Occlusions and Binocular Stereo

Davi Geiger[1], *Bruce Ladendorf*[1] *and Alan Yuille*[2]

[1] Siemens Corporate Research, 755 College Rd. East, Princeton NJ 08540.
[2] Division of Applied Sciences, Harvard University, Cambridge, MA, 02138. USA

Abstract.

Binocular stereo is the process of obtaining depth information from a pair of left and right cameras. In the past occlusions have been regions where stereo algorithms have failed. We show that, on the contrary, they can help stereo computation by providing cues for depth discontinuities.

We describe a theory for stereo based on the Bayesian approach. We suggest that a disparity discontinuity in one eye's coordinate system *always* corresponds to an occluded region in the other eye thus leading to an *occlusion constraint* or *monotonicity constraint.* The constraint restricts the space of possible disparity values, simplifying the computations, and gives a possible explanation for a variety of optical illusions. Using dynamic programming we have been able to find the optimal solution to our system and the experimental results support the model.

1 Introduction

Binocular stereo is the process of obtaining depth information from a pair of left and right camera images. The fundamental issues of stereo are: (i) how are the geometry and calibration of the stereo system determined, (ii) what primitives are matched between the two images, (iii) what *a priori* assumptions are made about the scene to determine the disparity and (iv) the estimation of depth from the disparity.

Here we assume that (i) is solved, and so the corresponding epipolar lines (see figure 1) between the two images are known. We also consider (iv) to be given and then we concentrate on the problems (ii) and (iii).

A number of researchers including Sperling[Sperling70], Julesz [Julesz71]; Marr and Poggio[MarPog76] [MarPog79]; Pollard, Mayhew and Frisby[PolMayFri87]; Grimson[Grimson81]; Ohta and Kanade[OhtKan85]; Yuille, Geiger and Bülthof[YuiGeiBul90] have provided a basic understanding of the matching problem on binocular stereo. However, we argue that more information exists in a stereo pair than that exploited by current algorithms. In particular, occluded regions have always caused difficulties for stereo algorithms. These are regions where points in one eye have no corresponding match in the other eye. Despite the fact that they occur often and represent important information, there has not been a consistent attempt of modeling these regions. Therefore most stereo algorithms give poor results at occlusions. We address the problem of modeling occlusions by introducing a constraint that relates discontinuities in one eye with occlusions in the other eye.

Our modeling starts by considering adaptive windows matching techniques [KanOku90], and taking also into account changes of illumination between left and right images, which provide robust dense input data to the algorithm. We then define an *a priori* probability for the disparity field, based on (1) a smoothness assumption preserving discontinuities, and (2) an occlusion constraint. This constraint immensely restrict the possible solutions of the problem, and provides a possible explanation to a variety of optical illusions that

so far could not be explained by previous theories of stereo. In particular , illusory discontinuities, perceived by humans as described in Nakayama and Shimojo [NakShi90], may be explained by the model. We then apply dynamic programming to exactly solve the model.

Some of the ideas developed here have been initiated in collaboration with A. Chamboll and S. Mallat and are partially presented in [ChaGeiMal91]. We also briefly mention that an alternative theory dealing with stereo and occlusions has been developed by Belhumeur and Mumford[BelMum91].

It is interesting to notice that, despite the fact that good modelling of discontinuities has been done for the problem of segmentation (for example, [BlaZis87][GeiGir91]), it is still poor the modeling of discontinuities for problems with multiple views, like stereopsis. We argue that the main difficulty with multiple views is to model discontinuities with occlusions. In a single view, there are no occlusions !

2 Matching intensity windows

We use adaptive correlation between windows. At each pixel, say l on the left, we consider a window of pixels that include l. This window is rectangular so as to allow pixels from above and below the epipolar line to contribute to the correlation (thereby discouraging mismatching due to misallignment of epipolar lines). The correlation between the left and right windows, $\| \mathbf{W}_l^L - \mathbf{W}_r^R \|$, is a measure of similarity. A major limitation of using large windows is the possibility of getting "wrong" correlations near depth discontinuities. To overcome this limitation we have considered two possible windows, one (window-1) to the left of the pixel l and the other (window-2) to the right (see figure 1). Both windows are correlated with the respective ones in the right image. The one that has better correlations is kept and the other one discarded. Previous work on adaptive windows is presented in [KanOku90].

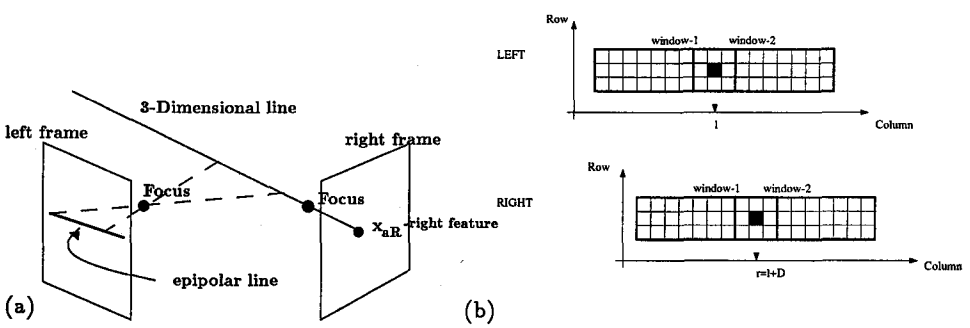

Fig. 1. *(a) A pair of frames (eyes) and an epipolar line in the left frame. (b) The two windows in the left image and the respective ones in the right image. In the left image each window shares the "center pixel" l. The window-1 goes one pixel over the right of l and window-2 goes one over left to l.*

2.1 Probability of matching

If a feature vector in the left image, say $\mathbf{W}_l^L$, matches a feature vector in the right image, say $\mathbf{W}_r^R$, $\| \mathbf{W}_l^L - \mathbf{W}_r^R \|$ should be small. As in [MarPog76][YuiGeiBul90], we use a

matching process M_{lr} that is 1 if, a feature at pixel l in the left eye matches a feature at pixel r in the right eye, and it is 0 otherwise. Within the Bayes approach we define the probability of generating a pair of inputs, $\mathbf{W}^L$ and $\mathbf{W}^R$, given the matching process M by

$$P_{input}(\mathbf{W}^L, \mathbf{W}^R|M) = e^{-\sum_{l,r}\left\{M_{lr}[\|\mathbf{W}_l^L - \mathbf{W}_r^R\|] + \epsilon(1-M_{lr})\right\}}/C_1 \qquad (1)$$

where the second term pays a penalty for unmatched points ($M_{lr} = 0$), with ϵ being a positive parameter to be estimated. C_1 is a normalization constant. This distribution favors lower correlation between the input pair of images.

2.2 Uniqueness and an occlusion process

In order to prohibit multiple matches to occur we impose that

$$\sum_{l=0}^{N-1} M_{l,r} = 0, 1 \quad \text{and} \quad \sum_{r=0}^{N-1} M_{l,r} = 0, 1\,.$$

Notice that these restrictions guarantee that there is at most one match per feature, and permits unmatched features to exist. There are some psychophysical experiments where one would think that multiple matches occur, like in the two bars experiments (see figure 5). However, we argue that this is not the case, that indeed a disparity is assigned to all the features, even without a match, giving the sensation of multiple matches. This point will be clearer in the next two sections and we will asume that *uniqueness* holds. Than, it is natural to consider an occlusion process, O, for the left (O^L) and for the right (O^R) coordinate systems, such that

$$O_l^L(M) = 1 - \sum_{r=0}^{N-1} M_{l,r} \quad \text{and} \quad O_r^R(M) = 1 - \sum_{l=0}^{N-1} M_{l,r}\,. \qquad (2)$$

The occlusion processes are 1 when no matches occur and 0 otherwise. In analogy, we can define a disparity field for the left eye, D^L, and another for the right eye, D^R, as

$$D_l^L(M)(1 - O_l^L) = \sum_{r=0}^{N-1} M_{l,r}(r - l) \quad \text{and} \quad D_r^R(M)(1 - O_r^R) = \sum_{l=0}^{N-1} M_{l,r}(r - l)\,. \quad (3)$$

where D^L and D^R are defined only if a match occurs. This definition leads to integer values for the disparity. Notice that $D_l^L = D^R_{l+D_l^L}$ and $D_r^R = D^L_{r-D_r^R}$. These two variables, $O(M)$ and $D(M)$ (depending upon the matching process M), will be useful to establish a relation between discontinuities and occlusions.

3 Piecewise smooth functions

Since surface changes are usually small compared to the viewer distance, except at depth discontinuities, we first impose that the disparity field, at each eye, should be a smooth function but with discontinuities (for example, [BlaZis87]). An effective cost to describe these functions, (see [GeiGir91]), is given by

$$U_{eff}(M) = 2\gamma - \sum_{l} ln(1 + e^{\gamma - \mu(D^L_{l+1} - D^L_l)^2}) - \sum_{r} ln(1 + e^{\gamma - \mu(D^R_{r+1} - D^R_r)^2})$$

where μ and γ are parameters to be estimated. We have imposed the smoothness criteria on the left disparity field and on the right one. Assigning a Gibbs probability distribution to this cost and combining it with (1), within the Bayesian rule, we obtain

$$P_{stereo}(M|\mathbf{W}^L, \mathbf{W}^R) = e^{-V_{eff}(M)}/Z \tag{4}$$

where Z is a normalization constant and

$$V_{eff}(M) = \sum_{lr} \Big\{ M_{lr} \Big[\| \mathbf{W}^L_l - \mathbf{W}^R_r \| - \frac{1}{N} ln(1 + e^{[\gamma - \mu(D^L_{l+1} - D^L_l)^2]}) - \frac{1}{N} ln(1 + e^{[\gamma - \mu(D^R_{r+1} - D^R_r)^2]}) \Big] + \frac{1}{N} \frac{\epsilon}{2} (O^L_l + O^R_r) \Big\}. \tag{5}$$

where we have discarded the constant $2\gamma + \epsilon(N-1)N$. This cost, dependent just upon the matching process (the disparity fields and the occlusion processes are functions of M_{lr}), is our starting point to address the issue of occlusions.

4 Occlusions

Giving a stereoscopic image pair, occlusions are regions in space that cannot be seen by both eyes and therefore a region in one eye does not have a match in the other image. To model occlusions we consider the matching space, a two-dimensional space where the axis are given by the epipolar lines of the left and right eyes and each element of the space, M_{lr}, decides whether a left intensity window at pixel l matches a right intensity window at pixel r. A solution for the stereo matching problem is represented as a path in the matching space(see figure 2).

4.1 Occlusion constraint

We notice that in order for a stereo model to admit disparity discontinuities it also has to admit occlusion regions and vice-versa. Indeed most of the discontinuities in one eye's coordinate system corresponds to an occluded region in the other eye's coordinate system. This is best understood in the matching space. Let us assume that the left epipolar line is the abscissa of the matching space. A path can be broken vertically when a discontinuity is detected in the left eye and, can be broken horizontally when a region of occlusion is found. Since we do not allow multiple matches to occur by imposing *uniqueness* then , almost always, a vertical break (jump) in one eye corresponds to a horizontal break (jump) in the other eye (see figure 2).

Occlusion constraint: A discontinuity in one eye correspond to an occlusion in the other eye and vice-versa.

Notice that this is not always the case, even if we do apply *uniqueness*. It can be violated and induces the formation of illusions which we discuss on the section 7.

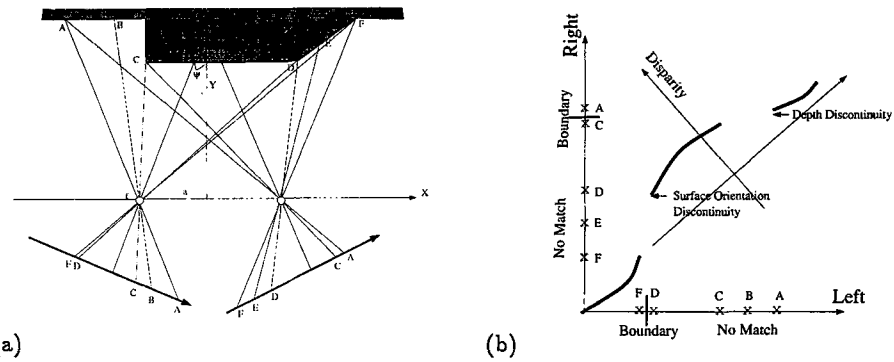

Fig. 2. *(a) A ramp occluding a plane. (b) The matching space, where the left and right epipolar lines are for the image of (a). Notice the symmetry between occlusions and discontinuities. Dark lines indicates where match occurs,* $M_{lr} = 1$.

4.2 Monotonicity constraint

An alternative way of considering the *occlusion constraint* is by imposing the monotonicity of the function $F_l^L = l + D_l^L$, for the left eye, or the monotonicity of $F_r^R = r + D_r^R$. This is called the *monotonicity constraint* (see also [ChaGeiMal91]). Notice that F_l^L and F_r^R are not defined at occluded regions, i.e. the functions F_l^L and F_r^R do not have support at occlusions. The monotonicity of F_l^L, for an occlusion of size o, is then given by

$$F_{l+o+1}^L - F_l^L > 0, \quad \text{or} \quad D_{l+o}^L - D_l^L > -o, \quad \forall\, l$$
$$\text{where} \quad O_{l+o+1}^L = O_l^L = 0 \qquad \text{and} \qquad \sum_{l'=l+1}^{l+o} (1 - O_{l'}^L) = 0 \tag{6}$$

and analogously to F_r^R. The *monotonicity constraint* propose an ordering type of constraint. It differs from the known *ordering constraint* in that it explicitly assumes (i) occlusions with discontinuities, horizontal and vertical jumps, (ii) uniqueness. We point out that the monotonicity of F^L is equivalent to the monotonicity of F^R. The *monotonicity constraint* can be applied to simplify the optimization of the effective cost (5) as we discuss next.

5 Dynamic Programming

Since the interactions of the disparity field D_l^L and D_r^R are restricted to a small neigborhood we can apply dynamic programming to exactly solve the problem.

We first constrain the disparity to take on integral values in the range of $(-\theta, \theta)$ (Panum's limit, see [MarPog79]). We impose the boundary condition, for now, that the disparity at the end sides of the image must be 0.

The dynamic program works by solving many subproblems of the form: what is the lowest cost path from the beginning to a particular (l, r) pair and what is its cost? These

subproblems are solved column by column from left to right finally resulting in a solution of the whole problem (see figure 5). At each column the subproblem is considered requiring a set of subproblems previously solved. Because of the *monotonicity constraint* the set of previously solved subproblems is reduced. More precisely, to solve the subproblem (l, r), requires the information from the solutions of the previous subproblems (x, y), where $y < r$ and $x < l$ (see shaded pixels in figure 5). Notice that the *monotonicity constraint* was used to reduce the required set of previously solved subproblems, thus helping the efficiency of the algorithm.

6 Implementation and Results

A standard image pair of the Pentagon building and environs as seen from the air are used (see figure 3 (a) and (b)) to demonstrate the algorithm. Each image is 512 by 512 8-bit pixels. The dynamic programming algorithm described above was implemented in C for a SPARCstation 1+; it takes about 1000 seconds, mostly for matching windows ($\approx$ 75% of the time). The parameters used were : $\gamma = 10$; $\mu = 0.15$; $\epsilon = 0.15$; $\theta = 40$; $\omega = 3$; and the correlation $\| \mathbf{W}_l^L - \mathbf{W}_{l+DL_l}^R \|$ has been normalized to values between 0 and 1. The first step of the program computes the correlation between the left and right windows of intensity. Finally the resulting disparity map is shown in figure 3. The disparity values changed from -9 to $+5$.

The basic surface shapes are correct including the primary building and two overpasses. Most of the details of the courtyard structure of the Pentagon are correct and some trees and rows of cars are discernible. As an observation we note that the disparity is tilted indicating that the top of the image is further away from the viewer than the bottom. Some pixels are labeled as occluded and these are about where they are expected (see figure 3).

7 Illusions and disparity at occlusions

In some unusual situations the *monotonicity constraint* can be broken, still preserving the *uniqueness*. We show in figure 4 an example where a discontinuity does not correspond to an occlusion. More psychophysical investigation is necessary to asserts an agreement of the human perception for this experiment with our theory. This experiment is a generalization of the double-nail illusion [KroGri82], since the head of the nail is of finite size (not a point), thus we call it the double-hammer illusion.

7.1 Disparity limit at occluding areas and Illusory discontinuities

At occluded regions there is no match and thus we would first think not to assign a disparity value. Indeed, according to (3) and (5) a disparity is just defined where a match exist, and not at occlusions. However, some experiments suggest that a disparity is assigned to the occluded features, like in the two-bars experiment illustrated in figure 5.

The possible disparity values for the occluded features are the ones that would break the *monotonicity constraint*. This is known as Panum's limit case. Nakayama and Shimojo [NakShi90] have shown that indeed a sensation of depth is given at the occluded features according to a possible limit of disparity. If indeed, a disparity is assigned to the occluded regions than a disparity discontinuity will be formed between the occluded and not occluded regions. We have produced a variation of the Nakayama and Shimojo experiments where indeed a sensation of disparity at occluded features and illusory contours are produced (see figure 5). We then make the following conjecture

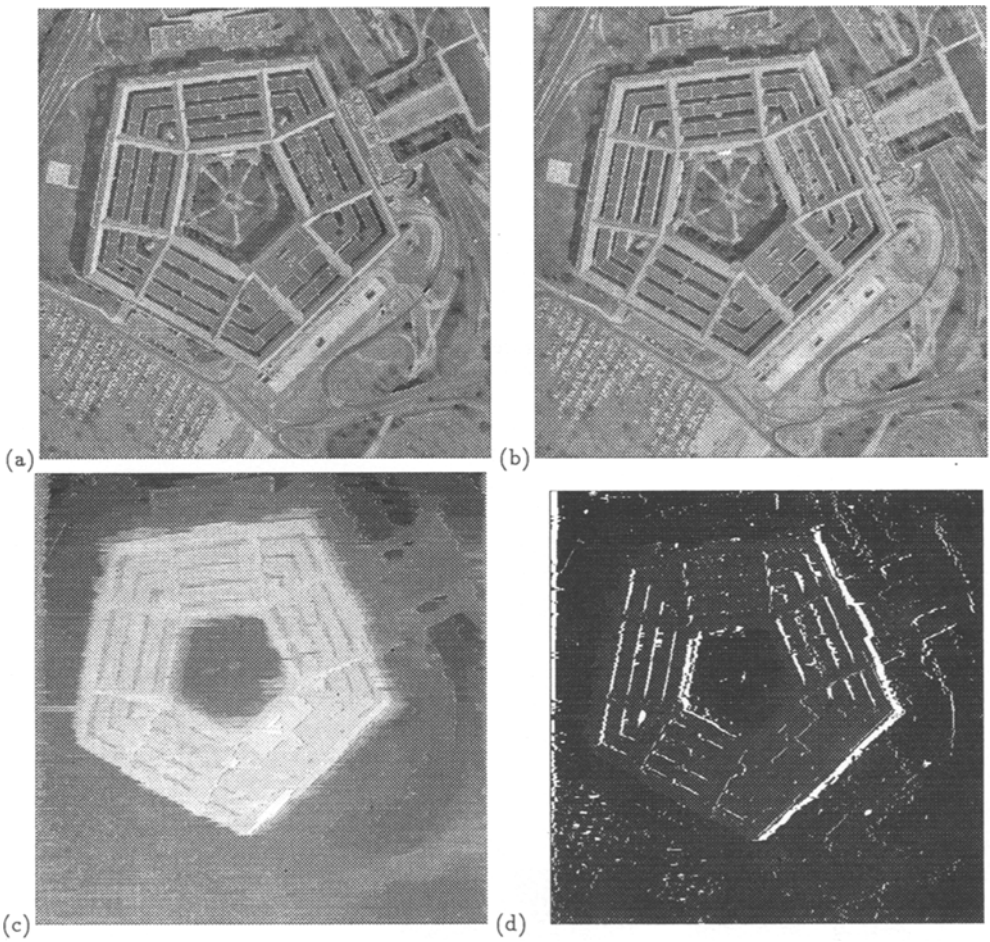

Fig. 3. *A pair of (a) left and (b) right images of the pentagon, with horizontal epipolar lines. Each image is 8-bit and 512 by 512 pixels. (c) The final disparity map where the values changed from -9 to $+5$. The parameters used where: $\gamma = 10$; $\mu = 0.15$; $\epsilon = 0.15$;$\theta = 40$; $\omega = 3$. In a SPARCstation 1+, the algorithm takes about 1000 seconds, mostly for matching windows ($\approx$ 75 of the time). (d) The occlusion regions in the right image. They are approximately correct.*

Conjecture 1 (occluded-disparity) *The perceived disparity of occluded features is the limit of their possible disparity values (Panum's limit case), if no other source of information is given.*

This conjecture provides a method, that we have used, to fill in the disparity for occluded features without having to assing a match.

Acknowledgements: We would like to thank A. Chamboll and S. Mallat for the stimulating conversations, and for their participation on the initial ideas of this paper and D. Mumford for many useful comments.

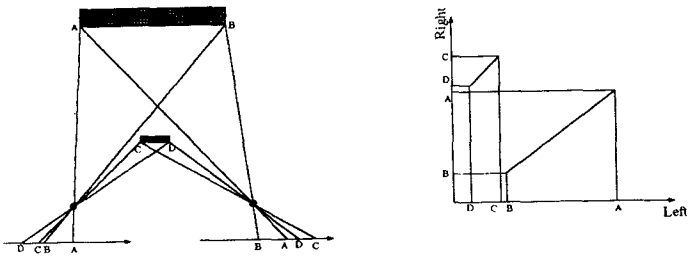

Fig. 4. The double-hammer illusion. This figure has a square in front of another larger square. There is no region of occlusion and yet there is a depth discontinuity.

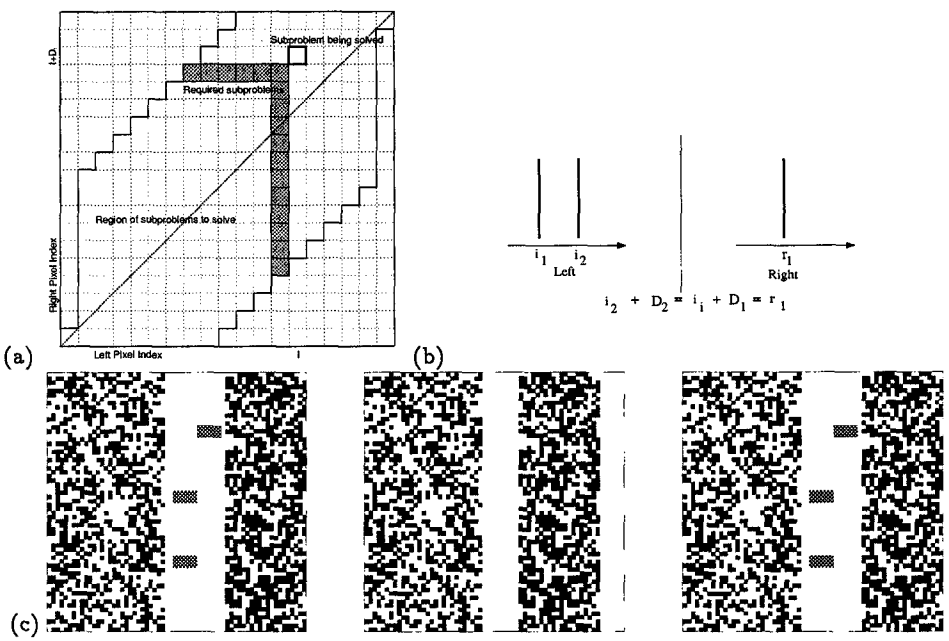

Fig. 5. *(a) An illustration of the dynamic programming. The subproblem being considered is the* $(l, l + D_l^L)$ *one. To solve it we need the solutions from all the shaded pixels. (b) When fused, a 3-dimensional sensation of two bars, one in front of the of the other one, is obtained. This suggests that a disparity value is assigned to both bars in the left image. (c) A stereo pair of the type of Nakayama and Shimojo experiments. When fused, a vivid sensation of depth and depth discontinuity is obtained at the occluded regions (not matched features). We have displaced the occluded features with respect to each other to give a sensation of different depth values for the occluded features, supporting the disparity limit conjecture. A cross fuser should fuse the left and the center images to preceive the blocks behind the planes. An uncross fuser should use the center and right images.*

References

[BelMum91] P. Belhumeur and D. Mumford, *A Bayesian treatment of the stereo correspondence using half-occluded region*, Harvard Robotics Lab, Tech. Repport: December, 1991.

[BlaZis87] A. Blake and A. Zisserman, *Visual Reconstruction*, Cambridge, Mass: MIT Press, 1987.

[ChaGeiMal91] A. Champolle, D. Geiger, and S. Mallat, "Un algorithme multi-échelle de mise encorrespondance stéréo basé sur les champs markoviens," in *13th GRETSI Conference on Signal and Image Processing*, Juan-les-Pins, France, Sept. 1991.

[GeiGir91] D. Geiger and F. Girosi, "Parallel and deterministic algorithms for mrfs: surface reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-13, no. 5, pp. 401–412, May 1991.

[Grimson81] W. E. L. Grimson, *From Images to Surfaces*, Cambridge, Mass.: MIT Press, 1981.

[Julesz71] B. Julesz, *Foundations of Cyclopean Perception*, Chicago: The University of Chicago Press, 1971.

[KanOku90] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiments," in *Proc. Image Understanding Workshop DARPA*, PA, September 1990.

[KroGri82] J.D. Krol and W.A. Van der Grind, "The double-nail illusion: experiments on binocular vision with nails, needles and pins.," *Perception*, vol. 11, pp. 615–619, 1982.

[MarPog79] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proceedings of the Royal Society of London B*, vol. 204, pp. 301–328, 1979.

[MarPog76] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283–287, 1976.

[NakShi90] K. Nakayama and S. Shimojo, "Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points," *Vision Research*, vol. 30, pp. 1811–1825, 1990.

[OhtKan85] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 2, pp. 139–154, 1985.

[PolMayFri87] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "Disparity gradients and stereo correspondences," *Perception*, 1987.

[Sperling70] G. Sperling, "Binocular vision: a physical and a neural theory.," *American Journal of Psychology*, vol. 83, pp. 461–534, 19670.

[YuiGeiBul90] A. Yuille, D. Geiger, and H. Bulthoff, "Stereo, mean field theory and psychophysics," in *1st. ECCV*, pp. 73–82, Springer-Verlag, Antibes, France, April 1990.

This article was processed using the LaTeX macro package with ECCV92 style