To appear in the IEEE Computer Society Workshop on Perceptual Organization in Computer Vision, June 26th 1998

Representation of linear structures using perceptual grouping

Laurent ALQUIER, Philippe MONTESINOS
LGI2P - Parc Scientifique G.BESSE
NIMES, F-30000
E-Mail : alquier@eerie.fr
Phone: (33) 4 66 38 70 00
FAX: (33) 4 66 30 70 74

Abstract

The robust detection of curvilinear structures from contours represents an important cue for most computer vision systems. However, the analysis of 2D images of a scene is a difficult task, made of under constrained, ill posed problems. Perceptual Organization is one contribution of psychovisual theories to computer vision. It provides generic approaches for the perception of scenes, necessary to reduce constraints and solve ambiguities.

The purpose of this paper is to propose an efficient method, based on Gestalt rules, to detect salient contours and extract elements of representation from complexe scenes. The proposed grouping strategy is hierarchical and divided in three levels of organization.

The first level is based on the saliency network framework. A quality function based on curvature, co-circularity, continuity and orientation is optimized throughout a network of locally connected elements. The purpose of this level is to reduce the complexity of visual tasks by selecting the most salient linear structures from contours. We propose a generic formalism for the conception of such networks, a more stable family of quality functions, a new algorithm for optimization and criteria for selection of a set of salient groups after optimization. Salient groups of edge elements play a role of focus of attention for the second level of organization. Hypotheses of segments, arcs and points of interest are proposed from each salient group and grouped following rules of parallelism, proximity, continuity to create a reduced set of representative elements of contours. Finaly, we illustrate the final level of organization with an application to junction detection and matching on digital and real scenes.

The main caracteristic of our approach is the separation between a general strategy of organization and grouping modules specialized for a defined task. Salient elements are defined by generic properties. A certain amount of ambiguities and redundancies is necessary to allow the detection of multi-scale structures. This work insists on the manipulation of complexe scenes, on usual systems. It has been applied to various type of scenes, from satellite and medical imaging to indoor and outdoor scenes. The quality of results confirm the robustness of this approach in cluttered environments.

Keywords

Perceptual grouping, scene representation, shape recognition, salient curve detection, reconstruction, feature matching, combinatorial optimization, dynamic programming, hierarchical representation, saliency network.

1 Introduction

One possible definition for Computer Vision is the automatic inference of decisions from pictures with the help of computers. In this context, extracting one or many representations from the image of a scene plays a fundamental role.

From antic philosophies to modern psycho-visual theories, two major conceptions of visual perception have lead to different approaches for Computer Vision. The *Reconstructive* approach [Marr, 1982] [Tarr and Black, 1994] seeks to build an internal representation of the scene and perform reflection from that representation. The *Purposive* approach [Aloimonos *et al.*, 1988] argues that decisions and actions can be performed from direct perceptions and partial representations [Aloimonos, 1994] [Bajcsy, 1988] [Ballard and Brown, 1992] . Recent work about visual perception at biological and psychological levels tend to show that both approaches are part of the process of vision. These lead to the emergence of the *Systemic* approach [Jolion, 1994] [Christensen and Madsen, 1994] taking visual perception as a complex system between the viewer and its environment.

In the context of *Reconstructive* 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 N-P complexity when applied directly to images of low level primitives. Though it is well known and solved in particular cases, such as industrial applications [Batchelor and Whelan, 1997] with well defined objects and environments, it remains an open problem in less well defined systems.

In order to reduce this complexity, it is often necessary to make assumptions about the observed environment, the illumination applied to it, the type of action expected (motion planning ? recognition ?) or even about efficiency constraints (real time ?). Other ways exist, like the study of the human visual system.

1.1 Perceptual Grouping and Computer Vision

Human Vision shows the existence of numerous mechanisms, the purpose of which is to continuously guide perception through the constant flow of visual information. Many of these mechanisms have been applied to Computer Vision. At a biological level, multiple views, focus of attention or artificial retina with polar resolution have proved to be helpful in solving ambiguities of the visual search. At a psychological level, one possible contribution comes from Perceptual Grouping. Taking its roots in Gestalt Psychology [Wertheimer, 1958] , Perceptual Grouping is related to the immediate organization, by human vision, of visual cues into perceptually significant groups.

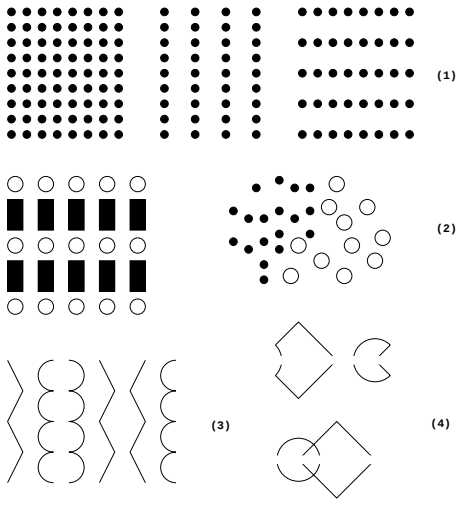

Figure 1. Gestalt rules - (1) Proximity (2) Similarity (3) Symmetry (4) Closure and Continuity

This phenomenon happens immediately, without prior knowledge of the content of a scene. Very simple psycho-visual experiments clearly show the importance of this mechanism, as well as the

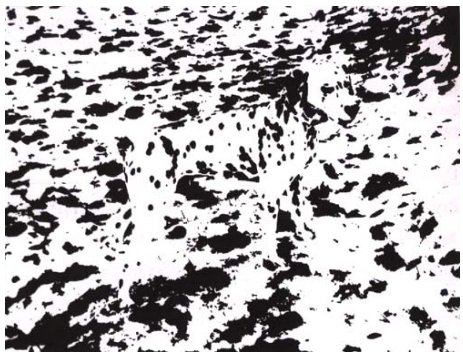

Figure 2. *Gestaltqualität* - **Inner quality of a shape**

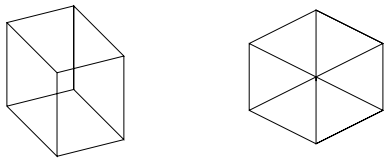

Figure 3. *Pragnanz* - **3D or 2D interpretation according to the visual context**

influence of relationships between objects in a picture. From these relationships were derived a set of collaborative or competing rules of grouping such as proximity, similarity, symmetry, continuity or familiarity - Fig. 1.

Perceptual organization plays an important role in reducing the complexity of visual search by guiding the perception of "structure" before interpreting the content of the observed scenes. This process is related to more general principles of *Gestaltqualität*, quality of "good shape" - Fig. 2 - and *Pragnanz*, a tendency toward simplicity, stability and familiarity according to the visual context - Fig. 3.

The role of Perceptual Grouping in Computer Vision suffered from the lack of formalism and the descriptive nature (as opposed to predictive theories) of Gestalt Psychology. In order to use it in computational terms, many attempts have been made to provide a formalism for perceptual organization, each representing only a fraction of these mechanisms [Witkin and Tenenbaum, 1983] .

Despite these difficulties of representation, the application of perceptual grouping to computer vision has been subject to various approaches [Sarkar and Boyer, 1993b] .

This complex combinatorial problem is usually 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 can model grouping as the construction of graphs based on geometric properties such as suggested by Lowe [Lowe, 1985] [Horaud *et al.*, 1990] . On the other hand, it can be viewed as the definition of measures of energy or probability, and the optimization of this measure over every possible groups [Hérault, 1991] [Guy and Medioni, 1996] [Williams and Thornber, 1997] .

2 Representation of linear structures with perceptual grouping

The work described in this paper uses the perspective of perceptual grouping to structure images into significant visual elements. We propose a complete system to extract visually important features from the contours of an image and give useful elements of representation of the scene for a higher recognition process. The importance of contours as visual cues has been already widely

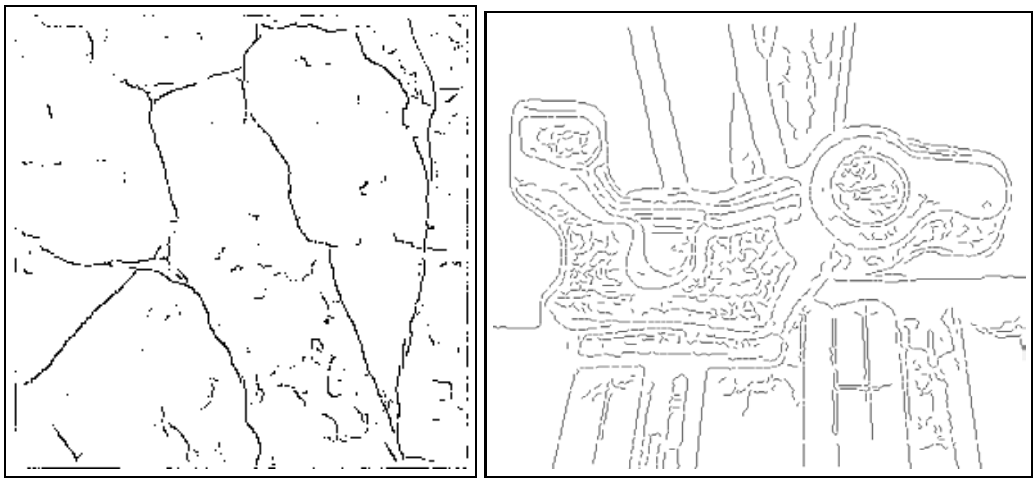

Figure 4. Examples of detection of contours

demonstrated as we are reminded each time we observe a line drawing.

Our method separates three levels of perceptual organization. The *Elementary level* organizes pixels of contours into salient chains. Its purpose is to tackle down the complexity of feature extraction with a robust estimation of structural saliency of contour elements. At the second level, *Intermediate* groups such as Segments, Arcs and Points of Interest are extracted from salient chains. Finally, these elements of representation are organized into *High level* groups with an example of structural matching between two images. As opposed to looking for a single framework to achieve the extraction of structures, one of the features of our method is the use of techniques of different nature for each level of grouping.

3 Elementary Grouping

First we seek to extract the salient curvilinear structures from images after the detection of contour or ridges. The role of this preliminary step is to reduce the combinatorial complexity of the visual search by extracting the main elements of interest from the scene. In order to group pixels of contours into linear structures, this level has to cope with the various perturbations and discontinuities introduced by the detection of contours. Finally, the robustness expected from this level has to be compatible with efficient computation - Fig. 4.

We propose to use optimization techniques for this level, in particular those involving a measure of saliency as they can provide an efficient focus of attention on important structures in the image. More precisely, the method we developed for this level is inspired by a Saliency Network of locally connected elements as defined by Shashua and Ullman in [Shashua and Ullman, 1988] . From this optimization scheme, we define generic principles for the construction of salient networks and the efficient extraction of salient groups after optimization.

3.1 Saliency Networks

A generic definition of a saliency network is a graph of locally connected elements. The nodes of this graph are taken from a set of visual primitives $\mathcal{P}$. The arcs are connecting elements between primitives, defined in a certain neighborhood $\mathcal{V}$ around each primitive. Within this context, a group is defined as a path between a certain number of nodes. A quality function $\mathcal{F}$ is also defined in order to rate the compatibility of a group with a set of structural relationships. The graph is then used to optimize this function and compute a measure of saliency $\mathcal{S}$ for each primitive P, defined as the quality of the best group crossing P.

The optimization scheme, inspired by Dynamic Programming [Montanari, 1971] allows an iterative construction of the best groups crossing each primitive. One of the most important properties of the graph is to perform an evaluation of the saliency of each primitive in a relaxation process involving only local computations. The output of the network is a saliency map rating the likelihood

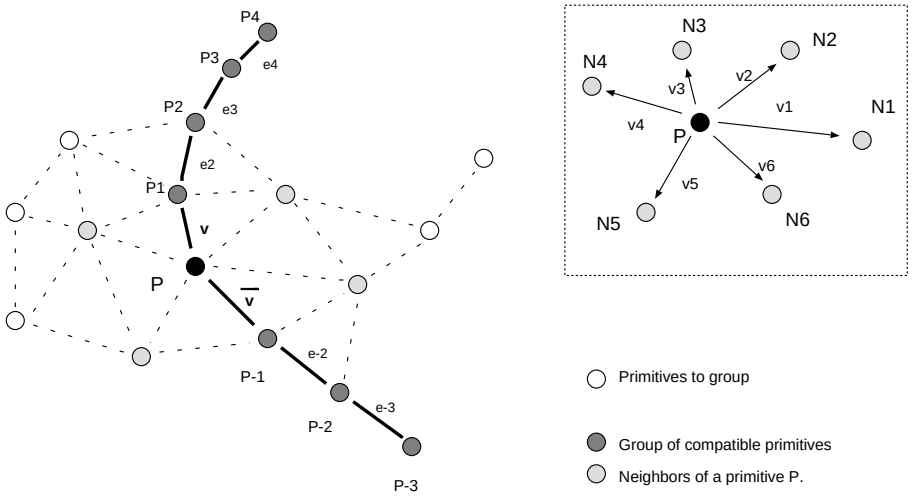

Figure 5. Example of group between locally connected primitive in a Saliency Network. The linear group crosses P through two directions, $v = e_1$ and $\overline{v} = e_{-1}$. The saliency for P is the maximum value for the qualities of possible groups crossing P.

for each primitive to belong to more global structures.

Our contribution to this method is to provide a new formalism for the quality function, a different optimization procedure and finally, a set of heuristics for the extraction of the most salient groups from the network.

3.2 Construction of Saliency Networks

The choice for a generic expression of saliency networks is necessary to make the method applicable to different primitives than pixels of contours. As a result of this conception of such networks, we applied the method to the grouping of pixels and to chains of pixels.

Neighborhood around primitives

In both case, the definition of neighborhood plays an important role in the quality of the results. The connecting elements between a primitive and its neighbors can be considered as a set of possible directions for groups crossing this primitive. A "small" neighborhood can result in giving a crude aspect to the final groups. On the contrary, a neighborhood too large will make the method too demanding in computational resources to be applicable [1].

Structural relationships and quality function

The quality function can be viewed as the measure of the compatibility between primitives and elements of connection for a group. The function defined in the original method was a weighted sum of local contributions for each element of the group. The weights combined constraints involving the global curvature of a group of pixels and the amount of discontinuities along the group. We argue that a better stability in the detection of salient structures can result from the choice of internal and external influences using the same formalism as the energy of 'snakes' [Kass *et al.*, 1987] [Lai and Chin, 1993] .

- External influences are constraints imposed by the image on the groups. They can relate for example to the amount of discontinuities, grey level values along a group, gradient values or common orientation of consecutive elements of contour.
- Internal influences are structural relationships related to the shape of the group. They can involve evaluations of global curvature or co-circularity for example.

[1] If v is the average number of neighbors around each primitive, it is necessary to store v values of quality for each primitive. This can become a serious limitation when using large pictures, such as satellite or medical images.

As the influences are opposed by construction, salient groups will be the best compromise between these influences. The final quality function is a linear combination of each normalized influence. Let K be the number of relationships $R_i(\cdot)$ for a group γ. The quality function is defined by :

$$F(\gamma) \;=\; \sum_{i=1}^{K} \alpha_i R_i(\gamma) \quad with \quad \forall i,\ \alpha_i < 1 \quad and, \quad R_i(\gamma) < 1 \tag{1}$$

The parameters α_i allow a fine tuning of the respective influences of each relationship on the quality function. The difference with the model of active contours is that we seek to maximize the quality function instead of minimizing an energy.

Measure of saliency

From the quality function, we can derive the saliency of a primitive P as the quality of the best group starting from P in the direction of the elements $v_i \in \mathcal{V}(P)$:

$$\mathcal{S}(\,P\,) = \mathop{\mathbf{Max}}_{\gamma \in \delta^n(P)} S^n(\gamma) \tag{2}$$

where $\delta^n(P)$ is the set of every possible group [2] of length n crossing the primitive P. More formally, a group γ crosses P following two directions defined by the elements v and $\overline{v}$. The saliency of the group can be written as a bi-lateral function for each structural relationship :

$$S^n(\,\gamma\,) = \sum_{k=0}^{K} \alpha_i (\, R_k^n(\,v\,) + R_k^n(\,\overline{v}\,) + H_k(\,P, v, \overline{v}\,)\,) \tag{3}$$

Each $R_k^n(\,v\,)$ is the quality of one of the structural relationships defined earlier. The role of the functions $H_k(\,\cdot\,)$ is to correct possible artifacts due to the sum lateral contributions.

A recursive expression for each structural relationship is necessary to optimize it using the saliency network. The key to this method is the choice of a certain type of function, called *extensible function*.

Definition : A function $\psi_N(\cdot)$, defined over N values $e_i, e_{i+1}, \cdots, e_{i+N}$ is said *extensible* if and only if :

$$\mathop{\mathbf{Max}}_{\delta^N(e_i)} \psi_N(e_i, e_{i+1}, \cdots, e_{i+N}) \;=\; \mathop{\mathbf{Max}}_{e_{i+1} \in \delta(e_i)} \psi_1(e_i, \mathop{\mathbf{Max}}_{\delta^{N-1}(e_{i+1})} \psi_{N-1}(e_{i+1}, \cdots, e_{i+N}))$$

This definition, inspired by dynamic programming, makes it possible to reduce to $(N_v - 1) \cdot N$ out of $(N_v - 1)^N$ possibilities [3] the search space for groups of N elements starting from the element e_i. This type of quality function allows also a recursive construction of optimal groups around each primitive.

We assume that the functions $R_k^n(\,\cdot\,)$ are extensible, and that their recursive expression can be written as follows :

$$R_k{}^{(\,n+1\,)}(\,e_i\,) = Q_k(\,e_i\,) + \rho \cdot \mathop{\mathbf{Max}}_{e_j \in \delta(e_i)} \{P_k(\,e_i, e_j\,) \cdot R_k{}^{(\,n\,)}(\,e_j\,)\} \tag{4}$$

where $Q_k(\,e_i\,)$ is the local contribution from e_i to P and ($\rho < 1$) the attenuation of new contributions with distance.

$$\mathop{\mathbf{Max}}_{e_j \in \delta(e_i)} \{P_k(\,e_i, e_j\,) \cdot R_k{}^{(\,n\,)}(\,e_j\,)\}$$

is the best contribution to e_i from its neighbors. The function $P_k(\,e_i, e_j\,)$ is used to allow more or less confidence to this contribution according to the configuration between e_i and e_j. Both functions P_k and Q_k are derived from the recursive expression of the quality function.

[2] In a similar way, $\delta^n(v_i)$ refers to the set of possible groups of length n starting from P in the direction of the element v_i. If we note P_i the primitive connected to P via v_i, then $\delta^1(v_i)$ is written $\delta(v_i)$ and corresponds to the neighborhood of P_i

[3] If N_v is the average number of neighbors in $\mathcal{V}(P)$, $(N_v - 1)$ is the number of possible elements to extend a group arriving in a primitive from the element e_i.

3.3 Recursive optimization

The optimization process performs a relaxation on the recursive expression of each structural relationship $R_s^n(\cdot)$.

Each element e_i around a primitive is associated to a state variable $R_s^n(e_i)$ representing the saliency of the best group of length n starting from P in the direction of the element e_i. The variable is initialized by the local contributions of the elements.

$$R_s^{(0)}(e_i) \;=\; Q_s(e_i)$$

From equation (4), the value of the variable is updated by finding the pair of elements (e_i, e_j), $e_j \in \delta(e_i)$ contributing the most to the state of the element e_i.

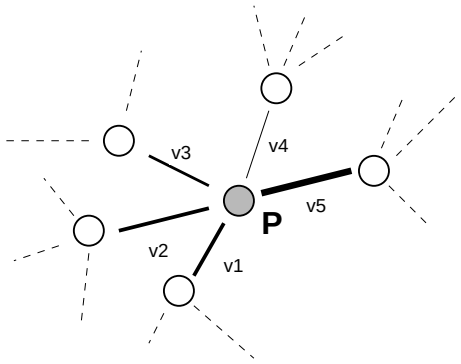

Figure 6. Example of a primitive P and five connecting elements

Pairing of elements

The choice of the best pair of elements at this level can have a strong influence on the result of the optimization. For example, in [Shashua and Ullman, 1991] , Shashua and Ullman used disjoint sets of pairs around each primitive to force the network to converge to a partition of optimized groups. However, this constraint was too strong since it excluded every intersection and junction.

In a similar fashion, our method finds the best pairs of elements before each iteration. This level is performed by defining an application between neighbors of the same primitive.

$$\begin{array}{rccc} \phi \;:\; & \mathcal{V}(P) & \longmapsto & \mathcal{V}(P) \\ & v_i & \longrightarrow & v_j, \quad v_j \neq v_i \end{array}$$

This application allows some pairs to be inhibitory to avoid unwanted configurations between elements. The figure 6 shows an example of a primitive with its connecting elements. In this example, v_5 provides the best contribution to the primitive, v_1, v_2 *and* v_3 provide the same contribution and v_4 provides the lowest contribution. The role of ϕ is to make sure that the pairs remain as reversible as possible while associating always the best elements. In this example, the pair (v_4, v_5) is inhibited as their corresponding primitives are too close, and v_4 should be paired with v_1, v_2 *or* v_3. By comparison, the original method would have paired all the elements with v_5.

Updating the saliency

Contributions from each element are diffused throughout the network using the following procedure.

$$R_k{}^{(n+1)}(e_i) \;=\; Q_k(e_i) + \rho \cdot \mathcal{C}^n(e_i)$$

If a reversible pair has been defined for e_i during the previous step, the corresponding element is used in priority :

$$\mathcal{C}^n(e_i) \;=\; P_k(e_i, \phi(e_i)) \cdot R_k{}^{(n)}(\phi(e_i)) \qquad (5)$$

If there is no existing pair, the neighbor contributing the most to the state of e_i is taken.

$$\mathcal{C}^n(e_i) = \underset{e_j \in \delta(e_i)}{\mathbf{Max}} \{P_k(e_i, e_j) \cdot R_k{}^{(n)}(e_j)\} \tag{6}$$

For each iteration, this global contribution takes into account more distant primitives. Along the optimization, primitives within global structures are enforced by contributions from other primitives of the structure whereas isolated primitives receive little contribution.

The output is a value for the saliency of each primitive in the network. In addition to this *saliency map*, a pair of elements (e_i, e_j) around a primitive P represents the best direction to follow for a group arriving in P from the direction e_i. This information is fundamental for the extraction of the most salient groups.

3.4 Selection of groups

As a result of the optimization, any primitive in the image can be a starting point for an optimized group following the best connections from a pair to another. The optimization reduces the number of possible groups to a single solution for each starting point.

The original method proposed by Shashua and Ullman does rather well for the detection of a single structure in a cluttered image. Unfortunately, it suffers from several drawbacks when applied to natural scenes with multiple salient structures. The following of the best connections is not enough to extract coherent structures from the scene. Experiments on various scenes showed how groups can easily 'jump' between salient structures, especially around junctions and occlusions.

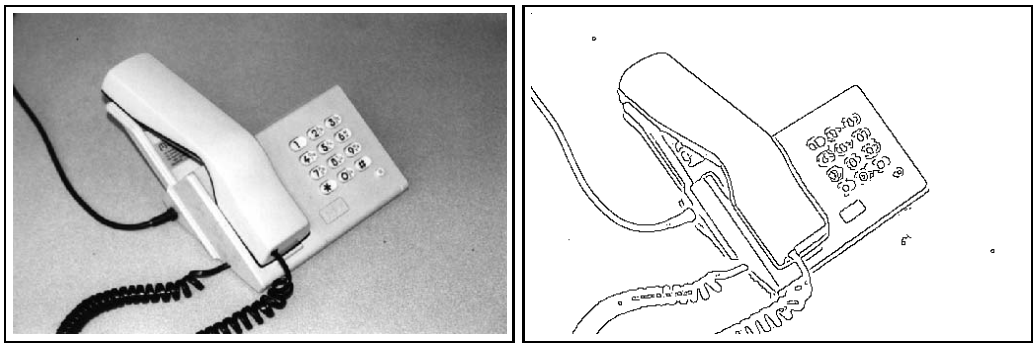

Figure 7. Contours detection from the intensity image of a scene.

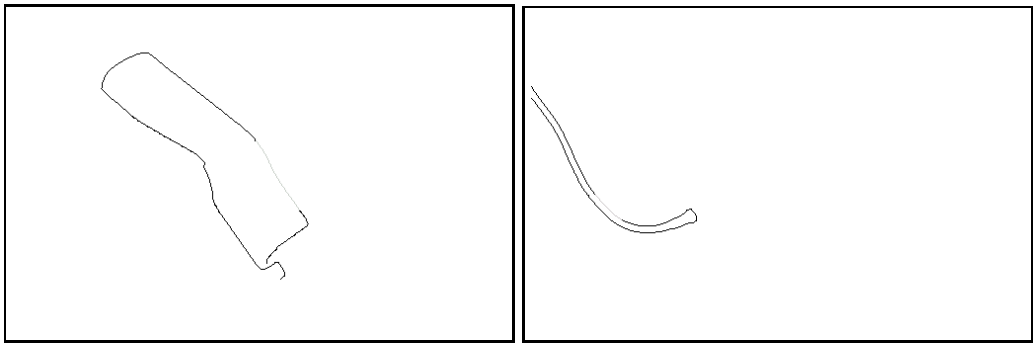

Figure 8. Examples of individual groups after optimization. The chain in grey represents the starting point of the goup in the saliency network.

We propose a different approach to extract multiple salient structures from these networks. As a single group is not enough to reflect a coherent structure in the image, we argue that the sum of multiple groups is enough to cover the salient structures and reject noise and accidental groups. The reduction of the number of optimized groups makes it possible to compute every group from

the primitives available. The final solutions are then selected according to their starting point and their global quality with a simple threshold of the following criteria.

- *Local saliency*

  This value represents the local importance of a primitive (intensity or gradient for a pixel or contour, normalized sum of gradients for a chain of pixels).

- *Accumulation*

  Each group votes for the primitives it aggregates. After following every possible group, the primitives with high votes have a higher probability to belong to important groups.

- *Global saliency*

  The sum of values for the quality function of every pair of elements in a group.

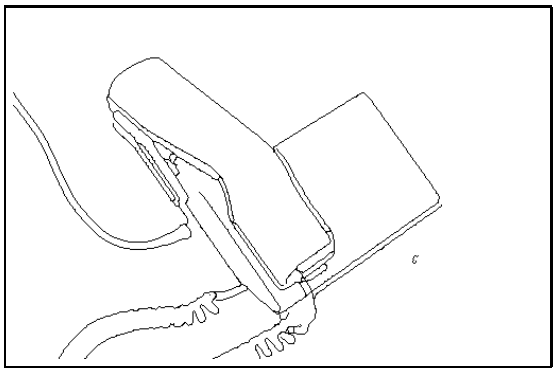

Figure 9. Final selection of 90 groups out of 560 chains in the original image.

Groups corresponding to any of these criteria are selected. Experiments are currently being carried on to define a single criterium from these thresholds.

3.5 Applications and results

The algorithmic complexity of the optimization process is directly related to the number of primitives, as well as the average number of neighboring elements. It is the order of $\mathcal{O}(v^2 \cdot N \cdot p)$, if N is the length of the expected groupings (it is also the number of iterations), p the number of primitives to group together and v the average number of neighbors around each primitive.

During the optimization, 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 the minimum of iterations required to fill gaps along the curves of the image. Usually, the number of iterations is set to the number of elements expected in the final groups.

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 directly related to the number of primitives to group together (on a SPARCstation 5, this time is about 1 second per iteration for 1000 chains to group and an average of 10 neighboring elements - the number of iterations ranges from 20 to 100 depending on the scenes). The only parameters involved into the optimization process are the coefficient of each contribution in the Quality function (α_i) and the number of iterations.

This methodology has been tested on the grouping of pixels, [Montesinos and Alquier, 1996] , and extended to the grouping of chains of pixels [Alquier and Montesinos, 1997] .

The grouping of pixels of contours uses a static neighborhood of 16 elements. The quality function involves terms of global curvature and co-circularity for internal influences and terms of continuity in intensity and orientations for external influences.

By comparison, the grouping of chains of pixels uses a dynamic neighborhood around end-points of chains. The external influences for the quality function involve terms of continuity and respective orientations at end-points. The internal influences use differences of angles and co-circularity.

Results on several synthetic and natural pictures show the relevancy of this approach. It is generic enough to be applied to more complex primitives, such as parallelograms or simple shapes.

- *Experiments on pixels*

  These extreme situations show the robustness of the method to perturbations. The results were selected manually to demonstrate the existance of a circular solution among every possible group.

- *Experiments on chains*

  In these test scenes, groups were selected using the thresholds. This procedure is a semi-automatic selection as opposed to a manual choice of the final groups. The selected groups are represented in black and the rejected primitives in grey. The most important outcome of the grouping of chains is the significant reduction of computational ressources needed for the grouping (to the order of 1 second by iteration as opposed to a minute for pixels on similar scenes).

- *Results on real Images*

  This scene, like the remaining figures of this paper, shows an example of semi-automatic selection of groups of chains. The chains in black represent 398 primitives detected as salient structures out of 1408 chains in the saliency network (in grey). The chains correspond to contours elements.

  The following results demonstrate the adaptability of the method to various scenes. Every class of example have been treated using the same parameters.

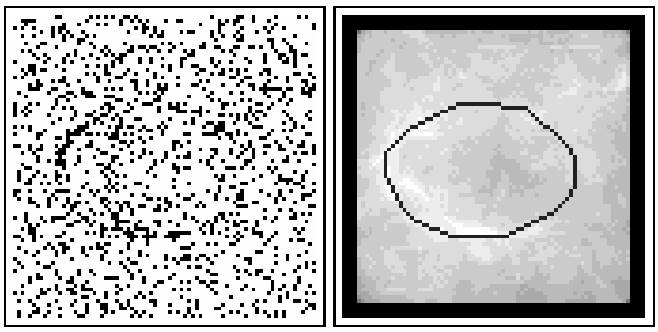

Figure 10. Grouping of pixels - Ellipse with white noise

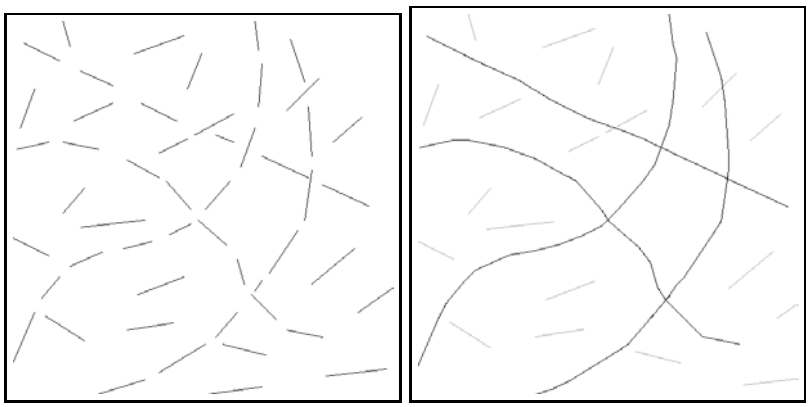

Figure 11. Grouping of salient chains

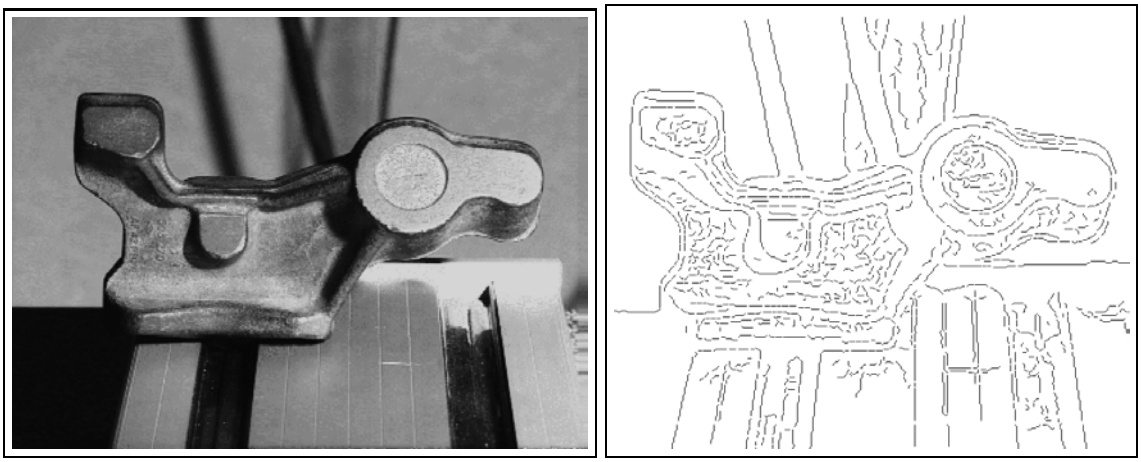

Figure 12. Original image - chains of contours - 1408 elementary chains

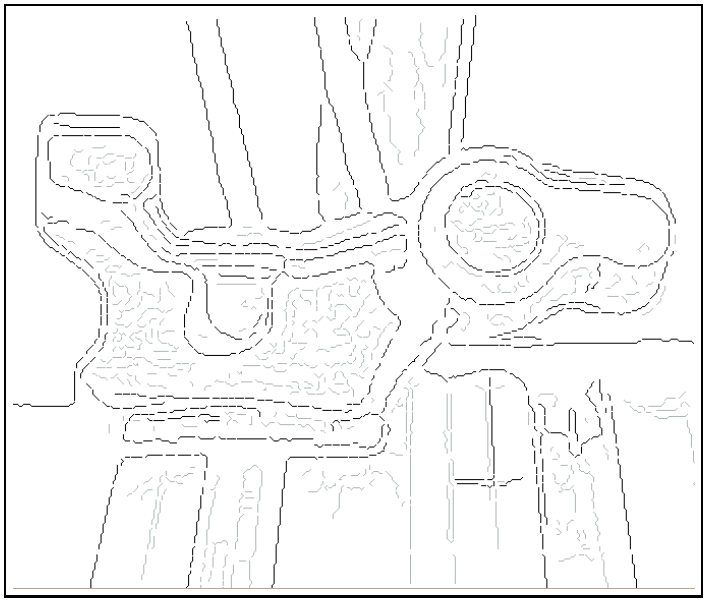

Figure 13. Groups - 398 salient chains

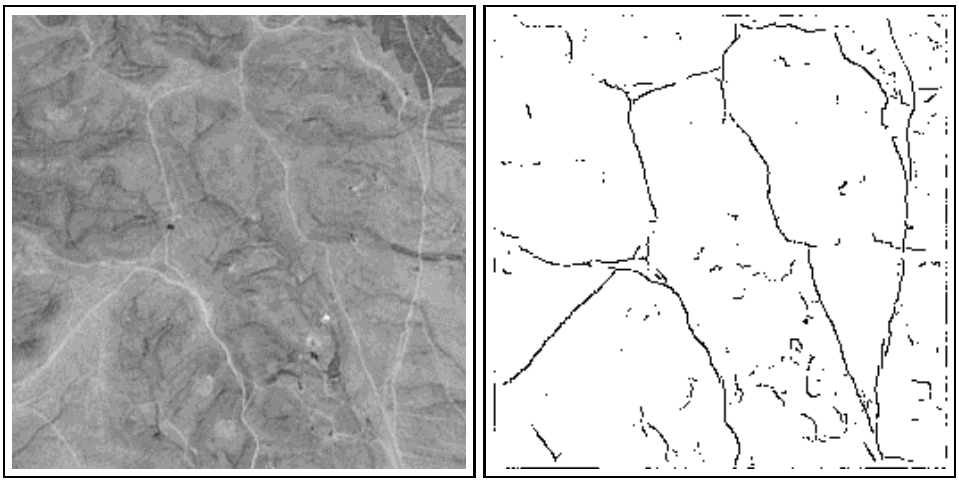

Figure 14. Original image and crest lines detection.

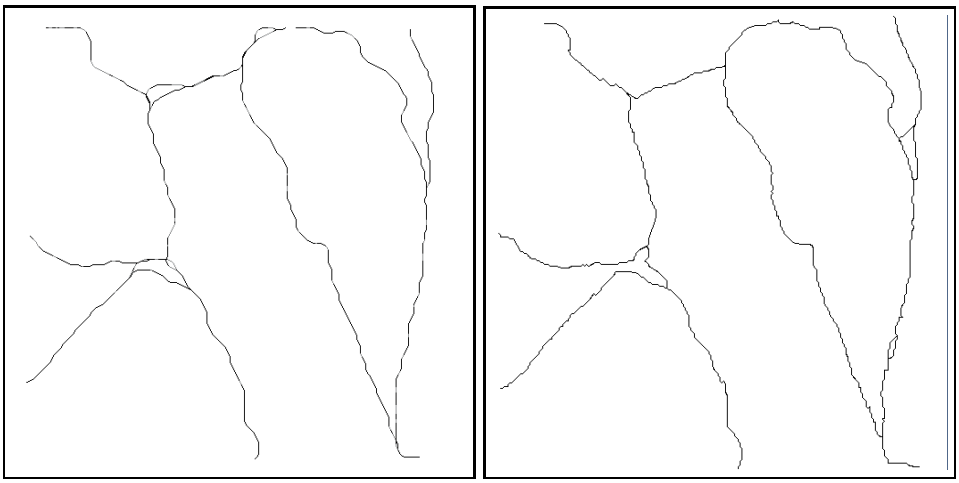

Figure 15. Selection of 16 salient groups of pixels (left) and 9 salient groups of chains (right)

4 Intermediate Grouping

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, groups after optimization and selection are longer, uniform, smooth, and allow curves.

4.1 Structural organization

The purpose of this second step is to build a hierarchy of primitives and relationships between them. Algorithmic methods of perceptual organization are well suited to the representation of groups according to relational graphs. The approach we propose for this level starts with the extraction of hypothesis from each selected group using classical segmentation techniques. It is important to note that the choice of these techniques is independent from our framework of perceptual grouping. This modular approach is necessary to easily adapt our method to a wide range of situations by choosing the most appropriate technique for each situation.

By order of complexity, the first primitives detected on the groups are 'points of interest' : simple junctions (collinearity, tangency, inflexion, corners), multiple junctions (vertices, occlusion). The

other primitives represent the way in which these points are connected to each other : 'segments' (straight lines) and 'arcs' (smooth curvature). We adopted a multi-resolution approach for the detection of 'segments' and 'arcs' in order to reduce ambiguities in the perception of each visual element. A certain degree of redundancies is tolerated to allow multiple interpretations of the same features.

These hypotheses are then organized according to perceptual rules to form graphs of geometric primitives. The groups extracted from the first level provide a focus of attention. Since the presence of multiple linear structures of interest produces a certain amount of overlap between groups, this step is necessary to remove redundant features for the same scale of segmentation. At this level, structural grouping between segments and between curves produces a simple set of the most probable geometrical features for a given scale.

4.2 Segments

The extraction of straight lines from the selected groups corresponds to the classic problem of the Polygonal Approximation of a chain of pixels. Two major approaches can be defined from the extensive bibliography on this problem [Garnesson and Giraudon, 1991] [Rosin, 1997] . The methods of *fusion* consist in aggregating points of the chain into segments minimizing a distance between the chain and a model of segment. The complementary methods concentrate on the localization of the best "cut" points on the chain.

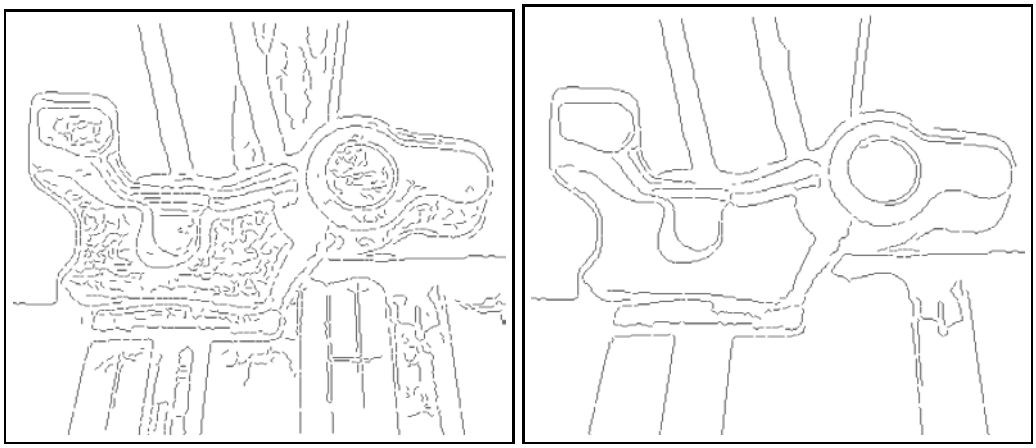

Figure 16. Initial chains - Selection of 34 salient groups from a network of 1408 chains.

Detection of straight features

The segmentation technique we adopted is the application of a fast and stable method by recursive division. A parameter of scale sets the amount of error allowed between portions of the chain and the corresponding detected segments. The final cut points are labeled as 'corners' and added to a list of points of interests.

Perceptual organization of segments

The segments detected from the selected groups are then organized to remove redundant information for a given scale. Overlapping segments are aggregated after evaluation of their respective positions, distances and orientations.

A process of relaxation on the length of remaining segments prepares them for the detection of junctions. For each iteration, we detect the intersections between segments and we derive a function of energy from these intersections. Contributions from segments to this energy are inversely proportional to the distance between the intersection and their end-points. The length of segments intersecting near their end-points is reduced iteratively in order to minimize this energy.

The output of this level is a set of straight lines allowing near intersections around their end-points and complete intersections around their middle-point. Incorrect segments on curved parts

are tolerated at this level. They will be compared to the detection of curved features at a higher level of grouping.

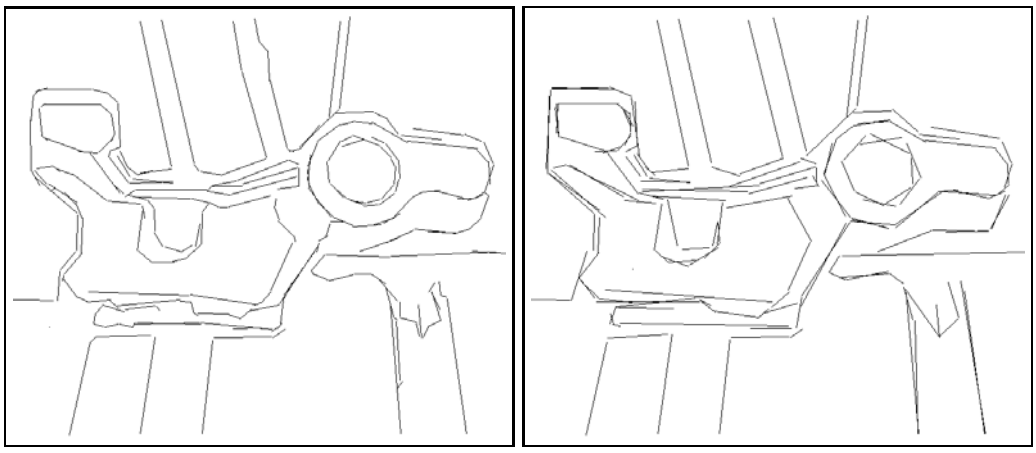

Figure 17. Detection of Segments with different scales. Left : 233 segments, error of 3 pixels. Right : 161 segments, error of 10 pixels.

4.3 Curves

The problem of interpreting salient arcs on a chain of pixels is an extension of the detection of segments since a curve can be always considered as a succession of tangent segments. The main issues of this type of detection have been addressed in [Saund, 1991] . Using multiple scales of detection, it is possible to identify the most salient arcs on a chain by minimizing the error between a model of arc and the chain, the magnitude of discontinuities and by applying geometric criteria to enforce the detection of unique arcs. This method however was limited to the approximation of a chain by circular arcs.

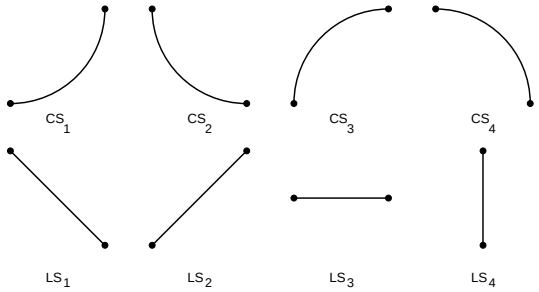

Figure 18. Eight elementary arcs based on the sign and variations of curvature and orientation.

Detection of curved features

Our detection of curved features concentrates on the estimation of curvature along the chain at a certain scale. A Gaussian derivative filter has been applied to the points of chains using a recursive implementation inspired by edge detection techniques [Deriche, 1990] . As a result of this filter, an estimate of orientation and curvature for each point in the chain can be used to segment the chain as a curve.

Points of extreme curvature are added to the list of points of interest as possible corners or inflexion points.

After elimination of parts of the curves with nearly zero of curvature [4] we define "L-arcs" the elementary arcs between two consecutive point of extreme curvature [Gao and Wong, 1993] . These arcs play the same role as the segments detected in the previous step.

Perceptual organization of arcs

The elementary arcs are grouped recursively in a way which is similar to that of the segments. Grouping according to different rules allows us to infer two classes of curves.

- *Arcs rouped by proximity and similarity of curvature.* They correspond to the aggregation of neighboring arcs with curvatures exclusively positive of negative.
- *Arcs grouped by Continuity and Co-curvature.* They correspond to possible circular arcs.

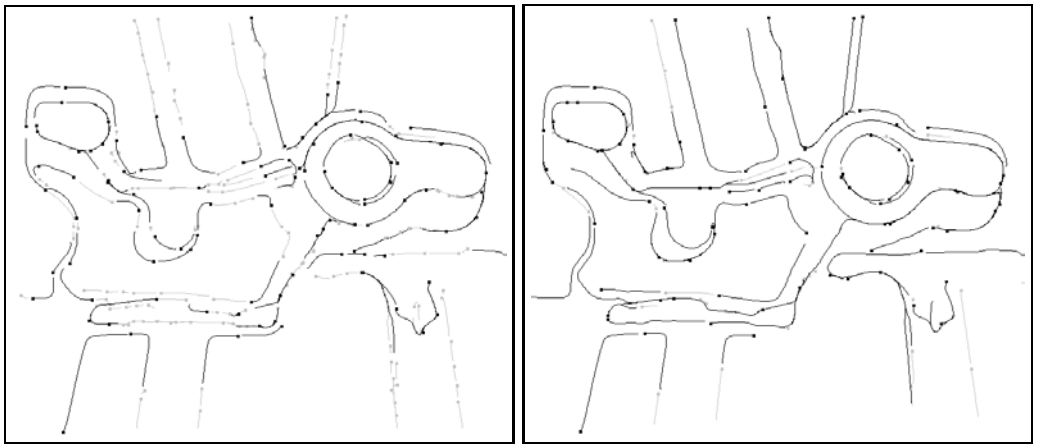

Figure 19. Detection of Arcs with different scales. Left : 281 arcs, scale $alpha = 0.5$. **Right : 175 arcs, scale** $alpha = 0.07$.

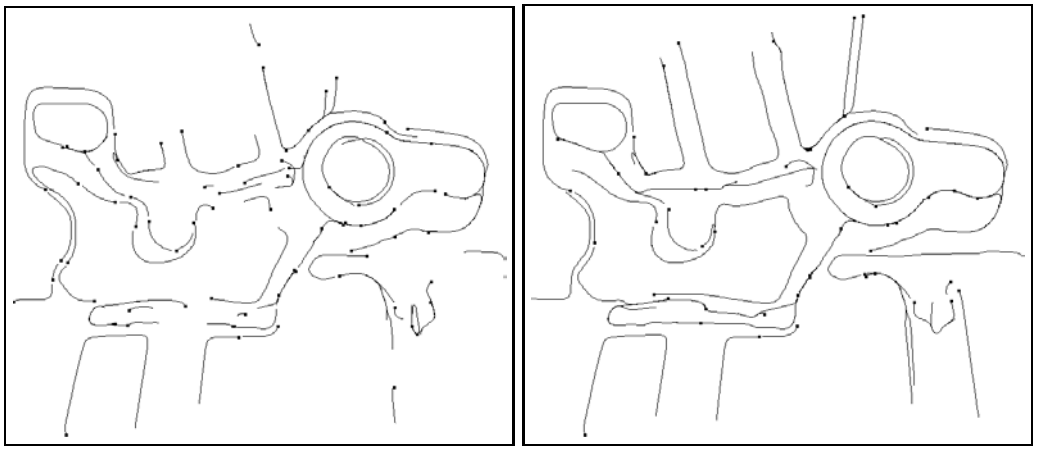

Figure 20. Groups of co-circular arcs with different scales. Left : 80 pairs of arcs, scale $alpha = 0.5$. **Right : 97 pairs of arcs, scale** $alpha = 0.07$.

[4] These parts have been already detected as 'segments'.

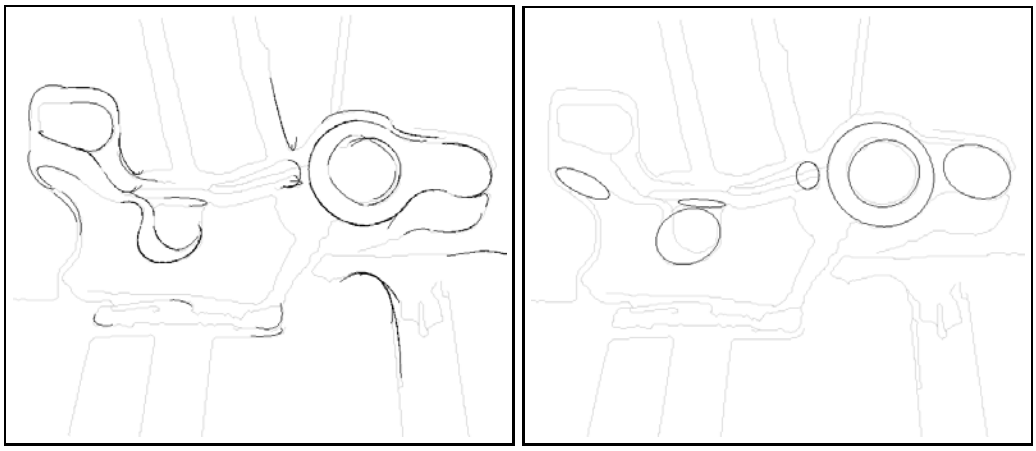

Figure 21. Fitting of 87 elliptic arcs and selection of 7 hypotheses of ellipses

The purpose of the two classes of arcs is to provide a higher level of grouping with features of the same nature, that can be easily compared and manipulated. Again, this level of grouping lacks the global information to allow a discrimination between curves and segments. The scales used for the detection of arcs and segments are unlikely to be related. The function of this level is then to provide the higher level of grouping with enough elements of representation to assist in making a decision.

5 High Level Grouping

The output of the previous level of grouping can be compared to a rough sketch of the geometric structure of the scene. Segments, arcs and junctions are salient elements of representation. The purpose of the high level grouping is to infer more complex elements of representation and to use these elements for visual tasks. Extensions to this level of grouping can also be strongly influenced by the application expected from the analysis of the scene.

- *Discrimination between visual elements.*

  A possible preliminary to many visual tasks is the definition of less ambiguous elements of representation. By comparing the elements detected at multiple scales and keeping the most stable features, it is possible to reduce the ambiguities between salient arcs and tangent segments. The final decision to solve ambiguities can be taken using the Minimum Description Length Approximation such as proposed in [Lindeberg and Li, 1997] .

- *Inference of complex objects*

  The most natural extension of the previous level of grouping is the detection of more complex features such as polygons, ellipses, or ribbons. These features can be directly extracted from the geometric primitives as it is the case for the detection of vanishing points or polygons from segments, or the construction of ellipses from elementary "C-arcs" - Fig. 22.

  They can also be inferred in a more generic way using inference frameworks such as Perceptual Inference Networks as proposed by [Sarkar and Boyer, 1993a] . Such networks represent knowledge about perceptual groups to detect as Bayesian graphs between relationships and use them to detect complex groups from more simple primitives.

- *Relationships between objects*

  This level can also be the starting point for the detection of more complex relationships between visual elements such as symmetry, or continuity and inclusion of regions as suggested by [Mohan and Nevatia, 1992]

5.1 Junction grouping

The points of interest detected during the two previous steps define simple junctions between segments and between curves. Since these features are inferred from single chains, they allow only the detection of simple junctions between consecutive primitives.

Detection of elementary junctions

This step is necessary to extract multiple junctions from the scene. First, double junctions are defined from the intersections of segments. In the same way as segments and arcs, we define a scale of detection related to a distance of the search allowed around endpoints of segments. This parameter allows the detection of "actual junctions", when end-points of segments actually meet, and "virtual junctions", when the intersection occurs outside one of the segments.

Multiple junctions

A simple comparison between segments around these junctions is enough to group them into multiple junctions. Simple junctions sharing the same location and common segments are aggregated into multiple junctions. Once the grouping of junctions has been performed, a measure of confidence is evaluated for each junction according to the presence of a point of interest which lies in a close proximity to their location. This confidence is used to validate the junctions with detected corners and extrema of curvature.

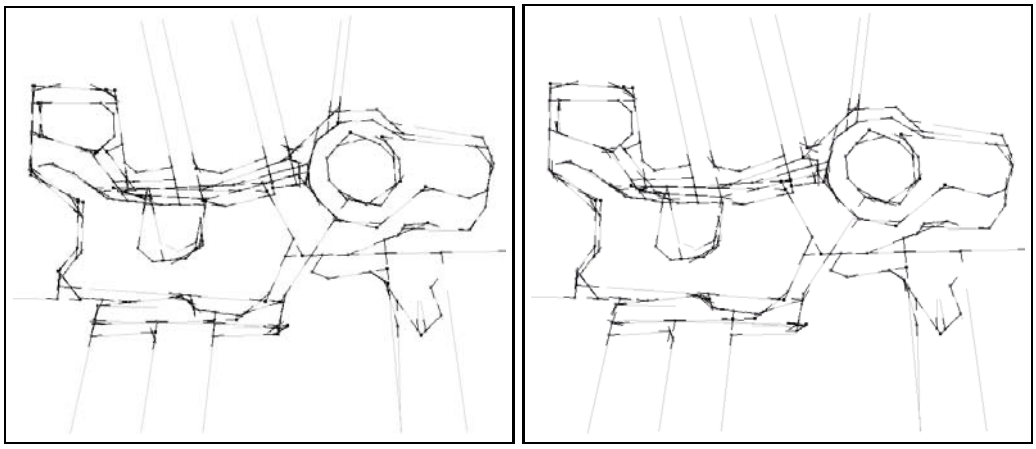

Figure 22. Left : Detection of 744 elementary Junctions from intesections between 169 segments. Right : 224 Junctions left after grouping.

5.2 Junction matching

We are currently extending the high level grouping to structural matching between elements of representation extracted from two scenes or between a scene and a model. The algorithm, inspired by [Chang and Aggarwal, 1997] , considers matching as a special case of temporal grouping. It is generic and robust enough to be adapted to junction matching provided the definition of measures of distances between junctions.

With the help of "perceptual neighborhoods" around features, the procedure iteratively matches junctions between the two scenes (matching inter-pictures) according to perceptual relationships such as intersections, proximity, symmetry, and uses these matching probabilities to activate or inhibit neighbors around each primitive (matching inter-primitive). As an illustration of this algorithm, figures 24 to 28 summarizes the levels of organization of our approach, from contour detection to junction matching.

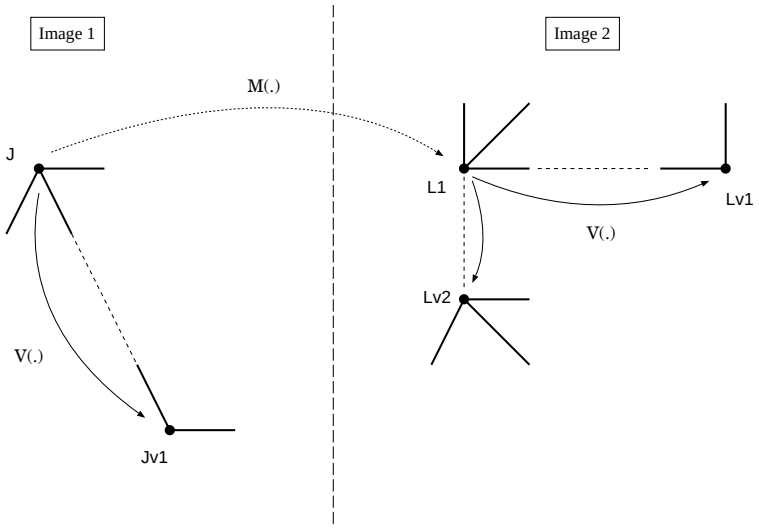

Figure 23. A junction and its Perceptual Match $V()$ and Temporal Match $M()$.

Figure 24. Example of junction matching between two scenes - Images of intensity.

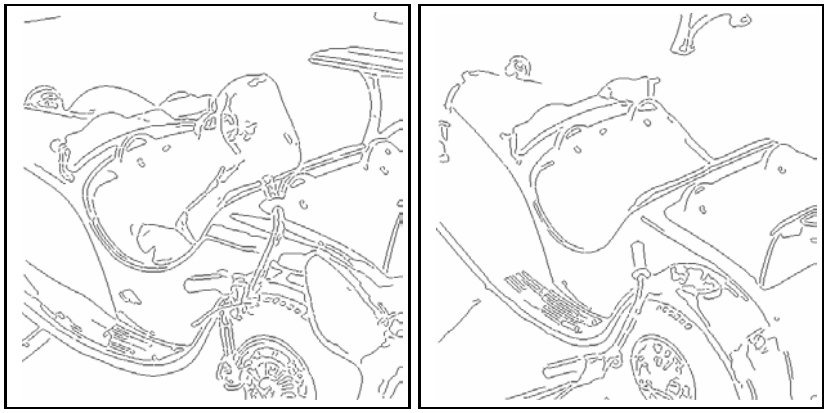

Figure 25. Contour detection.

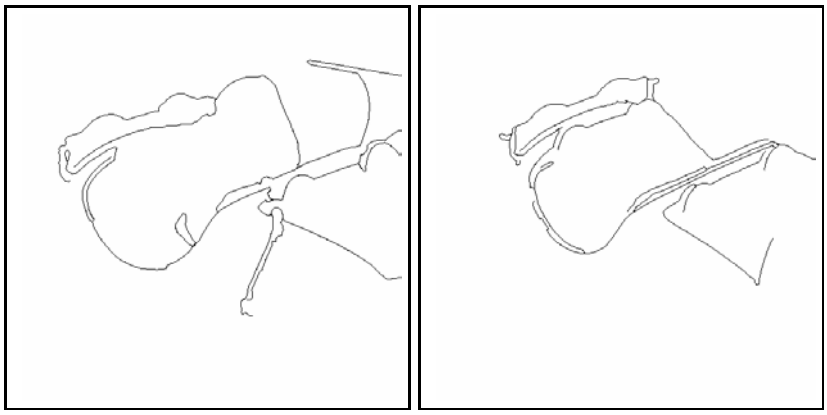

Figure 26. Selection of some salient groups after optimization of the saliency network.

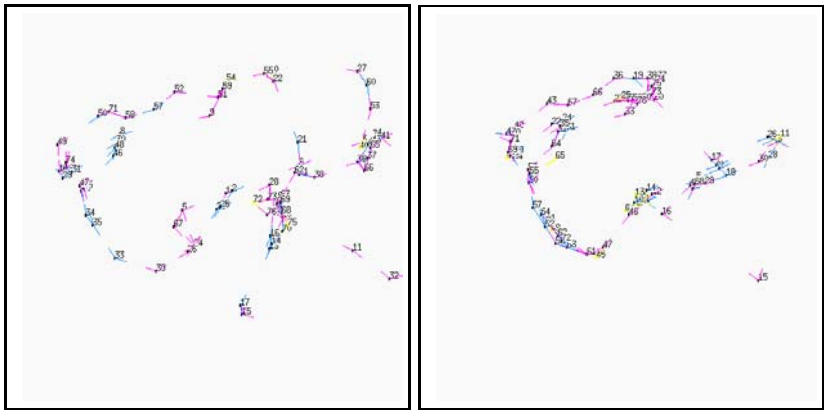

Figure 27. Junctions detected from intersections between salient segments.

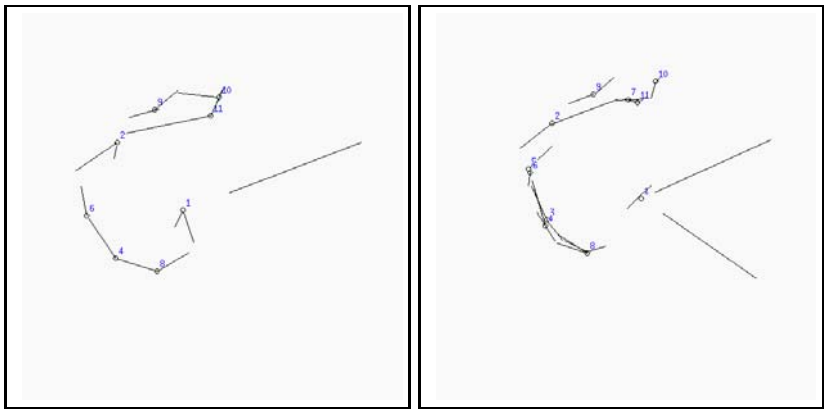

Figure 28. Most probable junction matches - selected from their temporal matching scores.

6 Conclusion

We presented a complete system for progressive perceptual organization of linear structures. As opposed to using a unique framework for the extraction of visual structures, our approach is composed of three levels using representations of perceptual grouping adapted to each type of organization.

The first level detects salient structures using a method derived from Saliency Networks. In addition to providing a generic expression of this type of optimization scheme, our contribution to the method includes a new formalism to evaluate the quality of a group, a different algorithm to enforce the convergence toward stable structures and a set of heuristics for the extraction of salient structures after optimization.

The salient groups detected in the first level provide a focus of attention for the detection of elements of representation. A modular approach to the second level of grouping allows our method to be easily adapted to various situations using more or less specialized modules of segmentation.

The output of this framework are elements of representation organized as graphs of relationships. Results for each level have been evaluated on artificial scenes, as well as indoor and natural images. They confirm the detection of salient structures at various scales of perception. The detection is efficient and stable, with little change of parameters from a scene to another, and robust to perturbations. The total computational time needed, including pre-processing, range between 5 and 15 minutes for scenes up to 900×700 pixels, tested on PCs and workstations [Alquier, 1998] .

Future extensions are concentrated on the detection of more complex features, and the development of multi-scale decisions to keep the most appropriate primitives. The short term applications of our approach are the use of structural matching for 3D reconstruction and automatic indexing of images according to perceptual features.

References

[Aloimonos, 1994] Aloimonos, Y. (1994). What i have learned. *CVGIP : Image Understanding*, 60(1):74–85. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Aloimonos *et al.*, 1988] Aloimonos, Y., Weiss, I., and Bandopadhay, A. (1988). Active vision. *International Journal of Computer Vision*, 2(1):333–356.

[Alquier, 1998] Alquier, L. (1998). *Analyse et représentation de scènes complexes par groupement perceptuel : Application à la perception de structures curvilignes.* PhD thesis, Université Montpellier II, Academie de Montpellier, Sciences et techniques du Languedoc.

[Alquier and Montesinos, 1997] Alquier, L. and Montesinos, P. (1997). Recursive perceptual grouping for 3d object reconstruction from 2d scenes. In *SCIA*. Scandinavian Conference in Image Analysis.

[Bajcsy, 1988] Bajcsy, R. (1988). Active perception. *Proc. IEEE (Special issue on computer visoion)*, 76(8):996–1005.

[Ballard and Brown, 1992] Ballard, D. H. and Brown, C. M. (1992). Principles of animate vision. *CVGIP Image Understanding*, 56(1):3–21.

[Batchelor and Whelan, 1997] Batchelor, B. G. and Whelan, P. F. (1997). *Intelligent Vision Systems for the Industry.* Number ISBN 3-540-19969-1. Springer-Verlag.

[Chang and Aggarwal, 1997] Chang, Y. L. and Aggarwal, J. K. (1997). Line correspondances from cooperating spatial and temporal grouping processes for a sequence of images. *Computer Vision and Image Understanding*, 67(2):186–201.

[Christensen and Madsen, 1994] Christensen, H. I. and Madsen, C. B. (1994). Purposive reconstruction. *CVGIP : Image Understanding*, 60(1):103–108. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Deriche, 1990] Deriche, R. (1990). Techniques d'extraction de contours. Technical report, INRIA Sphia-Antipolis.

[Gao and Wong, 1993] Gao, Q. G. and Wong, A. K. C. (1993). Curve detection based on perceptual organization. *Pattern Recognition*, 26(7):1039–1046.

[Garnesson and Giraudon, 1991] Garnesson, P. and Giraudon, G. (1991). Polygonal approximation : Overview and perspectives. Technical Report 1621, INRIA, Sophia Antipolis.

[Guy and Medioni, 1996] Guy, G. and Medioni, G. (1996). Inferring global perceptual contours from local features. *International Journal of Computer Vision*, 20(1):113–133.

[Hérault, 1991] Hérault, L. (1991). *Réseaux de neurones récursifs pour l'optimisation combinatoire : application à la théorie des graphes et à la vision par ordinateur*. PhD thesis, Institut National Polytechnique de Grenoble.

[Horaud *et al.*, 1990] Horaud, R., Veillon, F., and Skordas, T. (1990). Finding geometric and relational structures in an image. In *ECCV*, volume 90, pages 374–384. European Conference in Computer Vision.

[Jolion, 1994] Jolion, J. M. (1994). Computer vision methodologies. *CVGIP : Image Understanding*, 59(1):53–71.

[Kass *et al.*, 1987] Kass, M., Witkin, A., and Terzopoulos, D. (1987). Snakes: Active contour models. In *Third International Conference on Computer Vision*, pages 259–268.

[Lai and Chin, 1993] Lai, K. F. and Chin, R. T. (1993). On regularization, formulation and initialization of the active contour models. In *First Asian Conference on Computer Vision*, pages 542–545, Osaka.

[Lindeberg and Li, 1997] Lindeberg, T. and Li, M. X. (1997). Segmentation and classification of edges using minimum description length approximation and complementary junction cues. *Computer Vision and Image Understanding*, 67(1):88–98.

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

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

[Mohan and Nevatia, 1992] Mohan, R. and Nevatia, R. (1992). Perceptual organization for scene segmentation and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):616–635.

[Montanari, 1971] Montanari, U. (1971). On the optimal detection of curves in noisy pictures. *Communications of the ACM*, 14(5):335–345.

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

[Rosin, 1997] Rosin, P. L. (1997). Techniques for assessing polygonal approximations of curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):659–666.

[Sarkar and Boyer, 1993a] Sarkar, S. and Boyer, K. L. (1993a). Integration, inference and management of spatial information using bayesian networks : Perceptual organization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):256–274.

[Sarkar and Boyer, 1993b] Sarkar, S. and Boyer, K. L. (1993b). Perceptual organization in computer vision : A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):382–399.

[Saund, 1991] Saund, E. (1991). Identifying salient circular arcs on curves. *CVGIP Image Understanding*, 58(3):327–337.

[Shashua and Ullman, 1988] Shashua, A. and Ullman, S. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. In *IEEE, Second International Conference on Computer Vision*, pages 321–327, Tampa Florida.

[Shashua and Ullman, 1991] Shashua, A. and Ullman, S. (1991). Grouping contours by iterated pairing network. In Lippmann, R., Moody, J., and Touretzky, D. e., editors, *Advances in Neural Information Processing Systems*, volume 3, pages 335–341. Morgan Kaufmann publishers.

[Tarr and Black, 1994] Tarr, M. J. and Black, M. J. (1994). A computational and evolutionary perspective on the role of representation in vision. *CVGIP : Image Understanding*, 60(1):65–73.

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

[Williams and Thornber, 1997] Williams, L. and Thornber, K. (1997). A comparison of measures for detecting natural shapes in cluttered backgrounds. Ft. Lauderdale, FL. Assoc. of Researchers in Vision and Opthalmology (ARVO) Annual Meeting.

[Witkin and Tenenbaum, 1983] Witkin, A. P. and Tenenbaum, J. M. (1983). *Human and Machine Vision*, chapter On the role of structure in vision, pages 481–543. Beck, Hope and Rosenfeld.