Brain Inspired Cognitive Systems
August 29 – September 1, 2004
University of Stirling, Scotland, UK

STEREOSCOPIC DEPTH - A BIOLOGICAL INSPIRED JUDGMENT

Hugo Gravato Marques, IEETA – Institute of Electronic Engineering of University of Aveiro, Campus Universitário de Santiago, 3800-193 Aveiro, Portugal.

Eugénio Oliveira, Faculty of Engineering of University of Oporto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal.

ABSTRACT

The present paper aims to contribute to the research on depth perception using a biologically inspired model of stereoscopic vision. Stereogram images show that human beings are able to perceive depth just from the differences between the images coming from the retinas of both eyes. The research that was made aspires to find support for the idea that the way stereogram images are perceived by human beings is just a consequence of the way they see on real world environments; that is, to show that the information received by the brain over generations is sufficient to make us interpret stereogram images as having differences on depth.

Usually, to implement a stereoscopic depth mechanism on robots two parallel cameras are used. The images supplied by these cameras are very difficult to match by means of a biologically inspired algorithm since these ones require the use of local information, and the extension of information needed to match the images from two parallel cameras is too large. Human beings have the ability to move their eyes and when they look at a real world object both eyes converge to that object. The disparity method built was expected to make use of this behavior (eye convergence) in order to reduce the disparity computation to local and much simpler algorithms.

INTRODUCTION

Nowadays the idea that our brain has neurons dedicated to compute the differences between the images coming from the retinas of both eyes is generally accepted among neuroscientists. It is also generally accepted that those differences are most likely used to judge differences in depth. To support this idea cells that were selective for orientation and stereoscopic depth were found in area V3 of the visual cortex of macaques [1]. An easy way to prove that we are able to perceive depth from the differences of both retinas is the way the brain deals with stereograms, namely Random Dot Stereograms (RDS) introduced by Julesz in 1960 [2,3]. RDS images, because of their lack of meaning, do not contain any other information rather than some horizontal shifted points in some parts of one image that do not have a correspondent shift on the other. Our brain perceives these shifted areas as being at different distances giving a strong indication of the presence of stereoscopic vision.

Staring at stereograms involves directing our eyes to different images, or at least different parts of the same image. The most common situation is to make the viewer put his or her eyes parallel to each other or simply crossed, in such a way that different images are printed in both retinas. This procedure is a trick that we do with our eyes and brain considering the way that we usually see real-world objects. Typically our eyes are constantly moving and being directed to objects and, in order to perceive them clearly, both eyes must be converged to the same place[a]. This activity of the eye allows the brain to put in the same position the representation of the objects on both retinas (at the centre). It is the goal of this paper to show that the information received by the brain over centuries, every time eye convergence is carried out is enough to make us interpret stereogram images the way that we do.

When computer vision is considered to produce depth results from stereoscopic information, usually two parallel cameras are used. Even if, by using them, we can achieve good results, it is very difficult to create a biologically inspired algorithm that is able to produce satisfactory results because biological algorithms require local information and the extent of data needed to compute disparity from parallel cameras is too large. By using the ability to move the cameras we hope to surpass this barrier and be able at the same time to realize depth on real-world scenes as well as on stereogram images.

NOMENCLATURE

This article is about vision and it contains technical names like: disparitity, stereograms and stereopsis. Disparity is the difference between the images printed on both eyes. Stereogram is a 2D image conceived in such a way that it is interpreted as a 3D image. Finally, stereopsis is the perception of depth that we have from differences between the images printed on both eyes.

[a] Only binocular vision is implied.

BINOCULAR DEPTH CLUES IN REAL-WORLD SCENES

The use of binocular vision presents two main clues available to judge depth – disparity and eye convergence. The former, as previously noted, consists in computing the differences between the corresponding points on the images coming from the retinas of both eyes [3]. This is the situation presented in Figure 1. The greater the distance between corresponding points in the retina, the more distant the object is to the focus point. This is a primordial issue, since the evaluation that is made is relative to the focus point and not to the viewer, that is, the evaluation by itself does not inform us of how far an object is from the viewer, it just informs us how far it is from the viewpoint and in which direction[b]. The sample in Figure 1 shows what happens if we focus on some point (A), and all other points (B and C) are farther from the viewer than that point (A). However, a generalization is quite simple to make; one has only to imagine a point (D) closer to the viewer than point A. That point would appear both on the left part of the left retina and on the right part of the right retina. If this happens it means that the point is closer to the viewer, otherwise (like in Figure 1) it means that it is farther away.

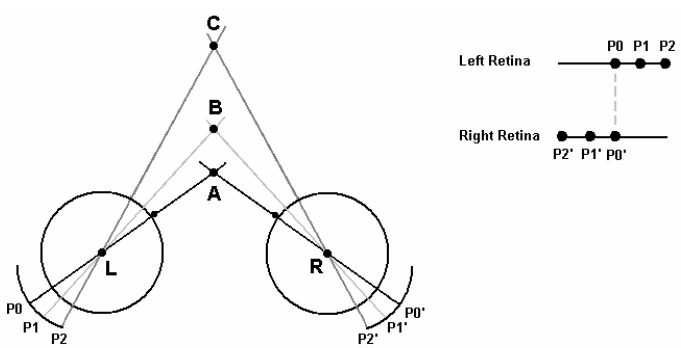

Figure 1: Stereoscopic vision mechanism in a real-world scene.

The second depth method is a simple measurement of the angle of both eyes. This is a much simpler idea than disparity. It is basically a measurement of the amount of rotation that both eyes need to perform (relatively to a fixed position[c]) when they are looking at a given point (Figure 2). The closer the object, the greater the rotation angle is - object A is closer than B, therefore the rotation needed to look at A is more extensive than to look at B.

PARALLEL VS. MOVING CAMERAS

As we can infer from the previous section, stereoscopic depth measuring implies calculating the amount of horizontal difference relative to the same point in both retinas. Once using biological algorithms mean computing results based on local information, using two parallel cameras brings up a very hard issue. The area where one point appears in both retinas can diverge a lot; in addition, it can diverge not only horizontally but vertically as well. For instance, if an object is very close to the viewer, a given point belonging to that object will be printed on the left boundary of the left eye and on the right boundary of the right eye, while if an object is too far way, that point will appear approximately in the same area on both retinas (Figure 3 and Figure 4). One can easily generalize Figure 4 by moving one of the points around and see that one point can be located at any position on one retina and at the same time to be located at any other position on half of the other one. This indicates a very strong limitation for biological algorithms which work with local information. The solution to this problem might be on the ability to move the eyes.

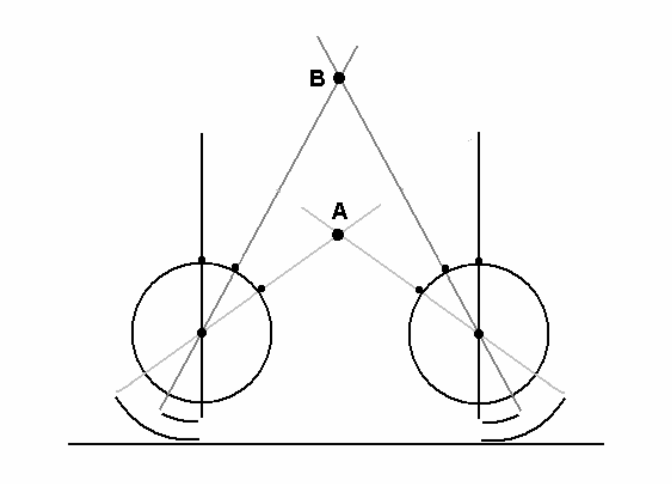

Figure 2: Eye convergence.

When we focus some specific point, the eyes converge to that point, which means that the point will be printed in the middle of both retinas (Figure 5). Having in mind that stereoscopic depth is based on the horizontal differences between the left and the right retinas, the presence of the point in the same location on both eyes does not give by itself any information about stereoscopic depth as there is no difference at all. However, having the position of both eyes, we are capable of judging the depth of the point as we saw in the previous section. The major achievement is that the convergence of the eyes into the same place makes the task of finding the correct match for each of the surrounding points much easier (regardless of their being farther or closer than the focus point). The points will most likely fall in close areas on both retinas; and they will be on the central area. If they do not, they are not very important to realize stereopsis anyway because, as we will see in Section *Visual Processing and Cells in the Visual Pathway*, the central area of the retina is the one that has the resolution needed to calculate it.

VISUAL ATTENTION AND EYE MOVEMENT

However, the moving eyes solution raises a different problem – how are we able to point both eyes to the same place? Or, in a slightly different way, how can we ensure they are pointed to the same place? It is not the aim of this paper to present a solution to this problem but, at least theoretically, it can be

[b] Closer or farther than the focus point relative to the viewer.

[c] For a better understanding, we should consider the angle between the current position of the eyes and their position when they are looking straight at the infinite.

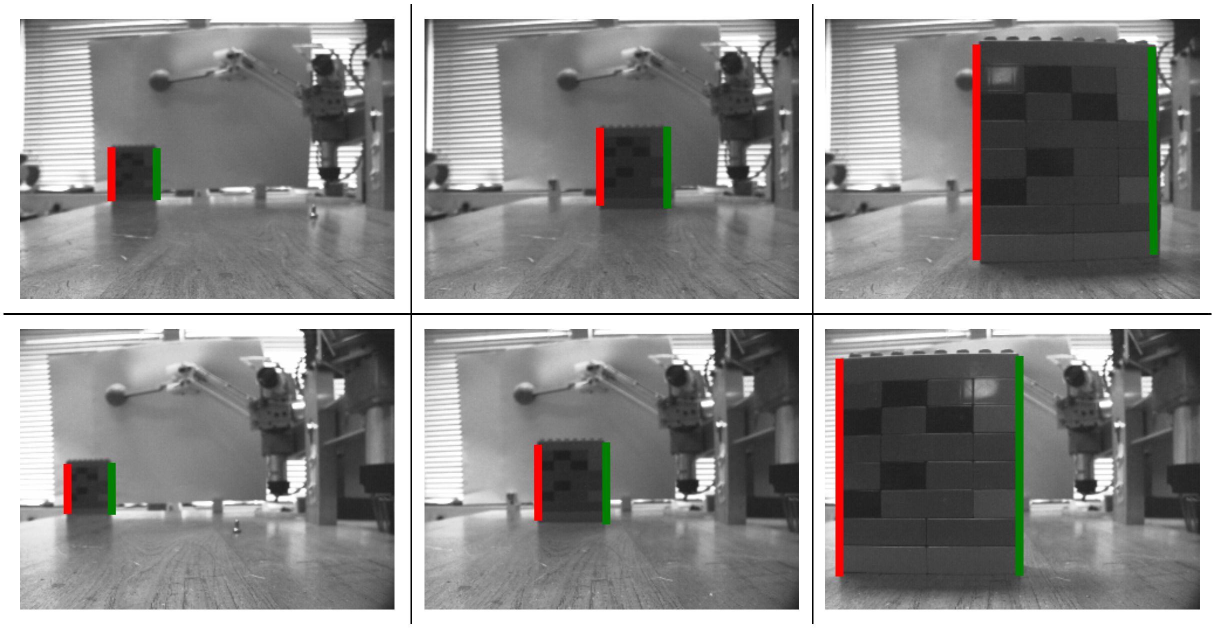

Figure 3: Pictures of objects at different distances taken by two parallel cameras. The upper pictures were taken from the left camera while the lower pictures were taken from the right camera. The distance of the objects get closer from the left to the right side.

said that it is a much easier problem to solve in a biologically inspired way than stereopsis with parallel eyes, since the eye muscles can receive information about the area where they have to be directed to; and that means working with local information rather than global.

The problem was approached simply by directing both eyes to an object of a given colour and then making them do small adjustments to make sure the object of attention was focused in the centre of the retina (Section *Attention Mechanism*). This was probably not the best way to come near of a solution for the convergence problem, mainly because it implied that all objects have different colours, however this was not a goal in the project we have been involved in, and it seemed clear at the time that it was the best way to test the disparity algorithm [4].

DIFFERENT AREAS OF THE RETINA WITH DIFFERENT GOALS IN THE BRAIN

The information that is received by our eyes is carried to the visual cortex in the brain via electrical impulses. There are two different visual pathways that carried those impulses – the magnocelular and the parvocellular pathways [5,3]. The magnocellular pathway is constituted by cells with large receptive fields that receive connections from M ganglion cells. The parvocellular pathway is composed by cells with smaller receptive fields that receive connections from P ganglion cells. The two pathways are presumed to have different goals in the brain. The magnocellular pathway, because its cells are very sensitive to brightness differences and have large receptive fields, is particularly suited to process information related with movement and spatial arrangement; while the parvocellular pathway, because its cells have a much smaller receptive field and are very sensitive to colour, is more suited to deal with information about shape and colour.

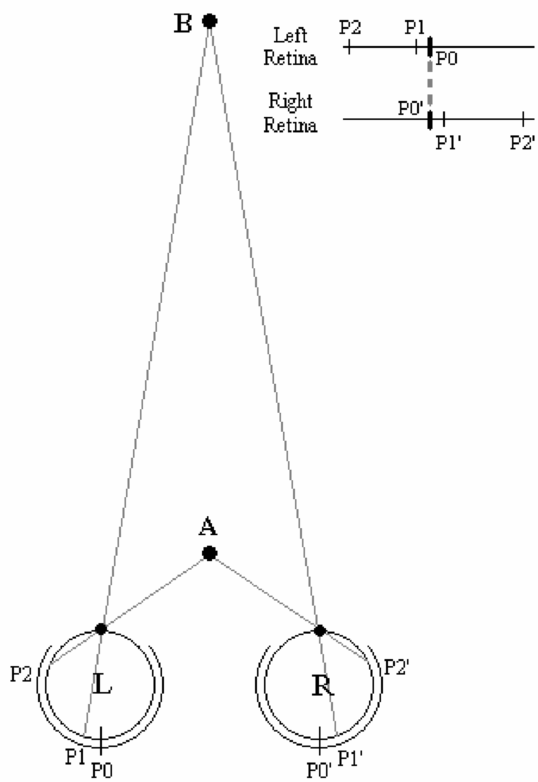

Figure 4: Vision without the ability to move the eyes (parallel eyes).

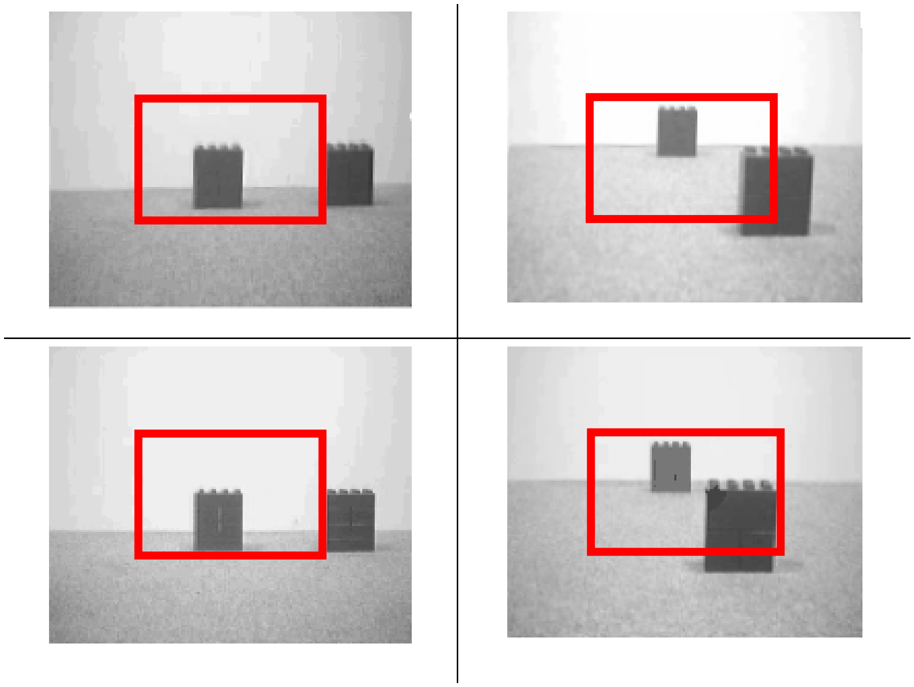

Figure 5: Pictures of objects at different distances taken by two moving cameras. The upper pictures were taken from the left camera while the lower pictures were taken from the right camera. The images in the first column represent the focus of the object by both cameras. The second column represents the differences between the images on the left eye and on the right eye when another object is closer to the eyes than the object that is being focused.

The eye itself includes areas with different resolution power [5]. The centre area of the retina – the fovea – is a small area that contains many photoreceptors and consequently it has a high resolution capacity (proper to process shapes accurately). The farther we get from the centre area – the peri-fovea - the less populated the retina becomes and therefore the lower resolution it has.

In our project, these different resolution areas turned out to be very useful. On the one hand, the peri-fovea area, because it encloses the larger area of the visual field, was used to get an idea of the scene configuration and to provide information to the attention mechanism. On the other hand, the fovea area because of its high resolution capabilities was ideal to compute disparity.

VISUAL PROCESSING AND CELLS IN THE VISUAL PATHWAY

The ability to judge stereoscopic depth requires a visual pre-processing step as there is information that does not contain by itself any stereoscopic value. This is the case of the inside part of an object. Let us suppose that we are trying to focus a given red object. If a given point inside the object is printed on the left retina, that point will have many correspondent points on the right retina since the whole object has the same colour. Thus, if we had just the contour of the object, it would be considerably easier to find the correct match. In the same way, the horizontal lines of a given contour do not give a decisive contribution to calculate disparity, since the differences are computed horizontally, and a point belonging to a horizontal line can be (incorrectly) matched with all other points within

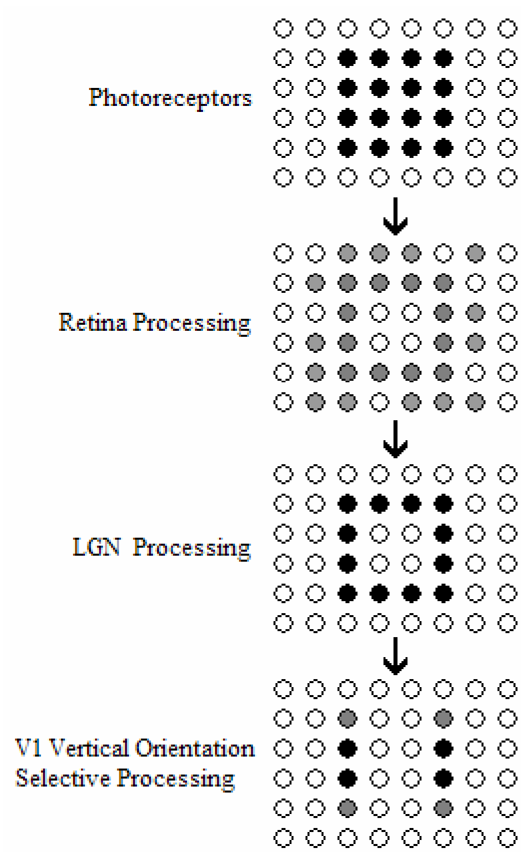

Figure 6: Visual Processing

the line. Thus, the closer to vertical a line is the more difficult it is to find an incorrect match in the area of the object.

To process the information that was coming directly from the cameras the computation made by the cells in the visual pathway was simulated [4]. In the visual field of mammals the retina cells are connected to LGN cells and LGN cells are connected to the V1 area of the visual cortex [5] (Figure 6). From the V1 area, cells connect directly, or by means of an intermediate layer, to all regions of the visual cortex. The cells on these regions were implemented in order to provide all the information necessary to the disparity algorithm. The retina cells were implemented to make the contour of the shapes presented to the eye. The LGN cells were built to represent, in a more accurate way, the information coming from the retina cells whereas V1 area cells were intended to be selective to vertical lines (or diagonals close to vertical)[d]. This last type of cells was directly used by the disparity algorithm.

ATTENTION MECHANISM

The idea behind the attention mechanism used the colour information coming from the cells in the visual pathway to point the eyes to the same place.

Before starting the attention mechanism, it was necessary to make the robot learn to associate a given colour with its RGB components. This was done by showing it different objects with different colours and by setting the colour that it should learn. At the end of the colour learning phase, one was able to put objects of different colours in the experiments area and make the robot point its cameras to the object of a given colour.

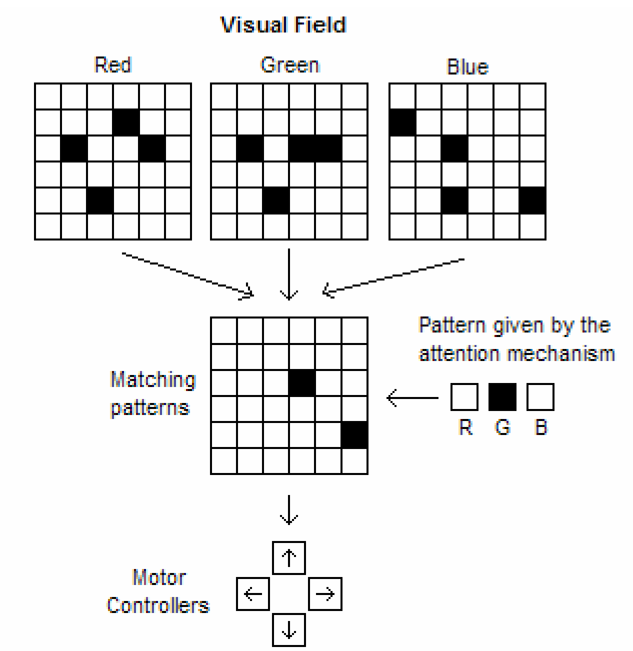

Figure 7: Attention Mechanism.

The mechanism worked in a very simple way (Figure 7). After providing the robot with the colour of the object that it should look, it immediately transformed the colour in the correspondent RGB pattern learnt previously. Next, a mapping of the location of matching patterns was built based on a scan made in the visual filed. Finally, that information was passed to the motor neurons that direct the cameras to the place of the object.

Finally to put the objects more or less in the same foveal position, an amount of the differences between the two foveas was calculated and one of the cameras was moved accordingly, a little bit to the right or to the left.

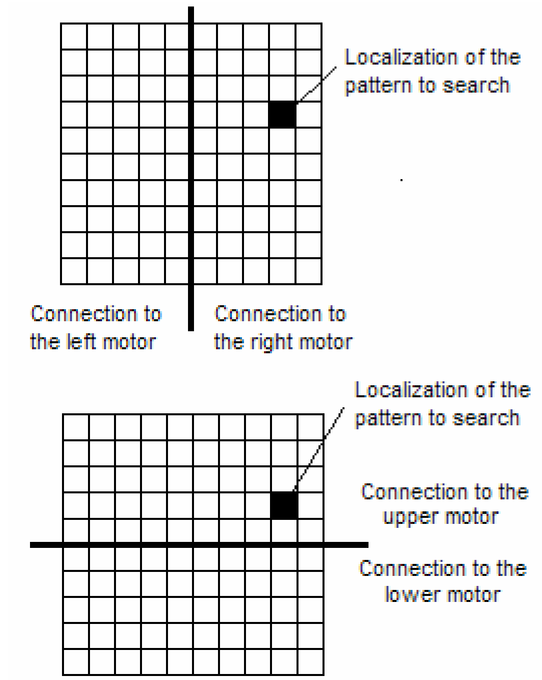

Figure 8: Connections between the motors and the visual field.

The motors received information corresponding with its own part of the visual field, that is, the left motor received information about the left part of the visual field and the right motor received information about the right visual field (Figure 8 - Top). The same happened about the upper and lower motor[e]. The motors moved to the closer matching pattern in the perpendicular of its direction. For example, in the (Figure 8) the right motor had to move 3 units and the upper motor 1 unit. Since only one object of each colour was present in the experiments area the cameras were forced to converge and, consequently, the object was in the fovea of both eyes.

[d] As previously observed horizontal lines do not contribute decisively to stereoscopic computation. We have just used vertical (or diagonals close to vertical) selective cells. To make the mechanism more flexible we would have to create cells selective to other orientations.

[e] These two motors had an extra constraint that did not allowed them to move separately on both eyes – these motors on the right eye were slaves of the ones on the left eye.

DISPARITY ALGORITHM

The cells built that were responsible for stereoscopic judgement received direct connections from vertical selective cells at the centre area of both eyes – the “virtual fovea”. As we have seen in section *Visual Processing and Cells in the Visual Pathway*, stereoscopic depth judgement implies working with high resolution because the differences between both eyes can be very small in some areas. Therefore, the information coming from the fovea, since it has high resolution, is more appropriate to compute stereoscopic differences.

The disparity algorithm works like this: each virtual disparity cell is connected to the correspondent V1 cell in the left eye (camera) and to fifteen cells[f] on the right eye (camera) – seven adjacent cells to the left, seven adjacent cells to the right and the correspondent V1 cell itself (Figure 9)[g]. Then, for each cell the following algorithm is applied:

```
BEGIN:
If (left_eye_cell is not active) then
    // nothing to be compared with
       Return no_response;


// left cell is active
If (righ_eye_corr_cell is active) then
// direct match found – no difference
    Return 0;

// gets the closest match in right eye input
// vector
var a = getClosestMatch();

If (a equals to NULL)
    // match not found
    Return no_response;

// gets the difference to the centre
var b = getDifference(a);
If (a is in the left part of the right eye)
    Return b; // closer
Else // a is in the right part of the right eye
    Return –b; // farther
END.
```

Thus, looking at the algorithm above as well as to Figure 9 we can see that points closer to the viewer than the focus point appear on the left part of the right eye whereas objects farther appear on the right side of the right eye[h]. The distance from a given point to the focus point can be measured by the distance to the centre of the input vector of the right eye.

[f] This is an example. The amount of connections depends directly on the resolution and sensitiveness desired. To judge stereoscopic depth accurately it is required more than one layer with different resolutions.

[g] All connections are placed in a horizontal line.

[h] If we look at Figure 1 we could be tempted to infer the opposite idea – objects closer to the viewer than the focus point appear on the right part of the right eye whereas objects farther appear on the left side of the right eye. However, we must remember that, contrary to what happens in the eye, the image taken from a camera is not reversed; or, in a more correct way, it is mirrored at the time that the photo is taken, and it is mirrored once again at the time that it is saved on the computer.

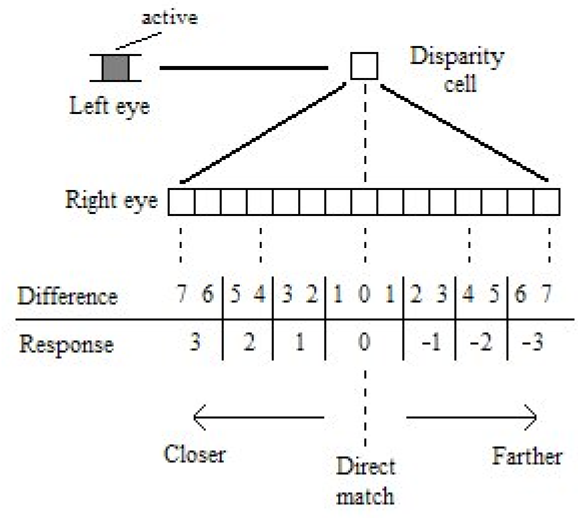

Figure 9: Virtual disparity cells’ input and responses.

TESTS AND RESULTS

The proposed algorithm was tested in a robot (Figure 10). The robot was equipped with two eyes (two web-cameras) that were able to take images from the real world and four “muscles” (motors) that allowed moving both eyes in an independent way. The “muscles” responsible for moving the eyes horizontally, had a 180º angle range, while the ones that were able to move them vertically, had 90º. The robot had yet a fifth “muscle”, simulating the neck, which was used to move the head horizontally (180º angle range).

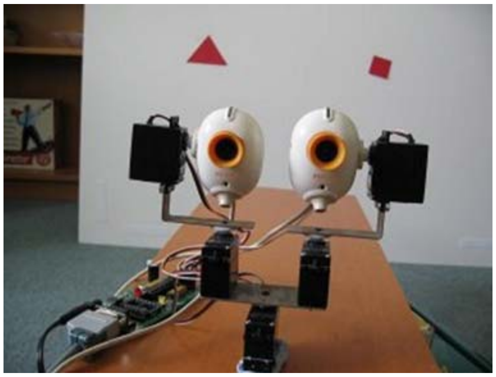

Figure 10: Picture of Frankenstein.

The first experiment started by making our robot (Frankenstein) to learn visual patterns and associate them with colours. This was done by showing it an object of a given colour and informing it what response should be given when that pattern was presented. Usually three within red, green, blue and yellow colours were learnt. After this first step, we put 3D objects in the real world scene (usually two of different colours) and asked Frankenstein to look at the object with the colour that we had selected. After that we picked the object that was not being the focus of attention by Frankenstein and put it near the other, but a slightly closer or farther. Frankenstein showed the ability to measure roughly the distances from the object moved to the object that was under focus; nevertheless, in almost all experiments he was able to judge if a given object was closer or farther the object focused. His failures could be solved by implementing multiple layers with different resolutions and by using colour information to avoid incorrect matches.

The second experiment began by pointing Frankenstein's cameras to the infinite, that is, parallel to each other. Then a large stereogram was placed in front of him in such a way that the right eye was just able to see the right part of the image and the left eye was just able to see the left part of the image. As we can see in Figure 11 the thinner bars on the right part are slightly displaced to the right and, on the left eye, a little to the left, while the larger bars were placed at regular intervals on the right eye and on the left one. Frankenstein interpreted the flat image as a 3D one – evaluating the thinner bars as being farther than the larger ones. This was the expected result since, on the one hand, the larger bars can be directly matched since there is no difference between them; while on the other hand, the thinner bars on the right side appear slightly shifted to the right compared with the ones appearing on the left side of the image. Thus, considering what was said in Section *Disparity Algorithm*, the difference must be judged as being farther in depth.

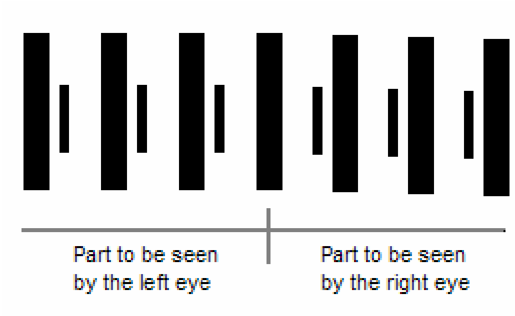

Figure 11: Stereogram image presented to Frankenstein.

CONCLUSIONS

The disparity method we have developed and used proved to be much simpler than the parallel cameras technique, regarding the eye adjustment method that was not fully studied. It was also proved that its implementation is possible using a biological inspired algorithm. The results obtained show that the algorithm is able to interpret stereoscopic depth both in real world scenes and in stereogram images as well.

ACKNOWLEDGEMENTS

I would like to thank Professor Igor Aleksander for the excellent supervision of the work that lead to this article. My thanks also to the people in IIS group of DEEE of the Imperial College and of the Faculty of Engineering of University of Oporto that in some way took part in the training period that provided the opportunity for the production of this article. I would also like to thank Professor Nuno Lau and Sofia Pimentel for helping me revising it. Finally, I would like to thank PRODEP FCT program for the financial support given during my training period.

REFERENCES

[1] Adams, Daniel and Zeki, Semir. *Fuctional Organization of Macaque V3 for Stereoscopic Depth*. The American Psychology Society (2001), p.1.

[2] Marr, David. *Vision*. W. H. Freeman & Company (1982), pp.8-9.

[3] Kandel, Eric; Schwartz and Jessell, Thomas. *Principles of Neural Science*. McGraw-Hill/Appleton & Lange (2000), pp. 549-569, 560-561.

[4] Marques, Hugo; Aleksander, Igor and Oliveira, Eugénio. *3D Vision using Neural Networks*. FEUP (2003), pp.14-19, 22-28, 35-39. Internal Document.

[5] Zeki, Semir. *A Vision of the Brain*. Blackwell Science (1993), pp. 22-30, 186-195.