Unsupervised Statistical Method for Edgel Clustering with Applications to Shape Localization

F. Destrempes
DIRO, Département d'Informatique
et de Recherche Opérationnelle,
Université de Montréal
destremp@iro.umontreal.ca

M. Mignotte
DIRO, Département d'Informatique
et de Recherche Opérationnelle,
Université de Montréal
mignotte@iro.umontreal.ca

Abstract

In this paper, we present a statistical model for linking edgels, which yields two models for edgel clustering. The first model favors large clusters, whereas the second one favors small sub-clusters. We show how the first model can significantly help localizing a shape in an image. The second model might be useful for other applications, such as clustering edgels into objects.

1. Introduction

Partitioning edgels into clusters that represent objects in an image is a hard problem which is still open as of now.

Following [7], we observe that grouping edgels into clusters is equivalent to segmenting the set of pairs of edgels into two classes: "off" and "on". Indeed, this segmentation defines uniquely an adjacency matrix for a graph having the set of edgels as set of nodes. The clusters are then defined as the connected components of that graph.

In this paper, we present a statistical model for a feature closely related to the one presented in [7]. This feature indicates whether two edgels should be linked or not. We present two models for the marginal distribution of this feature on the classes "off" and "on". The first model favors large clusters, whereas the second one favors small subclusters. The actual segmentation is based on a threshold which is optimal in the sense of the Maximum Likelihood (ML) for the two marginal distributions. We can then use the centroids of the main clusters to initialize the procedure for localizing shapes presented in [4]. This strategy seems to improve significantly that procedure.

In Section 2, we present a method for pre-segmenting the contours of an image into edgels. In Section 3, we explain how to remove the long background edgels from the set of edgels. Section 4 is devoted to the statistical model for the linking feature. In Section 5, we describe in details the SEM algorithm [2] for the estimation of the statistical parameters. The segmentation method is presented in Section 6. The application of the clustering method to localization of shapes is explained in Section 7. Finally, we present experimental results in Section 8, and a brief conclusion in Section 9.

2. Pre-segmentation of the contours into edgels

Given an image, we use the statistical method described in [3] to detect the contours. Next, we remove the isolated points of the contours and decompose the resulting set into a family $\mathcal{P} = \{p_\alpha\}$ of disjoint non-empty paths with the property that each path is maximal (no two distinct paths have adjacent end points). Such, a decomposition is not unique, but we fix one. We then compute a polygonal approximation of each path p, from which we obtain a set of disjoint edgels $\mathcal{L}$. The length of an edgel l is denoted by $\lambda(l)$, and the average length of all edgels is denoted by λ.

A segmentation of the set T of pairs of edgels into the classes "off" and "on" defines uniquely an adjacency matrix for a graph Λ having the set of edgels $\mathcal{L}$ as set of nodes. The clusters are then defined as the connected components of that graph.

3. Statistical model for removing long edgels

We have observed that the random variable $\tilde{Z}$ given by $\lambda(l)$ follows an exponential law. See Figure 1 and Table 1 for an example. We consider the set $\mathcal{L}_+$ of relatively long edgels l defined by $\lambda(l) > 5\lambda$. Note that the probability that an edgel belongs to $\mathcal{L}_+$ is equal to e^{-5} according to the exponential law. Our hypothesis is that the edgels in $\mathcal{L}_+$ belong to the background rather than to objects. We thus substract $\mathcal{L}_+$ from the set of edgels $\mathcal{L}$, and re-estimate the average length λ.

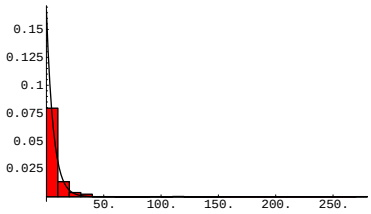

Figure 1: *Empirical density function for the length of edgels and the estimated distribution for Image 3 (see Figure 6).*

4. Statistical model for linking edgels

We adopt a modified version of the feature considered in [7]. Given a pair of edgels (l, l'), we define $\rho_{l,l'}$ to be the distance between the two nearest end-points of l and l', and $\hat{\rho}_{l,l'}$ to be the minimum of $\lambda(l)$, $\lambda(l')$, and λ. We also consider the angle $\theta_{l,l'}$ between the two edgels (the angle is normalized between $-\pi/2$ and $\pi/2$). We then define the normalized feature

$$y_{l,l'} \quad = \quad \exp(-\frac{\rho_{l,l'}}{\hat{\rho}_{l,l'}\cos^2(\theta_{l,l'})}).$$

The feature $y_{l,l'}$ is an indicator for linking the edgels l and l' in the graph Λ (to be constructed), a higher value being in favor of linking the two edgels. This feature is invariant under rotations, translations, and scaling of the set of edgels, and is symmetric in l and l'. Note that in our setting, $y_{l,l'}$ is not a probability, but rather a normalization between 0 and 1.

We have observed a concentration of the feature values around 0. In fact, in our tests, more than 98% of the pairs of edgels are less than twice the average value of the feature (which is around 0.0025). This empirical delta Dirac prevents from estimating the remaining relevant data. Thus, we remove the subset T_0 of T formed by the pairs of edgels (l, l') for which $y_{l,l'}$ is less than twice the average feature. We then consider only pairs of edgels in the set $T_1 = T \setminus T_0$. If $(l, l') \in T_0$, the edgels l and l' are not linked in the graph Λ.

We consider three classes of pairs of edgels in T_1: the class e_1 of pairs of edgels that should definitely not be linked in the graph Λ, the class e_3 of pairs of edgels that should definitely be linked in the graph Λ, and an intermediate class e_2.

Based on empirical results, we model the distribution of the random variable $Y_{l,l'}$ conditional to setting the linking label to e_1 by an exponential distribution

$$\mathcal{E}(y_{l,l'}; \tilde{\mu}_1).$$

On the other hand, we model the distribution of the random variable $Y_{l,l'}$ conditional to setting the linking label to each class e_j, with $j = 2, 3$, by a Gaussian distribution

$$\mathcal{N}(y_{l,l'}; \tilde{\mu}_j, \tilde{\sigma}_j^2)$$

where we expect $\tilde{\mu}_1 < \tilde{\mu}_2 < \tilde{\mu}_3$.

Figure 2 shows an example of the empirical distributions and the estimated distributions. The corresponding estimated parameters are presented in Table 2.

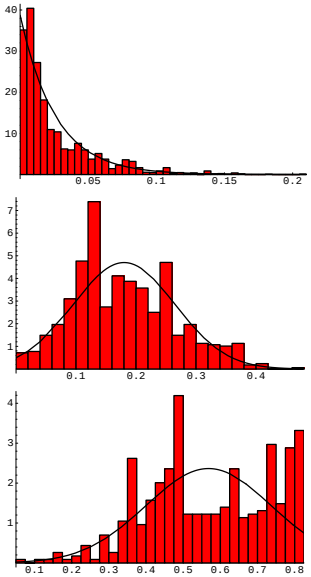

Figure 2: *Empirical density functions of the feature for linking edgels and the distributions estimated by the SEM procedure for Image 3 (see Figure 6). From Top to Bottom: pairs of edgels in class e_1, in class e_2, and in class e_3.*

number of edgels	average length of edgels
864	5.98102
number of short edgels	average length of short edgels
847	4.92304

Table 1: *Removal of long edgels for Image 3 (see Figure 6).*

class	π_i	$\tilde{\mu}_i$	$\tilde{\sigma}_i$
e_1	0.658774	0.0258291	
e_2	0.220731	0.18005	0.084858294
e_3	0.120495	0.568251	0.168721961

Table 2: *Values of the parameters estimated by the SEM algorithm for Image 3 (see Figure 6).*

5. Estimation of parameters

In order to estimate the vector of parameters $\Phi = (\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}_2, \tilde{\mu}_3, \tilde{\sigma}_3)$, we resort to the SEM algorithm [2]. We assume that the random variables $Y_{l,l'}$ are i.i.d. and follow a same law $\tilde{Y}$. We consider a hidden random variable $\tilde{X}$ taking its values in the set $\{e_1, e_2, e_3\}$. The aim of the algorithm is to obtain a decomposition of the form

$$P_{\tilde{Y}}(\tilde{y}) \quad = \quad \sum_{i=1}^{3} \pi_i P_{\tilde{Y}/\tilde{X}}(\tilde{y}/e_i)$$

with $P_{\tilde{Y}/\tilde{X}}(\tilde{y}/e_i)$ as in Section 4, which is optimal in the sense of the least square. The procedure can be summarized as follows in our context:

• **Initial segmentation**: We consider only the pairs of edgels (l, l') in T_1. We use the K-means algorithm described in [1], with $y_{l,l'}$ as only attribute. The class with smallest mean is labeled e_1, the class with greatest mean is labeled e_3, and the remaining one e_2.

Repeat until the current estimation of the parameter vector is close to the previous one:

• **Estimation step**: We use the ML estimator for an exponential distribution on the class e_1; namely, μ_1 is given by the mean of the feature. For the Gaussian distributions, we use the usual ML estimators on each class. Finally, we estimate π_i by the proportion of pairs of edgels having label e_i. This gives the current estimation of the parameter vector Φ.

• **Stochastic step**: For each pair of edgels $(l, l') \in T_1$, one realization of $\tilde{X}$ is simulated according to the probabilities $p'_i = \pi_i P_{\tilde{Y}/\tilde{X}}(y_{l,l'}/e_i)$, with $i = 1, 2, 3$, using the current estimation of the parameter vector Φ. We place $y_{l,l'}$ in class e_i with probability $p_i = p'_i/(p'_1 + p'_2 + p'_3)$.

6. Segmentation of edgels into clusters

We now consider the disconnected graph with set of nodes T_1. We define the observed random field $Y = \{Y_{l,l'}, (l, l') \in T_1\}$, with each random variable $Y_{l,l'}$ as in Section 4, and the label random field $X = \{X_{l,l'}, (l, l') \in T_1\}$, with each hidden random variable $X_{l,l'}$ taking its values in the set $\{c_1 =$ "off", $c_2 =$ "on"$\}$.

We consider two models. In the first model, the class "off" is formed of e_1, and the class "on" is formed of e_2 and e_3. This model favors large clusters. We then obtain the distributions

$$\begin{aligned} P_{Y_{l,l'}/X_{l,l'}}(y_{l,l'}/c_1) &= P_{\tilde{Y}/\tilde{X}}(y_{l,l'}/e_1) \\ P_{Y_{l,l'}/X_{l,l'}}(y_{l,l'}/c_2) &\propto \sum_{j=2}^{3} \pi_j P_{\tilde{Y}/\tilde{X}}(y_{l,l'}/e_j). \end{aligned}$$

In the second model, the class e_2 belongs to the class "off", thus favoring small sub-clusters. This model yields the distributions

$$\begin{aligned} P_{Y_{l,l'}/X_{l,l'}}(y_{l,l'}/c_1) &\propto \sum_{j=1}^{2} \pi_j P_{\tilde{Y}/\tilde{X}}(y_{l,l'}/e_j) \\ P_{Y_{l,l'}/X_{l,l'}}(y_{l,l'}/c_2) &= P_{\tilde{Y}/\tilde{X}}(y_{l,l'}/e_3). \end{aligned}$$

See Figure 3 for a comparison between the two models.

We then model the likelihood by the distribution

$$P_{Y/X}(y/x) \quad = \quad \prod_{(l,l') \in T_1} P_{Y_{l,l'}/X_{l,l'}}(y_{l,l'}/x_{l,l'}).$$

In either case, we view the clustering problem as finding the optimal segmentation x in the sense of the ML; i.e., which maximizes the likelihood $P_{Y/X}(y/x)$. The optimal solution is obtained by setting $x_{l,l'}$ to the class c_i that maximizes $P_{Y_{l,l'}/X_{l,l'}}(y_{l,l'}/c_i)$.

For the localization of shapes, we have adopted the first model in order to obtain a few large clusters. The second model could be useful for other applications, in which one would group together sub-clusters into objects. See Figures 5, 6, and 7 for examples of clustering using the first model, and Figure 8 for two examples of clustering according to the second model.

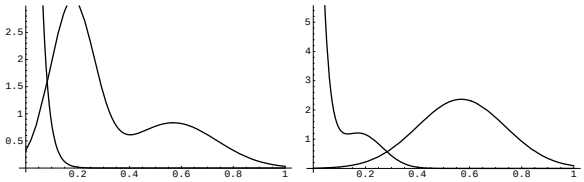

Figure 3: *Left: estimated distributions with the class "on" composed of e_2 and e_3. Right: estimated distributions with the class "off" composed of e_1 and e_2.*

7. Localization of shapes

A curve is represented as a sequence of points

$$\gamma \quad = \quad (x_1, y_1, ..., x_d, y_d)$$

where the d points $(x_1, y_1), ..., (x_d, y_d)$ are equally spaced on the curve.

We consider rigid deformations given by translation (τ_x, τ_y), scaling s, and rotation ψ, applied point-wise to a mean curve γ_0, as well as a vector of non-linear deformations ξ (see [4]) which is estimated thanks to the Probabilistic Principal Component Analysis [9]. This yields a vector of deformation

$$\theta \quad = \quad (\tau_x, \tau_y, s, \psi, \xi).$$

The resulting curve is denoted γ_θ.

We view the localization of a shape γ_0 in an image as finding its deformation θ that minimizes the statistical Gibbs field U presented in [4] on the range

$$\begin{aligned} 0 \leq \quad & \tau_x & \leq M, \\ 0 \leq \quad & \tau_y & \leq N, \\ 0.2\,d \leq \quad & s & \leq 0.6\,d, \\ 0 \leq \quad & \psi & \leq 2\pi \\ -3 \leq \quad & \xi_i & \leq 3 \end{aligned}$$

where $M \times N$ is the size of the image. Our assumption (and this seems to be the case) is that an optimal solution for the function U is the desired deformation of the shape. In order to solve that optimization problem, we resort to the stochastic algorithm [5].

However, due to its complexity, the function U is rather hard to optimize. Yet, it is likely that the object to be localized in the image is among the main clusters (though perhaps partially). Thus, in a first phase, we consider restricted domains given by the clustering method. Namely, for each of the main clusters, we take the restricted range

$$\begin{aligned} c_x - M/6 \leq \quad & \tau_x & \leq c_x + M/6, \\ c_y - N/6 \leq \quad & \tau_y & \leq c_y + N/6, \\ 0.2\,d \leq \quad & s & \leq 0.6\,d, \\ 0 \leq \quad & \psi & \leq 2\pi \end{aligned}$$

where d is the diameter of the image, and (c_x, c_y) is the centroid of the given cluster. Also, for that first phase, we optimize the simpler function U' (similar to [8]) defined by

$$U'(\tau_x, \tau_y, s, \psi, y) \quad = \quad \sum_{i=0}^{m} - \max\Big(\frac{\exp(-\sigma d(s_i))}{d(s_i)}, 1\Big)$$

where $d(s)$ is the distance from the point s to the set of contour points C. One could also use the heuristic function defined in [6].

Then, in a second phase, each of the n solutions found previously (with $n = 3$ in our tests), is used to initialize the same stochastic optimization algorithm, but applied to the function U on the entire domain. If ever the object were not among the main clusters, the second phase will allow the localization of the shape (though at the expense of more computations).

In [4], we would first optimize U' on the entire domain, and then optimize U on a restricted domain given by the solution found in the first phase. Note that an optimal solution to U' might not be the desired deformation of the shape. So, this strategy is not optimal, but this was the best we could do without the clustering method. See Table 3 for a comparison between the two strategies.

8. Experimental Results

The results obtained with a simple example are presented in this paper. The shape is a 20th century classical guitar. The data base consists of 10 pictures. Each curve is represented by a template of 70 equally spaced points. The training phase yields a reduced dimension of 5 for the non-linear deformations. We have tested the localization procedure on five images [4] with 40 different seeds for the stochastic optimization algorithm on each image. The percentage of seeds yielding to a good result within a given number of iterations for the first step and a fixed number of 500 iterations for the second step is indicated in Table 3. One can compare with the results obtained in [4]. We present two examples of the localization method in Figure 9.

Our strategy, exploiting the knowledge of the centroid of each object in the scene, allows to reduce significantly the computational cost of the stochastic search, and/or to increase the good localization rate. Moreover, the procedure for clustering edgels takes only a few seconds (from 3 to 7 seconds on a PC workstation 400MHz for the images shown in this paper).

9. Conclusion

In this paper, we have presented a statistical model for linking edgels and two models for edgel clustering. The first model produces large clusters, whereas the second one favors small sub-clusters. We have shown how the first model can significantly help localizing a shape in an image. The second model might be useful for other applications, such as clustering edgels into objects.

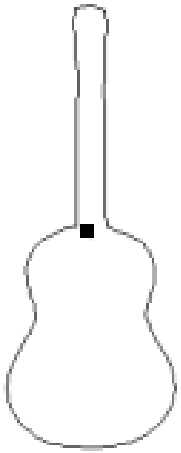

Figure 4: *Mean shape and its centroid.*

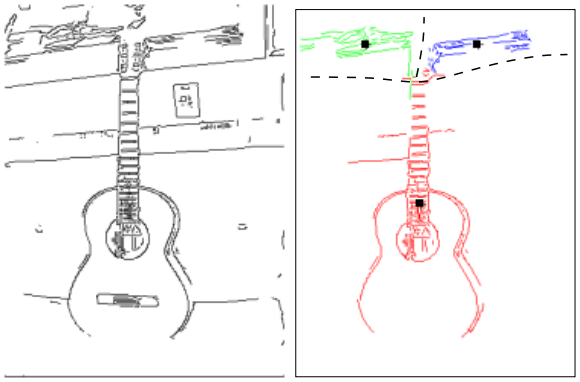

Figure 5: *Example of a segmentation into clusters according to the first model: the three main clusters and their centroids for Image 1.*

Image	2500 iterations	5000 iterations	10000 iterations	20000 iterations
1	97.5%	100%	100%	100%
2	47.5%	62.5%	82.5%	95%
3	55%	70%	80%	85%
4	90%	97.5%	100%	100%
5	57.5%	60%	65%	67.5%

Image	2500 iterations	5000 iterations	10000 iterations	20000 iterations
1	90%	97.5%	100%	100%
2	45%	52.5%	82.5%	85%
3	7.5%	7.5%	7.5%	7.5%
4	90%	100%	100%	100%
5	22.5%	27.5%	32.5%	37.5%

Table 3: *Proportion of the seeds for which convergence of the optimization algorithm occurred within the indicated number of iterations for each of the tested images. Top: using the strategy presented in this paper. Bottom: using the strategy presented in [4].*

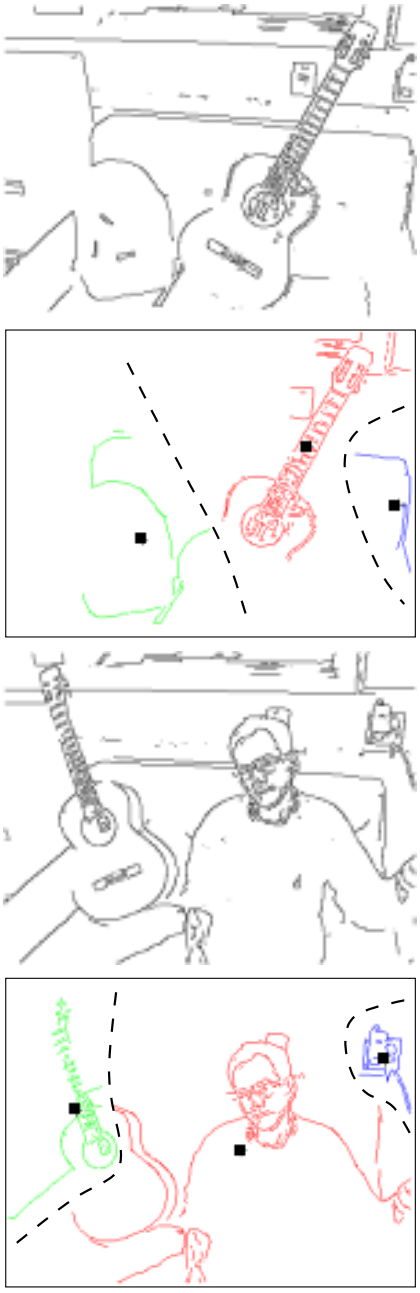

Figure 6: *Example of a segmentation into clusters according to the first model: the three main clusters and their centroids for Images 2 and 3.*

References

[1] S. Banks. *Signal processing, image processing and pattern recognition.* Prentice Hall, 1990.

[2] G. Celeux and J. Diebolt. L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. *Revue de statistiques appliquées*, 34(2):35 – 52, 1986.

[3] F. Destrempes and M. Mignotte. Unsupervised detection and semi-automatic extraction of contours using a statistical model and dynamic programming. In *ICSIP' 2002*, accepted.

[4] F. Destrempes and M. Mignotte. Unsupervised localization of shapes using statistical models. In *ICSIP' 2002*, accepted.

[5] O. Francois. An evolutionary strategy for global minimzation and its markov chain analysis. *IEEE Trans. on Evolutionary Computation*, 2:77 – 90, 1998.

[6] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 267 – 278, March 1996.

Figure 7: *Example of a segmentation into clusters according to the first model: the three main clusters and their centroids for Images 4 and 5.*

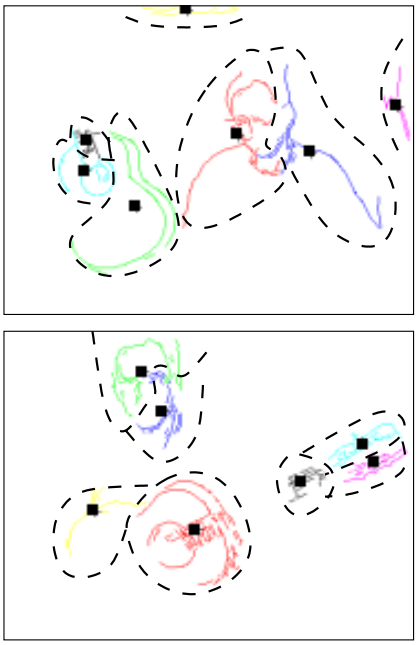

Figure 8: *Example of a segmentation into sub-clusters according to the second model: the seven main sub-clusters and their centroids for Images 3 and 4.*

Figure 9: *Example of localization of a shape: deformation of the mean shape for Images 3 and 5.*

[7] A. R. Kelly and E. R. Hancock. Grouping line-segments using eigenclustering. In *BMVC' 2000*.

[8] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(2):129 – 141, February 2000.

[9] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Technical report NCRG/97/003*, Juillet 1998.