey level bands is detected through the variation of the grey level band of the pixels. Notice that it is not that important that regions neither completely adjusts to the shape of the objects, nor that at a given moment two different objects appear overlapped in a same region. Consider that the proper relative motion of the objects will force those regions belonging to a same object to move in a uniform way, and those regions that hold different objects separate in the future.

From motion detection, we now introduce a representation that may help to establish further correspondences between different motion information. This representation finds its basis in the permanency memories mechanism. Precisely, this mechanism considers the jumps of pixels between bands, and it consists in a matrix of charge accumulators. The matrix, also called motion charge map, is composed of as many units in horizontal and vertical direction as pixels there are in an image frame. This way, a position (x, y) of the image is associated to a permanency memory charge unit. Initially all accumulators are empty; that is to say, their charge is minimal. The charge in the permanency memory depends on the difference between the current and the previous images grey level band value. An accumulator detects differences $diff(x, y, t)$ between the grey level bands of a pixel in the current and the previous frame:

$$diff(x, y, t) = \begin{cases} 0, & \text{if } B(x, y, t) = B(x, y, t-1) \\ 1, & \text{if } B(x, y, t) \neq B(x, y, t-1) \end{cases} \quad (2)$$

When a jump between grey level bands occurs at a pixel, the charge unit (accumulator) of the permanency memory at the pixel's position - $Ch(x, y, t)$ - is completely charged (charged to the maximum charge value max). This is the way to record that motion has just been detected at this pixel. This complete charge is produced when there is a jump to superior bands as well as to inferior bands. Thus, charge units of the permanency memory are able to inform on the presence of motion of the associated pixels. After the complete charge, each unit of the permanency memory goes decrementing with time (in a frame by frame basis) down to reaching the minimum charge value min , while no motion is detected, or it is completely recharged, if motion is detected again.

This behaviour is described by means of the following formula, where again $B(x, y, t)$ is the grey level band associated to pixel (x, y) at time instant t . dec is a fixed application-dependent quantity, which is decremented to the instantaneous charge of each charge unit each time that a frame is analyzed and no motion is detected. Thus, this quantity shows the discharge velocity of the permanency memory.

$$Ch(x, y, t) = \begin{cases} max, & \text{if } diff(x, y, t) = 1 \\ \max[Ch(x, y, t - 1) - dec, min], & \text{if } diff(x, y, t) = 0 \end{cases} \quad (3)$$

Values of parameters dec , max and min have to be fixed according to the applications characteristics. Concretely, values max and min have to be chosen by taking into account that charge values will always be between them. dec defines the charge decrement interval between time instants $t - 1$ and t . Thus, notice that the two-dimensional motion charge map stores motion information as a quantified value, which may be used for several classification purposes.

Thus, obviously, the evolution of charge in space depends on the velocity of the mobile in a direction. A slow mobile causes a short charge slope, as the object's advance from pixel to pixel may last various frames. During this time elapsed all affected units are discharging. In this case, between the charge and discharge of a unit, the mobile covers a short distance. On the other hand, a quick mobile causes that various memory units charge at the same time, such that there will be many more units affected by this motion. Thus, in this second case, between the total charge and discharge of a unit of the memory the mobile covers many pixels.

Fig. 1 shows all these issues. Fig. 1a and Fig. 1b 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. 1c you may observe the effect of these moving objects on the permanence memory.



Fig. 1. Motion charge map: (a) one image of a sequence, (b) same perspective after some seconds, (c) motion trails as represented on the bidimensional 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 cars one, is presented. Thus, permanency memories enable representing the motion history of the frames that form the image sequence, that is to say, there is segmentation from the motion of the objects present in the scene.

However, the dependency of the permanency memories from the segmentation in grey level bands imposes a restriction. The diminishment of the resolution in illumination levels produced by the segmentation in grey level bands does not exactly imply segmentation into objects. Some of the objects of the images are segmented into various regions, and physically distinct objects may be overlapped into a same region. Nevertheless, this issue is not that important when taking into account that our aim is to characterize motion of the objects and not their shape.

3.2 Stereovision Disparity Analysis

Motion-based segmentation into a two-dimensional motion charge map, as explained in the previous section, facilitates the correspondence analysis. Indeed, motion trails obtained through the permanency memories charge units are used to analyze the disparity between the objects in the stereo pair in a more easy and precise way. The set of all disparities between two images of a stereo pair is called the disparity map.

The retrieval of disparity information is usually a very early step in image analysis. It requires stereotyped processing where each single pixel enters the computation. In stereovision, methods based on local primitives as pixels and contours may be very efficient, but they are too much sensitive to locally ambiguous regions, such as occlusions or uniform texture regions. Methods based on areas are less sensitive to these problems, as they offer an additional support to obtain correspondences of difficult regions in a more easy and robust way, or they discard false disparities. Although methods based on areas use to be computationally very expensive, we introduce a simple area-based method with a low computational cost.

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. 2, 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. Looking at Fig. 2 it is possible to analyze the motion of each one of the three objects present in the permanency memories from their motion trails. This initial analysis is independent of the epipolar constraint studied. You may observe that object "a", which has a long trail and has his 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.

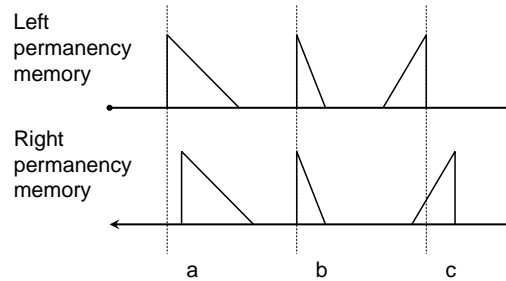


Fig. 2. Disparity by permanency memories

Also from Fig. 2, but now comparing between the motion trails in both epipolar lines, disparity is analyzed. Motion trail of object "b" presents a null disparity. Therefore, we can conclude that this trail corresponds to an object that is far away from the cameras. Remember that due to our parallel cameras configuration, pixels with a null disparity are located in the infinite. Object "a" has a little greater disparity. Finally, object "c" offers the greatest disparity.

This simple example draws three 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 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, the displacement value applied to the epipolar line is the disparity value. In third place, if we consider the representation of a mobile with a high velocity, various charge units of the permanence memories may charge simultaneously. This way, an object may correspond to various disparities. This is the reason why one single memory unit is not able to establish the disparity of an object. It is necessary to analyze the correspondence from the values of various units. The decision of all units has to validate the overall disparity value. The more efficient way to manage this is that each pixel chooses its disparity in such a way that the maximum of its neighbouring units confirm the 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 neighbourhood of each unit. The one displacement, which produces that a maximum number of surrounding elements confirm its correspondence, demonstrates to be the more trustful disparity value.

4 Data and Results

In order to test our algorithms, a real stereo sequence is shown. We show the results of applying our algorithms to a scenario called "OutdoorZoom", downloaded from labvision.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. The values of the main parameters used in our test series were: $dec = 128$; $n = 8$; $min = 0$; $max = 255$

Fig. 3 shows the result for some of the more representative results of applying our algorithms to the "OutdoorZoom" scenario. In row (a) the segmentation in grey level bands may be appreciated, in row (b) motion information as represented in the right permanency memory is offered, and in row (c) the final output, that is to say, the scene depth as detected by the cameras, is presented.

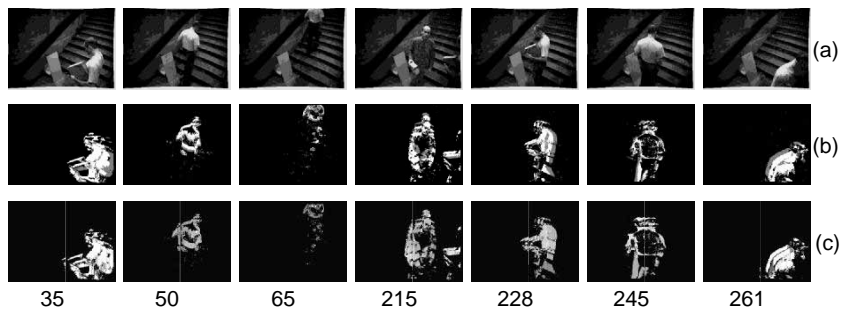


Fig. 3. Results for "OutdoorZoom" scenario

You may observe on Fig. 3 that light colours in row (c) means that persons are closer to the cameras. Black means there is no motion detected. The main information is available in columns (b) and (c). 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.

5 Conclusions

In this paper we have introduced a new method to retrieve depth based on motion and stereovision with a clear inspiration in Neurobiology. 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.

The biologically-motivated symbiosis between motion and stereovision enables getting two concrete aims. Firstly, it is possible to eliminate all static information in artificial vision systems. This is really important in dynamic systems (e.g. robotic vision), where the real important thing is the motion information in the environment. In second place, it is easier to correlate the motion information from both stereoscopic views, as motion is much more robust in eliminating ambiguities in the correspondence process.

Biological systems use simple cells to detect motion, by tuning characteristic stimuli; charging and discharging provide relative information of position and motion. In our case, the permanency effect permits to maintain the history of each movement of the scene. This effect is really simple and its results have been successfully explored [19],[20],[21],[22]. Finally, biological systems use the convergence of the eyes as a means to explore the individual's environment and to fuse the information coming from the two perspectives into the ocular dominance columns. In approach introduced, the inspiration leads to the mechanism of displacing one image over the other looking for the maximum number of corresponding elements, which provides an idea of the correspondence confidence.

6 Acknowledgements

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

References

1. M.Z. Brown, D. Burschka, G.D. Hager: Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**:8 (2003) 993–1008
2. R. Zabih, J. Woodfill: Non parametric local transforms for computing visual correspondence. *Third European Conference on Computer Vision* (1994) 150–158

3. W.E.L. Grimson: Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7** (1985) 17–34
4. I. Choi, J.-G. Yoon, Y.-B. Lee, S.I. Chien: Stereo system for tracking moving object using log-polar transformation and zero disparity filtering. 10th International Conference on Computer Analysis of Images and Patterns (2003) 182–189
5. V. Venkateswar, R. Chellappa: Hierarchical stereo and motion correspondence using feature groupings. *International Journal of Computer Vision* **15** (1995) 245–269
6. A.Y.K. Ho, T.C. Pong: Cooperative fusion of stereo and motion. *Pattern Recognition* **29**:1 (1996) 121–130
7. J. Liu, R. Skerjane: Stereo and motion correspondence in a sequence of stereo images. *Signal Processing: Image Communication* **5**:4 (1993) 305–318
8. G. Xu. Unification of stereo, motion and object recognition via epipolar geometry. 2nd Asian Conference on Computer Vision **I** (1995) 287–291
9. D.H. Hubel: *Eye, Brain, and Vision*. Scientific American Library (1995)
10. E.R. Kandel, J.H. Schwartz, T.M. Jessel: *Principles of Neural Science*. McGraw Hill (2000)
11. J.H. Martin: *Neuroanatomy: Text and Atlas*. Elsevier (1989)
12. Z.-L. Lu, L.A. Lesmes, G. Sperlingdagger: The mechanism of isoluminant chromatic motion perception. *Proceedings of the National Academy of Sciences of the United States of America* **96**:14 (1999) 8289–8294
13. W.H. Merigan, J.H.R. Maunsell: How parallel are the primate visual pathways? *Annual Review of Neuroscience* **16** (1993) 369–402
14. P.H. Schiller, N.K. Logothetis, E.R. Charles: Functions of the colour-opponent and broad-band channels of the visual system. *Nature* **343**:4 (1990) 68–70
15. G.F. Poggio, B. Fischer: Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology* **40** (1977) 1392–1405
16. H.A. Mallot, H. Bideau: Binocular vergence influences the assignment of stereo correspondences. *Vision Research* **30** (1990) 1521–1523
17. M.A. Fernández, A. Fernández-Caballero, M.T. López, J. Mira: Length-speed ratio (LSR) as a characteristic for moving elements real-time classification. *Real-Time Imaging* **9** (2003) 49–59
18. O. Faugeras: *Three-Dimensional Computer Vision, A Geometric Viewpoint*. The MIT Press (1993)
19. A. Fernández-Caballero, M.A. Fernández, J. Mira, A.E. Delgado: Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation. *Pattern Recognition* **36**:5 (2003) 1131–1142
20. A. Fernández-Caballero, J. Mira, M.A. Fernández, A.E. Delgado: On motion detection through a multi-layer neural network architecture. *Neural Networks* **16**:2 (2003) 205–222
21. A. Fernández-Caballero, J. Mira, A.E. Delgado, M.A. Fernández: Lateral interaction in accumulative computation: A model for motion detection. *Neurocomputing* **50** (2003) 341–364
22. A. Fernández-Caballero, M.T. López, M.A. Fernández, J. Mira, A.E., Delgado, J.M. López-Valles: Accumulative computation method for motion features extraction in dynamic selective visual attention. 2nd International Workshop on Attention and Performance in Computational Vision, *Lecture Notes in Computer Science* **3368** (2004) 206–215