The Disparity Pyramid: An Irregular Pyramid Approach for Stereoscopic Image Analysis

Rimon Elias
relias@site.uottawa.ca

Robert Laganière
laganier@site.uottawa.ca

School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada, K1N 6N5

Abstract

The disparity pyramid is a proposed solution for the problem of disparity estimation using a stereoscopic pair of images. This pyramid can be classified within the class of data-dependent irregular pyramids. The hierarchy of the pyramid is achieved by successive calculations of a set of levels. These calculations are done over a disparity range to get the difference in intensities between the two images. Unlike some other pyramids, and according to input values and/or distances between disparity vectors, the top level of this pyramid may consist of more than one cell. Thus, in terms of geometry, its shape looks like an incomplete pyramid with a flat top level and irregular sides.

1 Introduction

Disparity Estimation concerns finding pairs of point features in two perspective views of a scene such that each pair corresponds to the same scene point. Usually, the relationship between the viewpoints used to obtain the two images is unknown. In general, the two subsets of points are locally intensity matched and preserve the interpoint geometrical structure. As is true for most matching algorithms, the underlying principle for matching is geometrical similarity and/or intensity (or colour) similarity [5]. However, this similarity measure may cause some confusion when repeated many times within the same region and unfortunately this is the usual case. So we claim that the segmentation of the image according to the disparity variation would help eliminate this problem.

There is no standard approach to segment a scene. Many different types of image or scene parts can serve as the segmentation. Depending on which descriptions are based, there are many ways in which one can attempt to extract these parts from the image. However, we need tools that rapidly provide useful information and we claim that pyramids are such tools. For example, Jolion and Montanvert proposed the adaptive pyramid [6] which is an irregular pyramid used to segment the image according to the different gray levels included. Another similar pyramid called the stochastic pyramid was proposed in [10]. A few attempts using the regular pyramids in disparity estimation were introduced, e.g., refer to [2, 17, 14, 4, 16]. However, they were all constrained by the structure of the quad-pyramid in which one parent has exactly four children. For further discussion on the regular pyramids, refer to [7] and on stereo vision, refer to [11, 1, 3, 9, 13, 15, 8, 12]. One proposed technique using the concepts of disparity and irregular pyramids is presented in this paper.

2 Intensity-based matching

Applying to all the points in the image, a difference measure is computed over a window W. This window is determined by a disparity range $(dx_{min}, dx_{max}, dy_{min}, dy_{max})$. Let $f_1(x,y)$ be the image intensity at the point (x,y) in the first image and $f_2(x,y)$ be the image intensity at the point (x,y) in the second image. Then, over all the points in the window W, we apply the difference formula:

$$F_{xy}(dx,dy) = \sum_{i,j=-1}^{i,j<2} \{f_1(x+i,y+j) - f_2(x+i+dx,y+j+dy)\}^2 \quad (1)$$

where $dx \in [dx_{min}, dx_{max}]$, $dy \in [dy_{min}, dy_{max}]$, and F_{xy} is the minimum disparity value associated with (x, y) using the disparity vector (dx, dy). By applying this formula, we can determine which point in the second image is the closest one to (x, y) in terms of intensity value. Here, we should keep track of the values of dx and dy which result in this minimum difference. The last formula is the basis in determining the interest operator in our work as we will see later.

3 The disparity pyramid

3.1 Pyramidal architecture rules

A pyramidal architecture is completely defined if we specify how a new level is constructed and how two cells at different levels are linked together. The generation rules of this pyramid are similar to those of the adaptive pyramid [6]. The decimation process is based on two principles:

- Two neighbours at a given level cannot both survive at the next level;
- For each nonsurviving cell , there exists at least one surviving cell in its neighbourhood.

3.2 New level construction

Let I_h be the image at level h in the pyramid. The neighbours of a cell (i, j, h) is the list of cells, $Brothers_{ijh}$, sharing links with it. The support, $Support_{ijh}$, of a cell (i, j, h) is defined as the set of all the neighbours or brothers of (i, j, h) in addition to the cell itself, i.e., $Brothers_{ijh} \cup (i, j, h)$. At the base of the pyramid (I_0 is the input image), $Support_{ij0}$ is the 3×3 square array centered on the cell. The receptive field of a cell (i, j, h), $Receptive_{ijh}$, is the list of all descendants located on the base of the pyramid. Each cell is associated with four important variables:

- Two state variables p_{ijh} and q_{ijh};
- The output F_{ijh} of the *interest operator* which is calculated as discussed earlier (Eq 1) where the lower the value F_{ijh}, the better the chance of this cell to survive. Note that in a stochastic pyramid, F_{ijh} is the outcome of a random variable while in the adaptive pyramid, F_{ijh} is the variance of the gray levels in the receptive field;
- The last variable is a 2D vector $(disp_x, disp_y)^T_{ijh}$ which determines the disparity values associated with F_{ijh} in both directions.

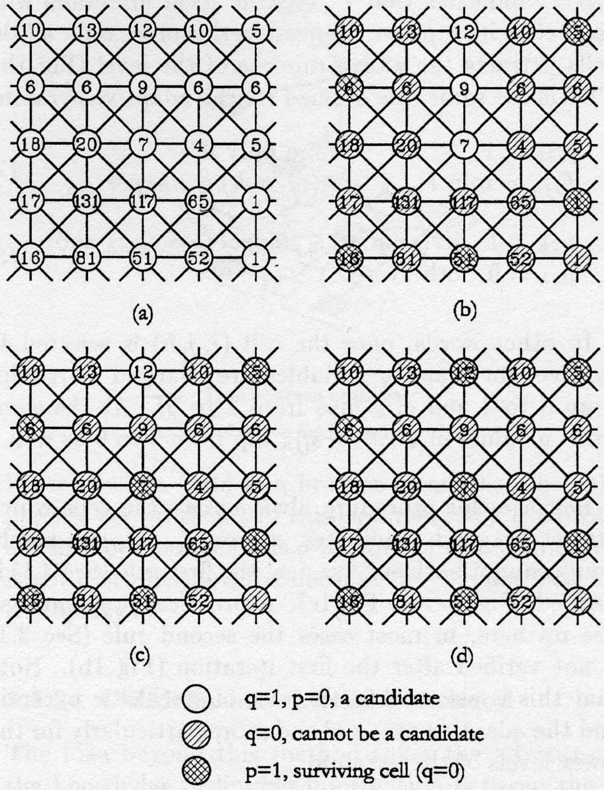

Figure 1: Decimation process: (a) initial graph with F_{ijh} values; (b) First iteration: global minima are extracted; (c) Second iteration: local minima are added; (d) Third iteration: local minima are added. *Note*: if q_{ijh}=0, this means that the cell cannot be chosen because either it has been chosen at a previous iteration or it is a brother of a surviving cell.

For the first level ($h = 0$), we considered the receptive field $Receptive_{ij0}$ of the cell $(i, j, 0)$ to be the cell itself. i.e., Eq 1 is applied only once for this cell. This is not the case for the upper levels where this formula is applied as many times as the number of cells of the receptive field.

3.3 The decimation process

The decimation process is the procedure by which the surviving cells are selected. We can define this process on the basis of the first three variables mentioned above, i.e., p, q, and F. An example is shown in fig 1.

First step in this procedure is to initialize the state variables p and q for each cell. Thus, for each cell (i, j, h), p_{ijh}=0 and q_{ijh}=1 as shown in Fig 1a. A cell (i, j, h) can survive to the next level $(h + 1)$ if the variable p_{ijh} is set to 1 during this procedure. The state p_{ijh}=1 indicates that F_{ijh} is the local minimum with respect to its support. Repeating this process for all the cells extracts the global minima of the level (Fig 1b). This is the same rule applied for the adaptive pyramid:

$$\begin{array}{ll} p_{ijh} = 1 & \text{if and only if} \\ F_{ijh} = \min F_{mnh} & \text{for}(m,n,h) \in Support_{ijh}. \end{array} \quad (2)$$

$$\begin{array}{ll} q_{ijh} = 1 & \text{if} \forall (m,n,h) \in Support_{ijh}, p_{mnh} = 0 \\ q_{ijh} = 0 & \text{otherwise} \end{array} \quad (3)$$

In other words, once the cell (i, j, h) is selected to survive, its p and q variables are changed (p_{ijh} flips from 0 to 1 and q_{ijh} flips from 1 to 0). At the same time, q values of $Brothers_{ijh}$ flip from 1 to 0 as well.

Because this procedure always skips the cells which cannot be good candidates, where q_{ijh}=0, two neighbours cannot both survive, i.e., the first rule (Sec 3.1) is verified (Fig 1b and Fig 1c). A practical problem may rise up here, in most cases the second rule (Sec 3.1) is not verified after the first iteration (Fig 1b). Note that this is also true in the case of stochastic pyramid and the adaptive pyramid and more particularly for the lower levels of the pyramid.

These cells, which have not been decided yet to survive or not, are indicated by the state variable q_{ijh} and the local minima of these cells are extracted :

$$\begin{array}{ll} p_{ijh} = 1 & \text{if and only if} \\ F_{ijh} = \min F_{mnh} & \text{for } (m,n,h) \in Support_{ijh}, \\ & \text{such that } q_{mnh} = 1. \end{array} \quad (4)$$

Steps 3 and 4 are repeated until there are no other cells to survive (Fig 1d). At this point, the second rule (Sec 3.1) is verified.

3.4 Inter-level communication

The parent at level $h+1$ is linked to its children at level h. In the disparity pyramid, as in the stochastic and the adaptive pyramids, the number of children depends on the information provided (random variable in the case of the stochastic pyramid or data-based in the case of the adaptive pyramid).

3.4.1 Which cell to choose?

As in the adaptive pyramid, we need a contrast measure to determine which two cells should be linked together. Note that, in the adaptive pyramid, the cell (i, j, h) is linked to the least contrasted surviving cell in its support. In the disparity pyramid, the disparity variation

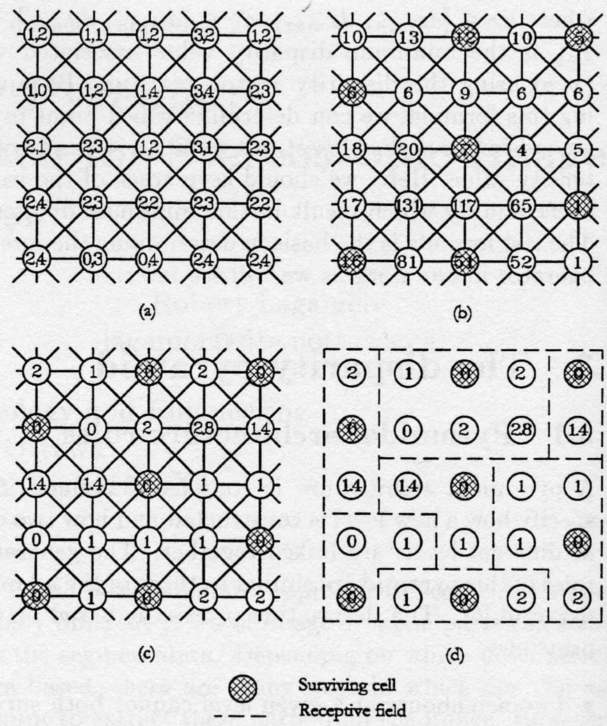

Figure 2: Disparity estimation: Which cell to choose? (a) Disparity vector values; (b) Minimum F values calculated as shown before and the choice of surviving cells is marked using the hatched circles; (c) Distance between each cell and the nearest surviving cell (in terms of disparity) in its neighbourhood; (d) Receptive field boundaries.

is used, i.e., the distance between disparity vectors of the two cells under consideration is calculated to help decide whether to choose the cell or not. Thus, the cell is linked to the nearest surviving cell (in terms of disparity). Here, the Euclidean's distance between their respective disparity vectors is computed.

A complete example of this procedure is shown in Fig 2. Fig 2a shows the disparity vectors associated with each cell. Fig 2b shows minimum values of F calculated using Eq 1 and the choice of surviving cells accordingly, while Fig 2c determines the Euclidean's distance between each cell and the nearest surviving cell in its neighbourhood. Finally, Fig 2d shows the receptive field boundaries for each surviving cell.

However, at this point, we may face a lot of cases in which we have to break these criteria. For example, if a nonsurviving cell has only one surviving cell in its support and the difference between them > *min_dist* (Sec 3.10), then we consider it as noise. This cell should be connected to the parent of the nearest brother (in terms of disparity).Again we may face another problem, if the parent of this brother is already created or not. That's why these locations, i.e., the noise, are indicated by the state variable $q = 2$ if the parent exists and $q = 3$ otherwise.

To link a nonsurviving cell (i, j, h) to (m, n, h)

$$\begin{array}{l} \text{if } dist(mnh, ijh) \leq min\_dist \\ \quad \text{if } parent_{mnh} \text{ exists} \\ \quad \text{then } q_{ijh} = 2 \\ \quad \text{else } q_{ijh} = 3; \end{array} \tag{5}$$

3.4.2 Linking cells

If (i, j, h) is a nonsurviving cell and (m, n, h) is the least contrasted surviving cell in its support $Support_{ijh}$, then:

$$\begin{array}{l} (i, j, h) \longrightarrow (m, n, h) \\ \text{iff } dist(ijh, mnh) = \min dist(ijh, klh) \\ \text{for } (k, l, h) \in Support_{ijh} \text{ s.t. } p_{klh} = 1 \end{array} \tag{6}$$

where

- $dist(ijh, mnh)$ is the minimum Euclidean's distance between $(disp_x, disp_y)_{ijh}$ and $(disp_x, disp_y)_{klh}$.

Always, the surviving cell (m, n, h) has at most one parent $(m, n, h+1)$ and at least one child $(m, n, h-1)$. Of course, there are no children for the cells in the base of the pyramid and no parents for those at the highest level. This cell is *not* linked to any other surviving cell at the same level.

3.5 New neighbourhood construction

There are two methods to construct the new neighbourhood, the first uses the coordinates of receptive field cells, and the other uses the children of the surviving cell.

3.5.1 Receptive-field-based neighbourhood

In this method, we should check the locations of the receptive field cells of both cells under consideration.

$$\begin{array}{l} \text{When } (I, J, 0) \in Receptive_{ijh} \\ \text{and } (M, N, 0) \in Receptive_{mnh} \\ \text{if } (|I - M| \leq 1) \text{ OR } (|J - N| \leq 1) \\ \text{then } (i, j, h) \text{ is a brother of } (m, n, h); \end{array} \tag{7}$$

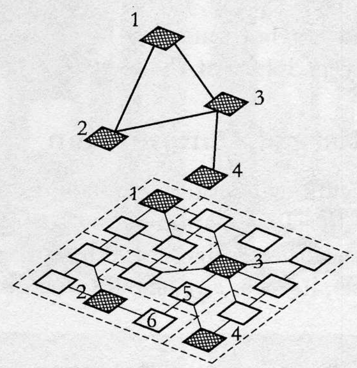

Figure 3: If a child is a brother of another then their parents are brothers. In this example cell 2 is a brother of cell 3 because cell 5 is a brother of cell 6 while cell 2 is not a brother of cell 4 because they do not have such a property.

3.5.2 Children-based neighbourhood

The idea beyond this method takes the advantage of the knowledge of linking information between the cell at the new level, the parent, and its children at the lower level.

$$\begin{array}{l} \text{When } (I, J, h-1) \text{ is a child of } (i, j, h) \\ \text{and } (M, N, h-1) \text{ is a child of } (m, n, h) \\ \text{if } (I, J, h-1) \text{ is a brother of } (M, N, h-1) \\ \text{then } (i, j, h) \text{ is a brother of } (m, n, h); \end{array} \tag{8}$$

This idea is illustrated in Figure 3.

3.6 State variables adjustment

After constructing the new neighbourhood $Brothers_{ijh}$ for each surviving cell (i, j, h) at the new level, we must recompute the values of the state variables p_{ijh} and q_{ijh} for (i, j, h) according to the new graph construction. Suppose that (m, n, h) is a neighbour of $(i, j, h,)$ then:

$$\begin{array}{l} \text{if } dist(ijh, mnh) < min\_dist \\ \text{then } p_{ijh} = 0 \text{ and } q_{ijh} = 1 \\ \text{else } p_{ijh} = 2 \text{ and } q_{ijh} = 0; \end{array} \tag{9}$$

where

- *min_dist*, (Sec 3.10), a parameter to the pyramid and its value is essential to distinguish between two different disparity vectors in the scene;

- $p_{ijh} = 2$ means that (i,j,h) is a root (Sec 3.8.1) and it should not be connected to any other cell as a child except itself (at the next level).

3.7 Statistical computation

Linking parents with children should, as expected, change the statistical parameters associated with the cell. Because of the change in the receptive field of the parent (i,j,h), its size, s_{ijh}, should change accordingly. Thus,

$$\begin{array}{c} \text{For } (m,n,h) \in Receptive_{ijh} \\ s_{ijh} = s_{ijh} + s_{mnh}; \end{array} \tag{10}$$

The other statistical parameter is the disparity values of the surviving cells. Based on the disparity range (dx_{min}, dx_{max}, dy_{min}, dy_{max}) (Sec 3.10) and the disparity equation (Sec 2), for every cell (i,j,h), we compute:

$$\begin{array}{l} \text{For } dx \in [dx_{min}, dx_{max}],\ dy \in [dy_{min}, dy_{max}] \\ \quad \text{For } (m,n,h) \in Receptive_{ijh} \\ \quad\quad F(dx,dy) = \sum_{(m,n,h)\in Receptive_{ijh}} F_{mnh}; \\ \\ F_{ijh} = \\ \min_{dx\in[dx_{min},dx_{max}],dy\in[dy_{min},dy_{max}]} F(dx,dy); \\ \\ (disp_x, disp_y)^T_{ijh} = (dx,dy)^T; \end{array} \tag{11}$$

where

- $F(dx,dy)$ is the sum of all disparity values associated with all cells $\in Receptive_{ijh}$. dx and dy indicate the disparity vector used to compute this value.
- $(disp_x, disp_y)^T_{ijh}$ is the new disparity vector associated with the cell(i,j,h) computed using the minimum value of the previous item, $F(dx,dy)$, over the whole disparity range.

3.8 Root extraction

3.8.1 What is the root?

At the highest level of the pyramid, one cell should exist for each disparity vector. This cell is called a *root*. Thus the root is a surviving cell all the way from the base that verifies a root predicate.

3.8.2 Root decision criteria

Like the adaptive pyramid we need procedures help decide if a cell is a root. Different factors should be considered. First, a root should be far enough from all the surviving cells in its support. This is decided with the help of min_dist parameter (Sec 3.10). In addition, its receptive field size must be large enough. This is achieved using min_size parameter (Sec 3.10). A small region is considered noise. Let (i,j,h) be a nonsurviving cell and (m,n,h) be the most similar surviving cell in its support. The cell (i,j,h) is called a root if and only if:

$$dist(ijh, mnh) > D(s_{ijh}) \tag{12}$$

where

- s_{ijh} is the size of (i,j,h) receptive field ($Receptive_{ijh}$);
- Function D is defined as:

$$D(s) = \begin{cases} min\_dist & \text{if } s > min\_size \\ min\_dist \times e^{\alpha a} & \text{otherwise} \end{cases} \tag{13}$$
$$\text{where } a = min\_size - s$$

3.8.3 Stopping criteria

All the above modules should be repeated iteratively until satisfying a certain stopping criterion. This is done by checking all surviving cells if they are turned into roots. This is an important property to this pyramid that all cells at the highest level should be roots, so for every cell (i,j,h) at the current level, we should check:

$$\begin{array}{l} \text{if } p_{ijh} = 2 \\ \text{then break the search and} \\ \quad \text{continue processing;} \end{array} \tag{14}$$

3.9 Practical problems

3.9.1 No decreasing function

One practical problem may rise up here if the input values cannot converge into one stable state. For example, if min_dist is not big enough to help decide whether the cell is a root or not, or to link it with another cell, this could lead to recomputing the same cells for the next level. Because we cannot control the inputs as they should vary according to the gray levels of the objects contained in the scene, we should force the processing to stop if the number of roots does not decrease.

3.9.2 More than one root

Even with this process, we may end up with more than one root for a unique disparity vector if all cells representing this vector disappear at the same level and/or min_dist is too small to distinguish the whole vector, rather it is distinguishing subparts from it.

Note that, the roots are detected at many levels according to the size of the their receptive fields. At the highest level of the pyramid, each root should represent one disparity vector of the scene.

3.10 Parameters to the pyramid

So we have 4 essential parameters to our pyramid:

- The value of α which is computed given the value of the size 1 (the highest contrast corresponds to the smallest size). In our experiments, we used α=0.2;
- The values of the disparity range (dx_{min}, dx_{max}, dy_{min}, dy_{max}) which specify the comparison window between both images.
- The value of the *min_dist* needed to distinguish between two different regions having different disparity vectors;
- The value of the *min_size* needed to decide whether or not this region is a real disparity vector or just a noise.

The second and the third parameters, i.e., the disparity range and the minimum distance depend on the location of the cameras, i.e., the view points of the scene. The last parameter, i.e., the minimum size depends on the size of the image.

4 Experimental results

To estimate the disparity between a pair of stereo images, the parameter *min_dist* is used to link regions with different disparity values under this threshold, i.e., < *min_dist*. The results are shown in two different ways. The first shows different point correspondences and the second presents a segmentation process of the scene according to the disparity values. Different types of stereo images are used including synthetic, outdoor and indoor scenes.

4.1 Point correspondence results

The test points are marked by the dotted white squares on both images. Fig 4 shows two examples of synthetic scenes. Notice that the second example represents a special case where the background is exactly the same in both images, i.e., the disparity vector of any point included in this background is $(0,0)^T$. However, this is not the case of real world scenes. The only difference between both images is the location of the small square which results in an occlusion area. Notice the error occurred in this area. An outdoor scene is presented in

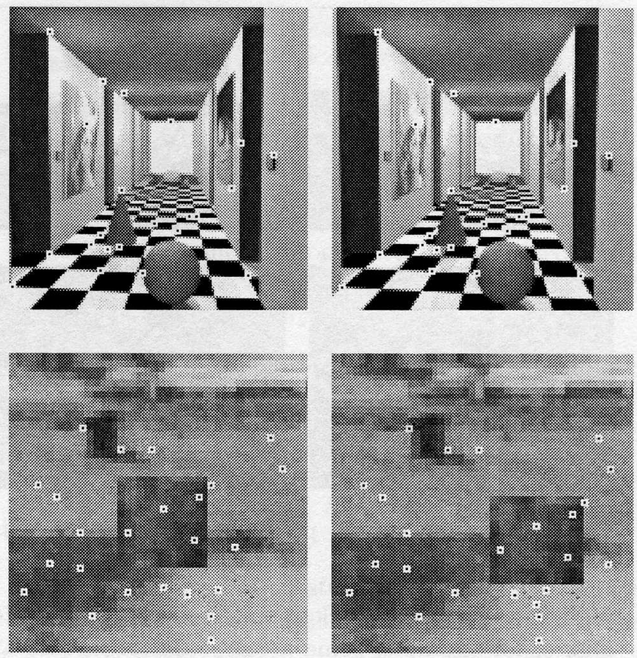

Figure 4: Synthetic scenes: point correspondence results.

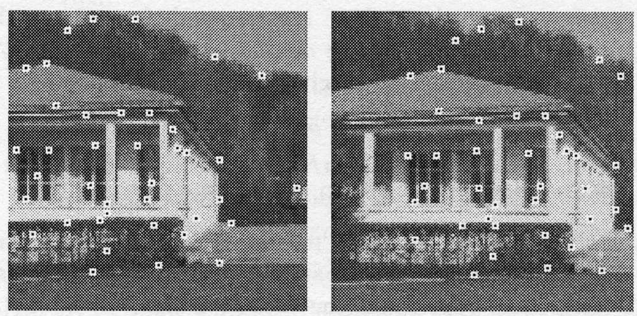

Figure 5: Outdoor scene (INRIA-Syntim © copyright): point correspondence results.

Fig 5 where some ambiguities take place. In this example, there is no distinction between the house and the trees located behind its ceiling. In other words, there is no enough features to distinguish between those two objects. This case is repeated again between the house and the fence in front of it which makes the fence as if it were a part of the elevation of the house. The disparity pyramid does not distinguish between different real objects included in the scene as long as their disparity values are the same, or they have a disparity difference $\leq$ *min_dist*. Again, we might expect some errors to appear in these areas.

In Fig 6, two pairs of indoor scenes are presented. The most important note to be mentioned here is the lack of features in some areas. Consider the first pair,

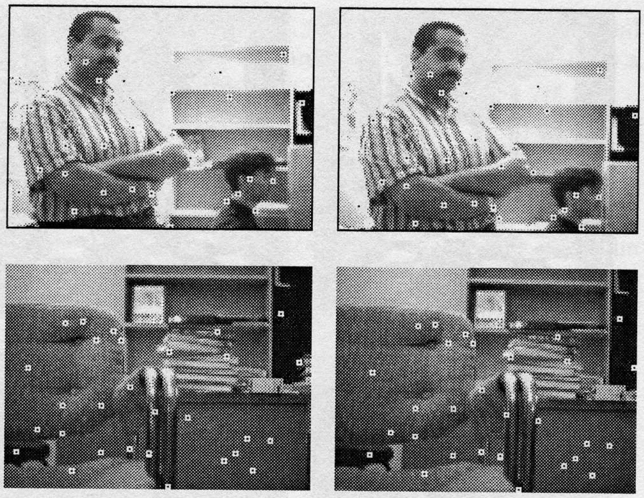

Figure 6: Indoor scenes: point correspondence results.

we notice that the background is overlighted which makes, for example, some areas of the book case appear as a continuity of the wall beside them. This results in a wrong perception and reduces the information needed to understand the scene. Consider the second pair, where there is no enough features on the wall behind the chair or on the back of the desk. As the previous case, we might expect some ambiguities in recognizing those areas.

4.2 Segmentation results

Each scene is segmented into layers of constant disparity values, or with disparity differences $\leq min\_dist$. Each layer represents the receptive field of a root at the highest level of the pyramid. Consider the second pair in Fig 7 where the segmentation process results in two layers of constant disparity values, as expected. The problem with the occlusion area is very clear in this example. A part of this area is linked to each of the two layers obtained. In Fig 8, as mentioned before, due to the lack of features, the fence and the tree area behind the ceiling are considered having the same disparity (or with a difference $\leq min\_dist$) as the house. The same problem appears in Fig 9 as expected, where some areas belonging to the wall and to the desk are incorrectly merged with the chair.

5 Conclusion

The results presented show that the proposed technique, could be used to estimate the disparity between a stereo pair of images. This is an irregular pyramid whose levels are built by successive extraction of surviving cells using the intensity difference values between

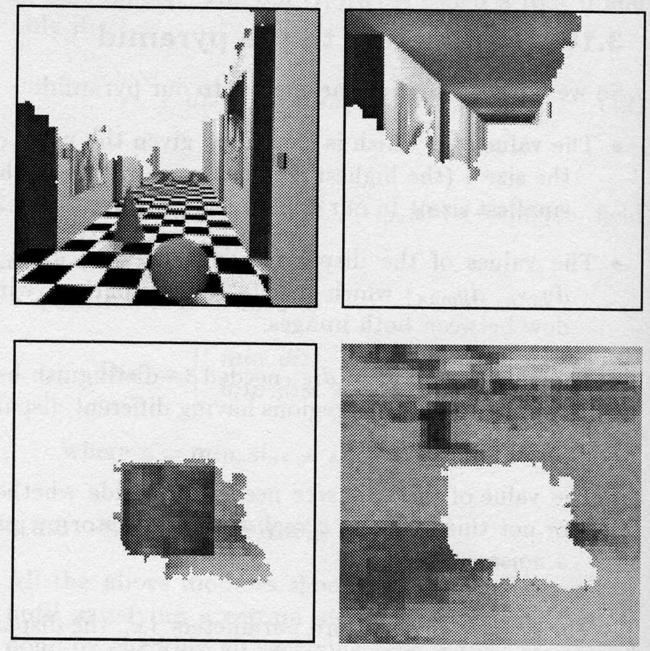

Figure 7: Synthetic scenes: segmentation results.

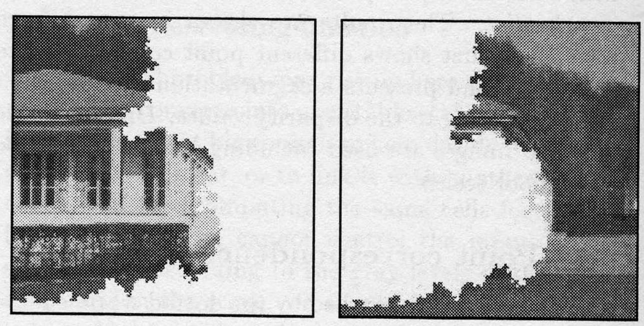

Figure 8: Outdoor scene (INRIA-Syntim © copyright): segmentation results.

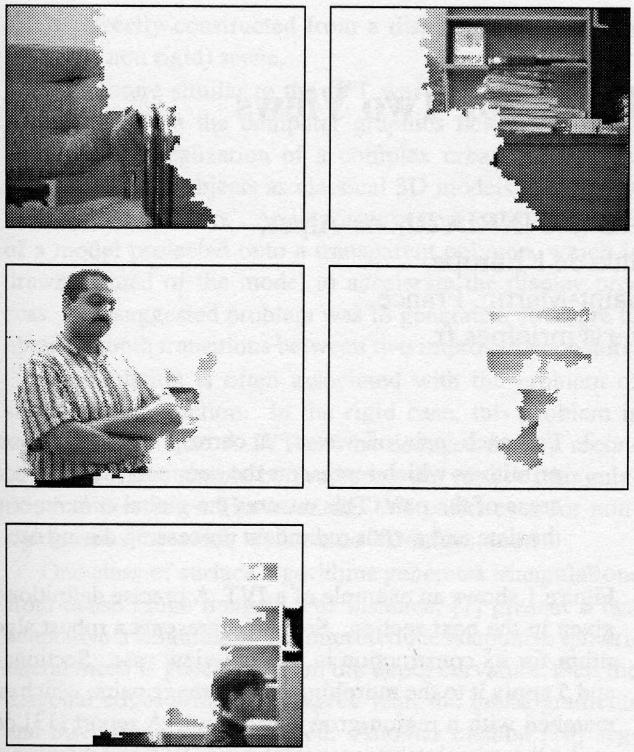

Figure 9: Indoor scenes: segmentation results.

the pair of images under consideration. This operation is limited within a disparity range including the whole possible set of disparity vectors of the scene objects. A nonsurviving cell is linked to the nearest surviving cell in its support. The distance between these two cells could be measured in terms of disparity variations. A root, associated with each component of the scene, should be far enough from all the surviving cells in its neighbourhood. The bottom-up process is implemented to extract the receptive field of each root.

References

[1] Nicholas Ayache. *Artificial Vision for Mobile Robots, Stereo Vision and Multisensory Perception.* The MIT Press, Cambridge, Massachusetts, 1991.

[2] Stephen Barnard. Stochastic stereo matching over scale. *International Journal for Computer Vision*, 3:17–32, 1989.

[3] Olivier Faugeras. *Three-Dimentional Computer Vision, A Geometric Viewpoint.* The MIT Press, Cambridge, Massachusetts, 1996.

[4] F. Heitz, P. Perez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *Image Understanding*, 59(1):125–134, 1992.

[5] Xiaoping Hu and Narendra Ahuja. Matching point features with ordered geometric, rigidity and disparity constrains. *IEEE Transactions on PAMI*, 16(10):1041–1049, October 1994.

[6] J.M. Jolion and A. Montavert. The adaptive pyramid: A framework for 2d image analysis. *CVGIP: Image Understanding*, 55(3):339–348, 1991.

[7] J.M. Jolion and A. Rosenfeld. *A Pyramid Framework for Early Vision.* Kluwer Academic Publishers, 1994.

[8] P. Limozin-Long. Vision stéréoscopique appliquée à la robotique. *Thèse, Université de Nice*, Octobre 1986.

[9] F. Lustman. Vision stéréoscopique et perception du mouvement en vision artificielle. *PhD thesis, Paris-Orsay*, 1987.

[10] P. Meer. Stochastic image pyramids. *Computer vision, Graphics, Image Processing*, 45:269–294, 1989.

[11] K. Nishihara and T. Poggio. Stereo vision for robotics. *Robotics Research, The First International Symposium*, pages 489–505, 1984.

[12] K. Ben Rhouma. Analyse de scènes par une méthode k2d en robotique avancée. *Thèse de Docteur Ingénieur, Université de Paris-Sud, Orsay*, 1984.

[13] A. R. Saint-Vincent. Perception et modélisation de l'environnement d'un robot mobile: une approach par stéréovision. *Thèse, Université Paul Sabatier, Toulouse*, 1986.

[14] Yoshihisa Shinagawa and Tosiyasu Kunii. Unconstrained automatic image matching using multiresolutional critical-point filters. *IEEE Transactions on PAMI*, 20(9):994–1010, September 1998.

[15] G. Toscani. Système de calibration optique et perception du mouvement en vision artificielle. *PhD Thesis, Paris-Orsay*, 1987.

[16] Guo-Qing Wei, Wilfried Brauer, and Gerd Hirzinger. Intensity- and gradient-based stereo matching using hierarchical gaussian basis functions. *IEEE Transactions on PAMI*, 20(11):1143–1160, November 1998.

[17] Juyang Weng, Narendra Ahuja, and Thomas Huang. Two-view matching. *ICCV'88*, pages 64–73, 1988.