Recursive Perceptual Grouping for 3D object reconstruction from 2D scenes

Laurent Alquier and Philippe Montesinos
LGI2P - Parc scientifique G. BESSE
NIMES, F-30000
alquier@eerie.eerie.fr

Abstract

The detection and the identification of an object in an image is an extremely complex combinatory problem. The objective of a shape recognition system is to reduce this complexity and make the problem solvable. In this paper, we present a two stage method for features extraction. First we start with a description of the scene using Perceptual Organization techniques. This step extracts a set of salient structures in the image, from a local to gobal level, without prior knowledge of the scene.

We show how this description is more appropriate for the shape extraction level. Along the analysis, increasingly complex features are extracted from this description, organized into a graph of structural relationships, and validated using the original image.

The final description of objects in the image can be used for matching a 3D model with a 2D scene or used to track objects in a sequence.

Keywords: Perceptual Grouping, Shape extraction, Dynamic Programming, Curves Detection.

1 Introduction

Shape recognition is one of the fundamental problems in Computer Vision. The interpretation of an image in terms of objects in a scene and the determination of relationships between these objects is a difficult task of Non Polynomial complexity when applied directly to images of low level primitives. Though it is well known and solved in particular cases, such as industrial applications with well defined objects and environments, it remains an open problem in less well defined systems. In order to reduce this complexity, visual systems need to organize the information embedded in pictures into a representation suitable for high level tasks. Many approaches exist depending on their level of application. Methods such as *simulated annealing* [20] or *Markov Random Fields* [17] operate directly on images after low level processings (if not directly on the original image). At the opposite extreme, methods of *statistical classification* [3] or *graph matching* [14] are more suitable for intermediate or high level descriptions.

To help with the reconstruction of a scene, many works have shown the importance of an intermediate level of scene description. In 1976, Marr [9] suggested the idea of a "primal sketch" involving both information from contour segmentation and the grouping of primitives such as curves or lines. Lowe [8] associated such groupings with a statistic feature, arguing that meaningful structures are unlikely to arise from the image by accident. Lowe's work showed how Perceptual Grouping could be used to efficiently structure images produced by inaccurate or biased low-level processing. Our work uses this assumption to effectively prune the search space of curves in images. Our work is also closely related to Ballard's statement [1] that the world has a predictable structure to a certain extent and can be used as an external memory in a system of *purposive* or *animate vision*.

Taking its roots in Gestalt Psychology [19], Perceptual Grouping in Computer Vision has been subject to various approaches. The objective is to reproduce the way human vision structures the representation of information in a picture before starting to interpretation it. Very simple psychovisual experiments clearly show the importance of

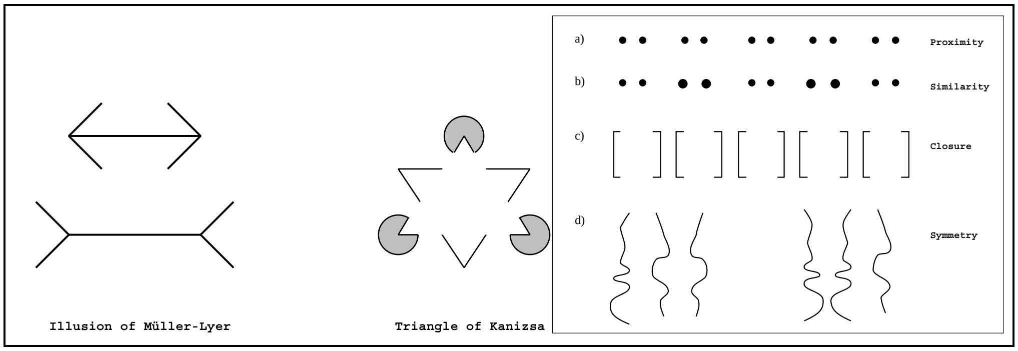

Figure 1: Examples of psychovisual experiments and Influence of the context over perception

this mechanism, as well as the influence of relationships between objects in a picture (Fig 1). These relationships can range from continuation or symmetry to similarity or object-background separation.

This complex combinatorial problem is mostly easily expressed as the grouping together of points of interest into larger structures (such as chains or regions). Other approaches are possible, according to the nature and complexity of the primitives used for grouping. These primitives can be grouped according to undetermined shapes (such as segments, curves, or regions) or parametric shapes (circles, squares, ellipses). Finally, the groupings are used to more easily initiate a model-based shape recognition system. Very few attempts have been made to group larger structures.

From a more practical point of view, two major approaches have been used to solve the problem of Perceptual Grouping. On one hand, algorithmic methods such as the construction of graphs based on geometric properties [13] have been used. On the other hand, optimization techniques, such as those suggested by Lowe have been favored. These approaches involve functions which take into account the saliency of curves in the same way the human eye would do. They can be either *static* [16] [5] [4] or *dynamic* [12] [7] [6].

The method described in this paper falls within the scope of Perceptual Organization with dynamic optimization techniques. Its purpose is to extract characteristic features from the image and give a useful description of the scene for a higher object recognition process. First in section 2, we describe how to extract the salient structures from images after the detection of contour or crest lines. This section discusses the choice of the visual qualities emphasized by groupings, the formalism to embed them into a quality function and the recursive scheme to optimize it. The optimal groupings are then selected according to their global quality. Section 3 gives an overview of the shape extraction process and presents some early results of our research. Finally, we discuss the applicability of our method to future situations.

2 Recursive Perceptual Grouping

As stated previously, the purpose of this preliminary step is to extract from the initial scene the most salient structures. By this preliminary extraction of these important stuctures, we are able to considerably reduce the complexity of recognizing and describing the objects of interest in the scene.

Low level processing can provide very different primitives for grouping ; regions, contours, crest-lines, and corners. Because structural relationships are well defined for curves (curvature, continuity, convexity), we chose to focus on the extraction of salient curves from pictures after the detection of either contours or crest-lines. A description of the scene based on curves is also particularly well adapted to the further use of objects models (usually described by vertices, edges and curves). Additionally, the use of such groupings efficiently solves problems arising from gaps and false detections due to noise or junctions in low level processings.

The high quality of the results for the grouping of pixels shows how robusts the method is to noise [11]. We have tested it on groupings of pixels and adapted it to the grouping of chains of pixels. We have done this in order to reduce time and computer resources and make it more suitable for practical applications.

In order to evaluate a possible grouping by a quality function, it is necessary to define the visual properties a grouping has to comply with. Each quality term represents a geometric property, which is expected to be high for a "good" grouping between a primitive and its neighbours. The kind of property suggested by the groupings depends on the type of scene ; for example, in the case of satellite pictures, roads are expected to be long, continuous and smooth curves, whereas straight lines and corners would be more expected for indoor scenes. It also depends on the choice of primitive to group together (the length of chains is an important factor for the grouping of chains, while there is no such information available for the grouping of pixels).

A parallel with the energy functions used in *active contours* or "snakes" gives a good formalism for the choice of the quality terms. A quality function composed of an internal term, representing the inner shape of the grouping, and an external term, representing the influence of the image on the grouping, increases the stability of the results. However, in contrast to the snakes approach, our optimization scheme is done locally and does not need to be initialized close to a solution.

For example, during a grouping of contour chains, external influences of the image are functions of the length of chains and the distance between the extremes of the chains to group together. The shape constraint is represented by the curvature of the link between the chains (a polynomial curve generated with respect to the orientations of each extreme - see Fig.3 about the grouping of chains).

The complete quality function is defined as a linear combination of the quality terms. Each quality term is normalized in order to allow a better control of their influence by a parameter relative to each other.

2.1 Definition of the quality function

It has been shown [15] that for a certain class of functions (known as "extensible functions"), each term can be written in a recursive way (a detailed discussion about the quality terms and algorithms can be found in [11]). Each quality term is written as a bi-lateral function associating a primitive to a pair of its neighbours, in order to define a trace *coming in* and *going from* this primitive (Fig.3).

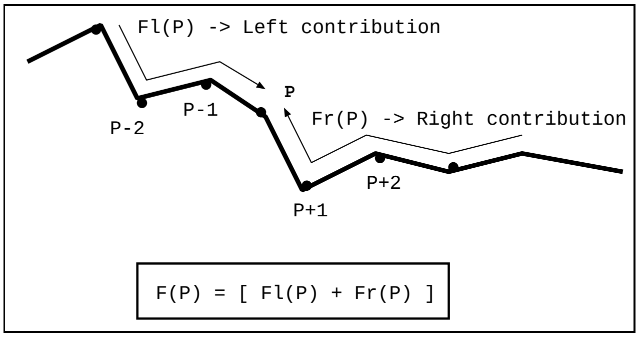

Figure 2: Notations used for a quality term on a dynamic curve during the optimization

A quality term $\mathcal{F}$ of a curve arriving in a pixel P is defined as the sum of local terms along the trace of the curve entering in and the curve exiting from P, with a factor $0 \leq \rho \leq 1$ representing the attenuation of the quality with distance. If we write the relation as a bi-lateral function of the trace, with a trace $\mathcal{F}_l(P)$ coming in P and a trace $\mathcal{F}_r(P)$ going from P, the quality becomes:

$$\mathcal{F}(P) = (\mathcal{F}_r(P) + \mathcal{F}_l(P) + \mathcal{H}(P)) \tag{1}$$

In general, $\mathcal{F}_r(P)$ and $\mathcal{F}_l(P)$ can be written as follows (but this is not the only possible expression of *extensible* functions) :

$$\mathcal{F}_l(P) = \ Q(P) + \rho \cdot Q_P(P-1) + \rho^2 \cdot Q_{P-1}(P-2) + ... \tag{2}$$

and :

$$\mathcal{F}_r(P) = \ Q(P) + \rho \cdot Q_P(P+1) + \rho^2 \cdot Q_{P+1}(P+2) + ... \tag{3}$$

$\mathcal{H}(P)$ is a function of correction [10] of the local quality in P (this function is necessary to compensate for the sum of $Q(P)$ coming from each contribution).

This expression written in a recursive way gives, for a grouping a of length n starting from P:

$$\mathcal{F}_l{}^{(n)}(P) = Q_P(P) + \rho \cdot \mathcal{F}_l{}^{(n-1)}(P-1) \tag{4}$$

where $Q(P)$ is the local quality term for the primitive P and $Q_P(P-1)$ represents the evaluation of a contribution from the primitive $(P-1)$ viewed from P. Each term of this quality function is representative of a long distance measure of the quality of the curve as well as a local measure of this quality.

2.2 Optimization of the groupings

The quality function is optimized iteratively, from a local to a global level, using a method related to Dynamic Programming [2] [15].

For each primitive, a connection is defined by a pair of neighbours representing the directions of arrival and departure for a possible grouping between the primitives. The pair of neighbours giving the best value for the quality function is selected. The recursive expression of the quality terms makes it possible to compute their values with a local part (defined by the local characteristics of the connection) and a global contribution provided by each neighbour. Along the iterations, the importance of individual primitives decreases with regard to primitives included in large structures.

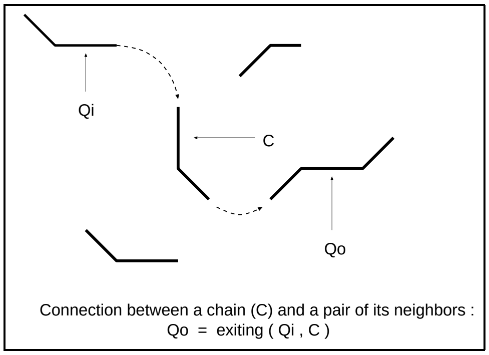
Qi
C
Qo
Connection between a chain (C) and a pair of its neighbors :
Qo = exiting (Qi , C)

Figure 3: Example of a connection between a chain and two neighbours

The algorithmic complexity is directly related to the number of objects of interest in the image after low level processing, as well as the average number of neighbours around each object. It is the order of $O(k \cdot N \cdot n)$, if N is the length of the expected groupings (it is also the number of iterations), n the number of primitives to group together and k the average number of neighbours around each primitive. In the case of grouping of pixels, the optimization is applied to each pixel in the image, with a constant neighbourhood. In the case of grouping of chains, the grouping process is reduced to the number of chains in the image, with a dynamic number of neighbours (depending on a fixed area around each extreme of the chains).

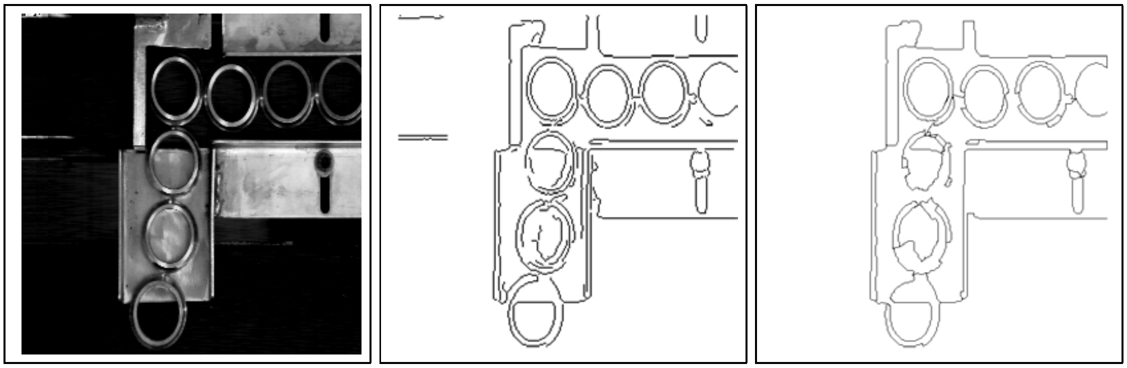

Figure 4: Example of some salient groupings from a contour detection

During the optimization process, the number of iterations is related to the distance between contour elements we want to connect. For example, in the case of pixels, the length of the widest gap gives us the minimum of iterations required to fill gaps along the curves of the image. We must remember that the optimized groupings tend to be smoothed as they receive more global contributions. Thus, a high number of iterations means a loss in the precision of the selected curves. Like the complexity, the time required for this process is directely related to the number of primitives to group together (on a SPARCstation 5, this time is about 10 seconds per iteration for 1000 chains to group - the optimization requires about 20 iterations in most situations). The parameters involved into the optimization process are the coefficient of each contribution in the Quality function (see 2.1) and the number of iterations.

Once the optimization has been performed, the curves are extracted by following the connections from one primitive to another. The number of possible groupings on the image is reduced to a single optimized grouping for each possible starting point. Primitives of high local quality are most likely to belong to large structures; they give the first approximation of a suitable solution. This selection is refined with regard to the global quality of the solutions. The global quality of a grouping is the sum of local qualities for each of its points. This definition divides the possible solutions into classes of groupings with equivalent qualities, thus reducing the amount of possibilities to a few classes.

3 Shape extraction

We showed how Perceptual Grouping can provide a reduced set of salient structures to the shape extraction level. The number of candidates is bounded by the number of primitives in the image before grouping. Unlike the chains produced by the classical piecewise polygonal approximation, groupings after optimization and selection are longer, uniform, smooth, and allow curves.

The purpose of this second step is to extract relational structures from the optimized groupings and build a hierarchy of primitives and relationships between them as suggested by Horaud and Skordas [13]. However, instead of building a relational graph from a polygonal approximation of the edge elements, our scheme build the graph incrementally, by creating hypotheses from the groupings and validating these hypotheses by comparing them to the original image.

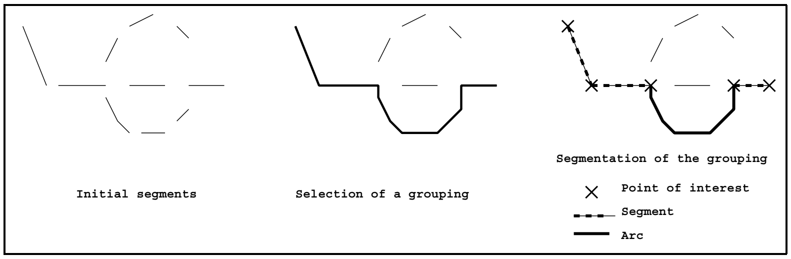

Figure 5: Selection and segmentation of a grouping

3.1 Segmentation into local hypotheses

Firstly, the groupings are segmented into consistent primitives. By order of complexity, the first primitives detected on the groupings are *points of interest* : endpoints (end of a curve or a segment), simple junctions (colinearity, tangency, inflexion), multiple junctions (corner, occlusion). These points correspond to vertices in the final description of the scene. The other primitives represent the way in which these points are connected to each other : *segments* (straight lines), *arcs* (smooth curvature) or *undetermined curves.*

The analysis of local relationships between consecutive primitives leads to the creation of local hypotheses about the geometric shape to which they belong. For example, from the configuration *Segment-Segment*, it is possible to draw the hypothesis of a straight line (if the segments are colinear), rectangles (if symmetrical), convex polygons or undetermined polygons. In the same way, the configuration *Arc-Arc* generates a hypothesis of a circle or ellipse (in the case of a convex curve) or a general curve. Figure 5 shows an example of segmentation of a grouping into segments and an arc.

The algorithm of segmentation is inspired by a method of polygonal approximation [18]. Each grouping is evaluated sequentially and divided according to the most important points of interest (that is, the points corresponding to important changes of curvature and orientation). Other cues can be used in this step to improve the segmentation; for example, we also used corner detection from a low level processing. The complexity of this algorithm is directely related to N the number of groupings and k the average number of primitives found for each grouping. This complexity is the order of $O((k \cdot N)^2)$. Within this context, the time required by the algorithm depends on the algorithms used to identify the geometric primitive from each division in the polygonal approximation (the implementation showed that most of the computation time is taken by the identification of the arcs of ellipses).

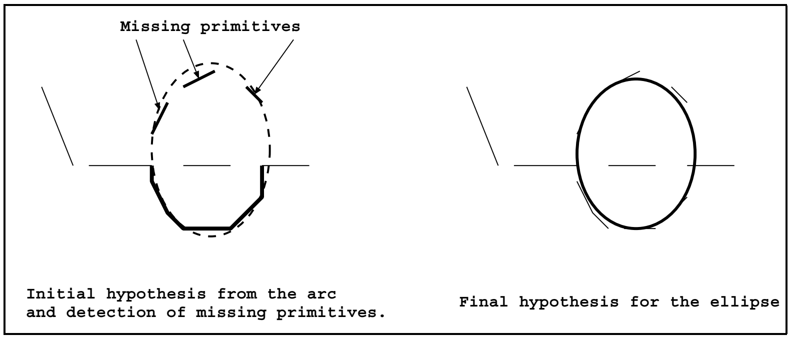
Missing primitives
Initial hypothesis from the arc
and detection of missing primitives.
Final hypothesis for the ellipse

Figure 6: Validation of an ellipse from the hypotheses of an arc

3.2 Results and perspectives

Figures 7 and 9 show some results for ellipse detection in two scenes using this method. Each example shows the original picture in grey levels, the edge detection used for the Perceptual Grouping, some of the main groupings after optimization and the shapes extracted from these groupings. The main groupings have been selected manually from the classes of solutions provided by the optimization level (semi automatic selection). They show clearly how a few groupings only are enough to extracts the ellipses, as opposed to the complexity of the this extraction from the edges only.

In figure 7, 23 groupings only have been used from the 251 chains created by the contour detection. 6 out of the 15 ellipses have been accurately found. 3 ellipses were not accurately found due to errors in the edge detection. The arcs have been detected in addition to the ellipses - they cover parts of the missing ellipses and can be used as starting points for a higher level of shape extraction.

In figure 9, the edge detection created 431 chains. A single grouping has been enough to find the ellipse.

Finally, each hypothesis of the geometric shape drawn from a primitive is compared to the original primitives for verification. According to the number of pixels from the contour which are matched by the shape (using a least squares optimization), the hypothesis is validated and then evaluated again, taking into account the chains matched.

In Figure 6, the initial ellipse is compared with neighboring segments from the original edge detection. The ellipse is re-evaluated for each segment within a reasonable distance from the previous hypotheses.

Current work consists in embedding the different detections of geometric shapes (polygones, ellipses, polynomial curves) into a single process and creating a final relational graph between these shapes.

The final objective of this method is to improve the matching between features from multiple images of a scene for a 3D reconstruction.

4 Conclusion

In this paper we presented a complete scheme for the description of a scene into geometric features. The shape extraction is strongly helped by a level of Perceptual Grouping. We showed that this preliminary stage efficiently extracts a reduced number of groupings, with each one being representative of a salient structure in the image. This reduced number of groupings has the function of *focus of attention* as it restricts the shape extraction process to the salient structures only. The final description of geometric shapes is provided as a graph of relationships between consistent elements in the image.

In order to further reduce the gap between this description and the model of a scene, it is now necessary to take into account the relationships between geometric shapes (adjacency, tangency, parallelism, and symmetry). This

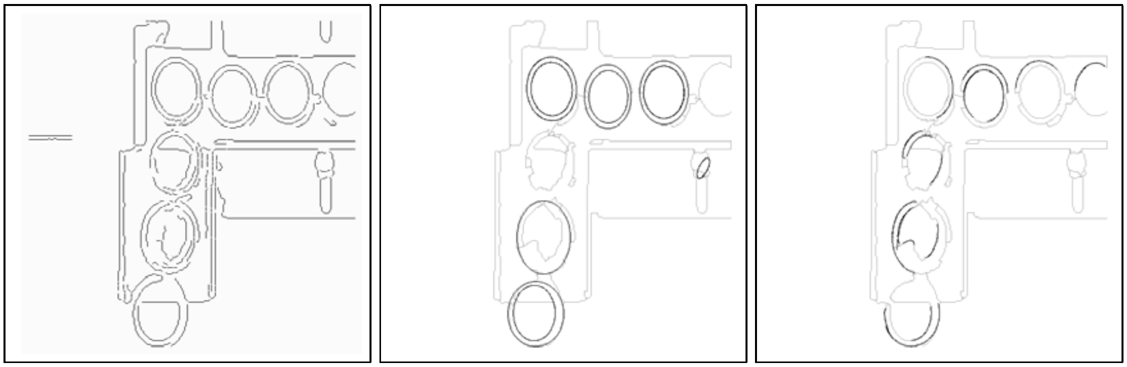

Figure 7: Examples of ellipses and arcs extracted from the scene.

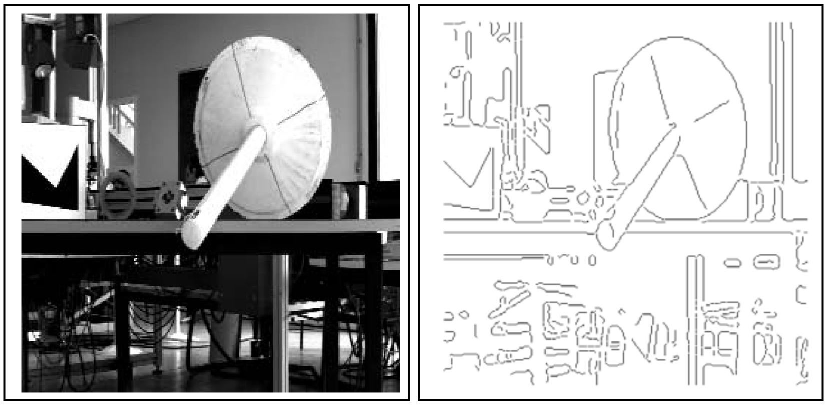

Figure 8: Scene 2 : Original image and edge detection

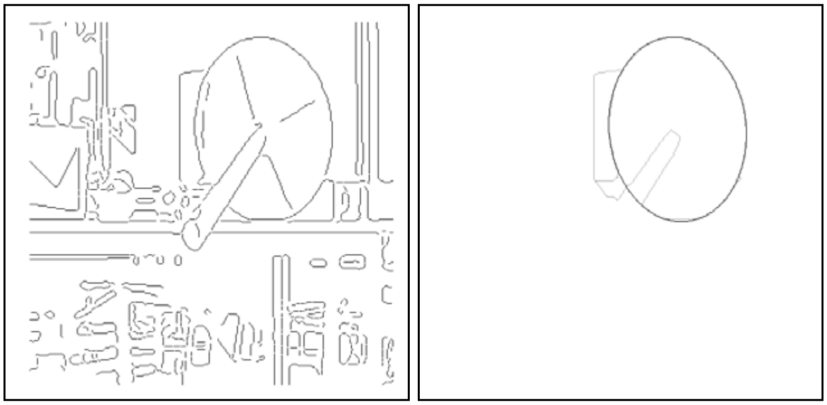

Figure 9: Ellipse detection on a grouping from Scene 2.

scheme of prediction-verification makes it possible to build more and more abstract geometric shapes to describe the scene. Itallows the further addition of contributions from other cues such as other low level processing (a segmentation in regions could be useful to validate some hypotheses) or possibly other views of the same pictures with different scales.

The graph of relationships will be useful for a projective reconstruction of shapes and objects. Ongoing work is investigating the use of this description to help feature matching between multiple images for a 3D reconstruction of the scene.

References

[1] D. H. Ballard and C. M. Brown. Principles of animate vision. In *CVGIP - Image Understanding*, volume 56, pages 3–21, Jul 1992.

[2] D. P. Bertsekas. *Dynamic Programming : Deterministic and Stochastic Models.* Prentice-Hall, INC., Englewood Cliffs, N.J. 07632, 1987.

[3] B. Bhanu. Shape matching and image segmentation using stochastic labeling. Technical report, USCIPI Report 1030, 1981.

[4] G.Roth and M.D.Levine. Extracting geometric primitives. In *CVGIP - Image Understanding*, volume 58, pages 1–22, Jul 1993.

[5] G. Guy and G. Medioni. Perceptual grouping using globally saliency enhancing operators. In *IEEE Transaction on Pattern Matching and Machine Intelligence*, pages 99–103, 1992.

[6] L. Hérault. *Réseaux de neurones récursifs pour l'optimisation combinatoire.* PhD thesis, Institut National Polytechnique de Grenoble, Février 1991.

[7] M. Kass, A. Witkins, and D. Terzopoulos. Snakes: Active contour models. In *Third International Conference on Computer Vision*, pages 259–268, June 1987.

[8] D. G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic publisher, Hingham MA 02043, USA, 1985.

[9] D. C. Marr. *Vision.* Freeman, Oxford, 1982.

[10] P. Montesinos and J.P.Fabre. Groupement perceptuel par optimisation recursive. In *Neuronimes 92*, Nov 1992.

[11] P. Montesinos and L.Alquier. Perceptual organization with active contour functions : application to aerial and medical images. In *ICPR - Pattern Recognition*, volume 2, Aug 1996.

[12] P. Parent and S. W. Zucker. Trace inference, curvature consistency, and curve detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 11, August 1989.

[13] R.Horaud, F.Veillon, and T.Shorda. Finding geometric and relational structures in an image. In *First European Conference on Computer Vision*, pages 23–27, Apr 1990.

[14] R.Horaud and T.Skordas. Stereo correspondance through feature grouping and maximal cliques. In *IEEE Transaction on Pattern Matching and Machine Intelligence*, volume 11, pages 1168–1180, 11 1989.

[15] A. Sha'ashua and S. Ullman. Grouping contours elements using a locally connected network. In *Neural Information Processing Systems*, 1990.

[16] Y. C. Shiu. Experiments with perceptual grouping. In *SPIE, Proc. Intelligent Robots and Computer Vision IX: Algorithms and Techniques,*, volume 1381, Boston, Massachusetts, 5-7 Nov 1990.

[17] S. Urago, M. Berthod, and J. Zerubia. Restauration d'images de contours incomplets par mod lisation de champs de markov. Technical report, INRIA, Rapport de recherche 1688, 1992.

[18] K. Wall and P.-E. Danielson. A fast sequential method for polygonal approximation of digitized curves. In *CVGIP - Image Understanding*, volume 28, pages 220–227, Feb 1984.

[19] M. Wertheimer. Untersuchungen zur lehe von der gestalt ii, translated as: "principles of perceptual organization". In *Readings in Perception, 1958*, pages 115–135, Princeton, N.J, 1923.

[20] S. W. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. In *Neural Computation*, volume 1, pages 68–81, Massachusetts Institute of Technology, 1989.