A Graph Labelling Approach for Connected Feature Selection

Jocelyn Marchadier[1], Sylvain Michelin[2], and Yves Egels[1]

[1] Institut Géographique National
2-4, Avenue Pasteur F-94165 Saint Mandé Cedex France
[2] Université de Marne-la-Vallée, Equipe Image, Institut Gaspard Monge,
5, Boulevard Descartes, F-77454 Champs sur Marne Cedex

Abstract. Many authors have already proposed linear feature extraction algorithms. In most cases, these algorithms can not guarantee the extraction of adjacency relations between extracted features. Object contours appearing in the analyzed images are often fragmented into nonconnected features. Nevertheless, the use of some topological information enables to reduce substantially the complexity of matching and registration algorithms.

Here, we formulate the problem of linear feature extraction as an optimal labelling problem of a topological map obtained from low level operations. The originality of our approach is the maintaining of this data structure during the extraction process and the formulation of the problem of feature extraction as a global optimization problem.

keywords Contour map, feature extraction, model selection, MDL

1   Introduction

A structural description of an image, relevant for example in an architectural context, is an issue of computer vision. Many algorithms use structural information. Examples of such algorithms are registration algorithms in data bases of images or three dimensional models [4], or calibration algorithms.

Elementary structural elements can be classified into three categories, according to their dimension, i.e. interest points, linear features and regions. More complex features, such as line pencils (vanishing points), collinear segments, etc, may be pertinent in specific applications.

Although the different kind of elementary features are complementary, they are usually extracted independently from each other. For example, some authors proposed techniques of interest point detection, such as corners and multiple junctions [8]. Other authors proposed linear feature extraction techniques from independent point sets [3], chained points [11], or thick point sets [14]. Segmentation algorithms, which partition an image into regions of homogeneous properties, are numerous. The problem of consistency between the different types of features is faced when one tries to organize those features into a global adjacency structure.

The perceptual grouping paradigm is widely used to solve this problem [15]. Techniques using perceptual organization are resistant to occlusions, but are very sensitive to noise and to parameter values. More recently, Fuchs and al. [6] proposed a multi-primitive extraction system where conflicts between features extracted independently are detected, thanks to a structure that stores adjacency relations between the different features.

Other coherent approaches are used in line drawing vectorization [9]. The main interesting point of those approaches consists in using a skeleton of a binary image obtained from the initial bilevel image. This skeleton is viewed as a graph of discrete points, which guarantees the consistency of the adjacency relations between the extracted line segments. Junction points and faces can be naturally deduced from the graph structure. Line segments are extracted with respect to local criteria, but the extension of these methods to other curves seems problematical.

In this paper, we propose a new method for linear feature extraction that enables the retrieval of topological relations between the extracted features. It can be applied to bilevel images as well as on multilevel images. The first step of the extraction process is the construction of a low level topological structure called contour map (section 2). The problem of linear features extraction is then viewed as an optimal labelling problem, which is derived in the second section of this article (section 3), and a simple sub-optimal algorithm is presented. We then propose an example illustrating the usefulness of the proposed contribution (section 4).

2   Contour Map

2.1   Definition

A contour map is a topological map embedded in the discrete plane : its vertices are points of $\mathbb{Z}^2$, its Jordan arcs are 4-connected curves, and its faces are 8-connected sets of points.

A topological map is a data structure representing a cellular decomposition of a plane, that codes the faces and the edges of that decomposition, and that permits to access efficiently to the adjacency relations between the different kind of elements. Such a structure has already been used for the edition of two-dimensional drawings [7], and for representing images with respect to an inter-pixel topology [5,1].

2.2   Construction

A contour map can be obtained from a straightforward algorithm introduced first by M. Pierrot Deseilligny and al. [16]. We suppose that the crests of an image I represents the contours of an original image. An image I is defined as a function with a discrete support (pixels (x, y)), such as all its values are different.

This condition can be fulfilled by defining I as a concatenation of functions. The image of the figure 2.c is obtained using the following tabulated function :

$$I(x, y) = [|Grad(x, y)|, D_-(x, y), x, y] \tag{1}$$

where $|Grad(x, y)|$ is the norm of the gradient of the initial image, and $D_-(x, y)$ is the shortest distance of a point (x, y) to a point (x', y') such as $|Grad(x, y)| > |Grad(x', y')|$.

The contour map is built thanks to a simple algorithm based on a local analysis of the 8-neighborhood of each pixel p of I. Each neighborhood is decomposed into sets of 4-connected components which values are greater than the value of the central pixel. We construct the contour map by adding an edge that connects the central pixel with the highest value pixel of each component.

Pierrot Deseilligny and al. demonstrated that a bijection exists between the faces of the topological map and the local minima of I. A homotopic transformation is then applied, which suppresses the pending darts of the previous combinatorial map until stability. This operation conserves the number of faces and of connected components of the map. The resulting map is called an elementary contour map. The edges of this map bind two 4-connected pixels.

On a real image, we obtain an over-segmentation of the initial image, as shown on the image of the figure 1. This is due to noise, textures, ... The relevant curves are noisy, but well localized. The contour map can be simplified with the use of pre-treatment such as a selective smoothing , or by applying a segmentation algorithm. This kind of operations does not necessarily conserve curves of interest.

3   Features Extraction and Labelling Problem

Feature extraction is viewed here as the segmentation of the contour map into 4-connected discrete curves, according to different models of curves. This guarantees that the connectivity of the map is preserved during the segmentation process and that no adjacency information is lost.

3.1   Model Selection

How can we choose the model which best fits a given sample of measured data among a set of possible models ? To formalize it, let $x = \underline{x} + e$ be a vector of measures, where e is the error vector and $\underline{x}$ is the vector of the exact values of the measures. A statistical model is known when we know the law of e. A model m of a feature (for example, a linear feature model) is a set of relations which can be represented by a function f that verifies $f(\underline{x}, p) = 0$ where p is a vector of parameters describing this feature.

Let $M = \{m_i, i = 1..N\}$ be the set of all possible models and f_i their associated functions. We are then seeking the feature of model $\dot{m}$ of M that is the most adapted to the data sample x. Let $s(x, m_i, p_i)$ be a criterion of model
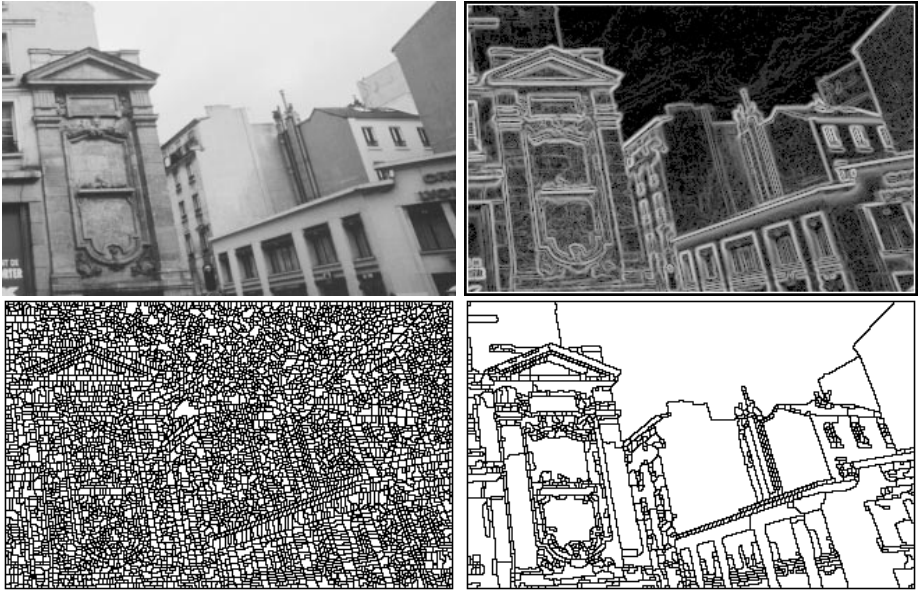
Fig. 1. Construction of a contour map. The upper left image is the initial image. The upper right image represents the modulus of the gradient of the previous one. The bottom left image shows the resulting contour map after the suppression of the pending arcs. Each black point of the image corresponds to an edge. The bottom right image was obtained by thresholding the locally maximal curves to suppress over-segmentation problems

selection, which is a function that measures the appropriateness of the feature of model m_i and of parameter p_i to the data x. For a given model m_i and a given data sample x, the minimization of this function gives an estimate $\widehat{p_i}$ of the parameters of the most appropriate feature of m_i. The best model $\dot{m}$ according to the data x is the model realizing the minimum of s on M.

$$\dot{m} = m_i / s(x, m_i, \widehat{p_i}) = \min_j s(x, m_j, p_j) \tag{2}$$

Note that the estimates $\widehat{p_i}$ of the different features are required to find the model $\dot{m}$. An example of such a criterion is given by Rissanen's Minimum Description Length Principle (MDL) (see [17], for a clear introduction). It has been largely used in the computer vision literature [2,10,13,18]. According to that principle, the feature which is the most adapted to the data is the one that minimize the length of the code of the data which are described with that feature. The model selection criterion is :

$$s(x, m_i, p_i) = -\log_2 P(x/p_i) + sizeof(m_i) \tag{3}$$

where $P(x/p_i)$ is the a posteriori probability of x knowing the parameters p_i of the feature, and $sizeof(m_i)$ is the size of the code of the feature which is equal to the amount of memory needed to stock the vector of parameters.

Note that for a given model, the description length criterion is similar to the maximum likehood criterion.

3.2   Optimal Labelling of a Contour Topological Map

We seek to segment an elementary contour topological map into a set of curves following known models, in order to find a continuous representation which "explains" the discrete map. Those curves can be represented by paths of the map, each path being associated to a label and a model of a given model set. A model selection criterion is used for selecting both the paths and their associated models. It enables to give a label to each edge of the elementary contour map. In order to simplify the problem, we impose that each edge must have only one label.

Let G be an elementary contour topological map, C the set of all possible paths of G, and M the set of all admissible curve models. We associate to each vertex of the elementary contour topological map some measures (coordinates, gradient direction of the initial image ...). We can then construct a vector composed of measures associated to each vertex of the path with a path c_j. We will use the abusive notation $s(c_j, m_i, p_{ij})$ which is the model selection criterion built with measures realized along the path c_j.

The set F of all admissible linear connected features is defined as a subset of $C \times M$ which elements respect the following constraint : the parametric representation of each point with respect to the model and its estimated parameters must conserve the order of the points along the path. This constraint is illustrated on figure 2.
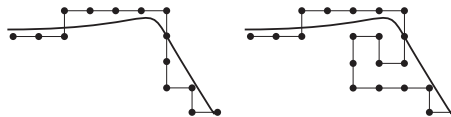

Fig. 2. Curve constraint. The drawings of this figure represent two 4-connected paths and possible continuous analogs. The left drawing is a correct possible curve, while the right drawing does not fulfill the constraint

Definition 1. *A subset E of F is independent if and only if the paths associated to the elements of E do not share an arc.*

Definition 2. *An independent subset E of F is of maximal size if and only if $\forall a \in F, E \cup \{a\}$ is not independent.*

It is straightforward to figure out that the paths included in E contain every edges of G. E is equivalent to a labelling of all the edges of G. Let define the optimality criterion :

Definition 3 (: Optimal labelling). *An optimal labelling of G according to a model set M and a model selection criterion s is the independent subset E of F of maximal size minimizing :*

$$\sum_{(c_j, m_i) \in E} s(c_j, m_i, \widehat{p_{ij}})$$

Note that there is not necessary an unique optimal labelling. Finding an optimal labelling seems to be a NP-complete problem in the general case, if no hypothesis on s, M, and G are made. Classical combinatorial optimization algorithms, such as the A algorithm, or relaxation, can be used to solve this problem, but the implementation of such algorithms is not efficient, partly due to the amount of data required.

3.3   Sub-Optimal Labelling Algorithm

As the labelling problem is complex, we do not seek to obtain an optimal solution. In this sub-section, we present an algorithm that can be used with any model set. This algorithm can be decomposed into three steps.

This first step enables a substantial reduction of the amount of computation required in the following steps. We construct a new topological map from the contour map such that edges are discrete lines, with at most two adjacent faces. Simple classical algorithms [12] can be used.

For each arc of the new map, a local analysis is realized, that enables to find the linear connected feature that minimizes the function

$$s(c_j, m_i, \widehat{p_{ij}}) - s(c_j, m_0, p_{0j}) \tag{4}$$

where c_j is a path that starts by the considered arc, m_i is a model of M, m_0 is the model of independent point, and p_{0j} is composed of the measures associated to each point of c_j. The model m_0 is the noise model, and is implicitly used when no model fits the considered path. The description length $s(c_j, m_0, p_{0j})$ equals the code length of p_{0j}.

The local analysis is driven by a deep first strategy. Each of the "optimal" linear feature is inserted into a priority queue, ordered according to the previous function.

An independent set E of curve of maximal size is then constructed, using the extracted features. While the priority queue is not empty, the first curve c_j of the queue is extracted. If $E \cup c_j$ is independent, then c_j is included in the set E. If it is not, c_j is split into n curves c'_{ij} and n' curves $c''_{i'j}$, such that $E \cup \{c'_{ij}, 1 < i < n\}$ is independent, and $\forall i' | 1 < i' < n', E \cup \{c''_{i'j}\}$ is not. The c'_{ij} are then inserted in the priority queue.

This algorithm is clearly non optimal. The local analysis may result in a curve that is locally, but not globally, optimum according to the model selection criterion. Moreover, the complexity of this algorithm is very high, each admissible curve being analyzed. To reduce the computational cost of this algorithm, the deep first analysis is realized until all the possible estimated models of curves fail a confidence test based on the the maximum residual distance between the points and the estimated curve. The threshold value can be chosen in such a way that it does not change the result of the algorithm.

4   Application

In this section, we give an example of application with a model set composed of two models : a line segment model and a noise model (section 4.1). We then present some results obtained on real images (section 4.2).

4.1   Models

In this sub-section, we derive the two previous models. Let c_j be a path of an elementary contour map. The data vector constructed along the path c_j is composed of the coordinates of each vertex of the path, and of the initial image gradient direction at each of those points. This direction is supposed to be close to the normal vector of an object contour. This measure is used to discriminate the real contours from the spurious ones.

The description length of the noise model is then :

$$s\left(c_j, noise, p_{0j}\right) = 3n_j t_{real} \tag{5}$$

where n_j is the number of points along the path c_j, and t_real is the amount of memory needed to store the chosen representation of a real.

The line segment model is composed of sub-matrix (one sub-matrix per point P) of the following form :

$$f(x_P, y_P, \theta_P; \theta, d) = \begin{pmatrix} x_P \cos\theta + y_P \sin\theta - d \\ \theta - \theta_P \end{pmatrix} = \begin{pmatrix} -r_{\widehat{d},P} \\ -r_{\widehat{\theta},P} \end{pmatrix} = 0 \tag{6}$$

where x_P and y_P are the coordinates of P, θ the angle between the normal vector of the line and the x axis, d the distance of the origin to the line, and $r_{\widehat{d},P}$ and $r_{\widehat{\theta},P}$ are the residuals of f.

The estimation of the parameters of the line can be calculated by non-linear constrained optimization [3]. In order to reduce the computation cost, we can use a linear least square estimator, considering the set of equations $ax + by = -1$, where a and b are the new parameters of the line, or an estimate obtained by sampling the data points, using techniques such that the least median of squares or RANSAC.

Once the optimal parameters has been obtained, the description length of the path by the line has to be computed. The residuals of the observations (x_P, y_P, θ_P) must then be calculated, which is costly. We will use the residuals of f, $r_{\widehat{d},P}$ and $r_{\widehat{\theta},P}$. We suppose that $r_{\widehat{d},P}$ follows a normal law of standard deviation σ_d, and that $r_{\widehat{\theta},P}$ follows a normal law of standard deviation σ_θ. Moreover, we suppose that residuals are independent. The description length of a curve c_j according to a line segment model is given, after a straightforward development, by :

$$
s\left(c_j, line, \begin{pmatrix} \widehat{\theta} & \widehat{d} \end{pmatrix}\right) \approx n_j \log_2\left(2\pi \frac{\sigma_d \sigma_\theta}{\epsilon_d \epsilon_\theta}\right) + \\
\frac{1}{2\ln 2} \sum_{P \in c_j} \left(\left(\frac{r_{\widehat{d},P}}{\sigma_d}\right)^2 + \left(\frac{r_{\widehat{\theta},P}}{\sigma_\theta}\right)^2\right) + (2 + n_j)t_{real} \tag{7}
$$

where ϵ_d and ϵ_θ are the resolution of the digitalization of θ and d, t_{real} is the code size in bits of a real, and n_j is the number of points in the curve c_j. The coding length of the model is $(2 + n_j)t_{real}$ because 2 reals are necessary to code the parameters θ and d, and a real is needed to code the parametric coordinate of each point along the segment.

4.2   Results

In this section, we discuss the results of the algorithm of sub-section 3.3 on the image of the figure 1 (figure 3) .

We first estimate the standard deviations σ_d and σ_θ. To do so, test lines can be manually selected, and their estimates can be computed. Those estimates can be used for an image sequence. In our example, 8 lines were used. σ_d was lying between 0.36 and 0.28 pixel, with a mean of 0.34 pixel and σ_θ between 0.05 and 0.386 radian with a mean of 0.14 radian.

In our application, the linear estimator is more robust than the least median of squares estimator. The used MDL criterion is strong enough to drop lines with outliers in the linear least squares case, Whereas a large number of random samples does not guarantee the repeatability of the extraction of local consistent labellings. It is well known [18] that the MDL criterion tends to accept outliers. Although not especially addressed here, this problem is very sensitive in our algorithm. The early acceptance of a feature with outliers (in this case, points belonging to another feature) can give a global labelling result that is not satisfactory. This problem is solved here by underestimating the standard deviations. Experiments show that a low σ_θ gives a stronger filtering criterion, while a low σ_d gives lines with well localized end points. The figure 3 demonstrates the resulting labels with a low σ_θ. The lower σ_θ, the most powerful the filtering effect is. The last figure gives an example of correct labelling.

This method is well adapted to images without contrasted random textures. It can be applied on other types of images, and have been tested succesfully on aerial images of urban environment. On any images, complex shapes of contours are split in simple linear models from the initial model set. When the model
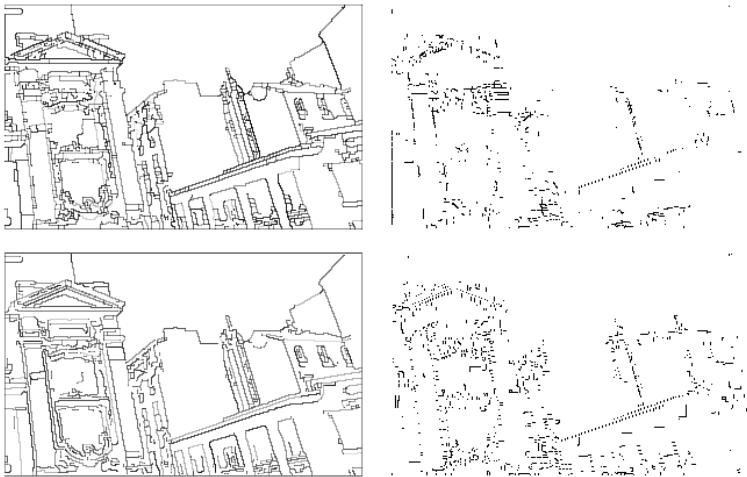
Fig. 3. labelled lines (right column) and noise (left column). Upper images : σ_d=0.34 pixel and σ_θ 0.14 radians, Lower images : σ_d=0.12 pixel and σ_θ=0.04 radians

set contains only the line segment model, the result is a polygonization of the contours. Although the method presented here is less performant in term of time than other methods dedicated to line detection, it can easily be extended to other types of linear features. Moreover heuristics can be designed for each model of the model set in order to reduce the amount of computation time.

5   Conclusion

In this contribution, we have presented the problem of linear feature extraction as a combinatorial optimization problem. We have also proposed a sub-optimal algorithm that has been applied to the extraction of lines segments in outdoor photographs. Although it does not perform perfectly, it demonstrates the usefulness of our approach. Moreover, only two parameters, which values can be estimated, are needed. Actual and future works concern the extension of this approach to face features.

References

1. Braquelaire, J.-P., Brun, L. Image Segmentation with Topological Maps and Interpixel Representation, Journal of Visual Communication and Image Representation, **9(1)** (1998) 62–79
2. Brunn, A., Weidner, U., Forstner, W., Model-based 2D Shape Recovery, in: Mustererkennung 1995, G.Sagerer et al. (eds.), Springer-Verlag, 260–268
3. Danuser, G., Stricker, M., Parametric Model Fitting : From Inlier Characterization to Outlier Detection, IEEE Trans. on PAMI, **20(2)** (1998) 263–280

4. Dickinson, S.J., Pentland, A. P., Rosenfeld, A., 3-D Shape Recovery Using Aspect Matching, IEEE Trans. on PAMI, **14(2)** (1992) 174–198
5. Fiorio, C., A topologically Consistent Representation for Image Analysis : the Frontiers Topological Graph, DGCI'96, Lectures Notes in Computer Sciences, no. 1176, (1996) 151–162
6. Fuchs, C., Frostner, W., Polymorphic Grouping for Image Segmentation, Proc. 5th ICCV, (1995) 175–182
7. Gangnet, M., Hervé, J.-C., Pudet, T., Van Tong, J.-M., Incremental Computation of Planar Maps, SIGGRAPH Proc., Computer Graphics, **23(3)** (1989) 345–354
8. Harris, C., Stephens, M., Combined Corner and Edge Detector, Alvey88, (1998) 147–152
9. Janssen, R. D. T., Vossepoel, A. M., Adaptative Vectorization of Line Drawing Images, Computer Vision and Image Understanding, **65(1)** (1997) 38–56
10. Leclerc, Y. G., Constructing Simple Stable Descriptions for Image Partitioning, International Journal of Computer Vision, **3** (1989) 73–102
11. Lee, J.-W., Kweon, I.S., Extraction of Line Features in a Noisy Image, Pattern Recognition, **30(10)** (1997) 1651–1660
12. Lindenbaum, M., Bruckstein, A., On recursive, O(N) Partitioning of Digitized Curve into Digital Straight Segments, IEEE Trans. on PAMI, **15(9)** (1993) 949–953
13. Lindeberg, T., Li, M.-X., Segmentation and Classification of Edges Using Minimum Description Length Approximation and Complementary Junction Cues, Tech. Rep. ISRN KTH/NA/P–96/01–SE, Jan. 1996.
14. McLean, G.F., Kotturi, D., Vanishing Point Detection by Line Clustering, IEEE Trans. on PAMI, **17(11)** (1995) 1090–1095
15. Mohan, R., Nevatia, R., Using Perceptual Organization to Extract 3-D Structures, IEEE Trans. on PAMI **11(11)** (1989) 1121–1139
16. Pierrot Deseilligny, M., Stamon, G., Suen, C., Veinerization : A New Shape Description for Flexible Skeletonization, IEEE Trans. on PAMI, **20(5)** (1998) 505–521
17. Rissanen, J., A Universal Prior For Integers and Estimation by Minimum Description Length, The Annals of Statistics, **11(2)** (1983) 416–431
18. Torr, P. H. S., Geometric Motion Segmentation and Model Selection, Philosophical Transactions of the Royal Society A, J. Lasenby. and A. Zisserman. and R. Cipolla. and H. Longuet-Higgins editors, pp. 1321