MULTISEGMENT DETECTION

Rafael Grompone von Gioi, Jérémie Jakubowicz

CMLA, ENS Cachan, France
{grompone, jakubowi}@cmla.ens-cachan.fr

Gregory Randall

IIE, Universidad de la República, Uruguay
randall@fing.edu.uy

ABSTRACT

In this paper we propose a new method for detecting straight line segments in digital images. It improves upon existing methods by giving precise results while controlling the number of false detections and can be applied to any digital image without parameter setting. The method is a nontrivial extension of the approach presented by Desolneux *et al.* in [1]. At the core of the method is an algorithm to cut a binary sequences into what we call a *multisegment*: a set of collinear and disjoint segments. We shall define a functional that measures the so called meaningfulness of a multisegment. This functional allows us to validate detections against an a contrario background model and to select the best ones. The result is a global interpretation, line by line, of the image in terms of straight segments which gives back accurately its geometry. Comparisons with state of the art methods will be performed (more examples are available on line).

Index Terms— Straight line segment detection, Number of False Alarms (NFA), Computational Gestalt

Fig. 1. Up-Left: A digital image. Up-Right: All DMM 1-meaningful segments. Clearly there are many more printed segments than we would be able to perceive, but all of them we do perceive are among these detected segments. Bottom-Left: DMM 1-meaningful segments after exclusion principle. Bottom-Right: Multisegment detections

1. INTRODUCTION

Straight segments give important information about the geometric content of images. These segments can be used as low-level features to extract information from images or can serve as a basic tool to analyze and detect more elaborate shapes. As features, they can help in several problems, as stereo analysis [2], crack detection in materials [3], and image compression [4].

Straight segment detection is an old and recurrent problem in computer vision. Faugeras and his collaborators investigated digital segments thoroughly and derived interesting applications [5]. To detect segments, they split edge maps into chains and do polygonal approximations. The most standard segment detection method uses the Hough Transform [6] to extract lines and then cuts them off into segments using gap and length thresholds. In this paper we will call it HTM.

HTM have serious drawbacks. Various thresholds must be set. When correctly set, they can lead to good results, but using fixed thresholds can lead to a significant number of false positives or false negatives. Moreover, the step that leads

from chain codes or lines to segments also involves some parameters. Using fixed values usually leads to misinterpretations. See Figs. 5.

In [1] Desolneux, Moisan and Morel proposed a segment detection method (DMM) based on controlling the number of false positives. The main idea is to count the number of aligned (gradient direction) points and find the segments as outliers in a non-structured background model. This method is based on a general principle of perception, according to which an observed geometric structure becomes perceptually meaningful when the expectation of its number of occurrences is very small in the absence of causal relations [7]. DMM gives a fairly good segment explanation of the scene. It has demonstrated to give neither false positive nor false negative. However, we shall see that it very often missed the right interpretation when aligned segments are present (see Fig. 1 bottom-left.)

All segment detection methods have an implicit 1D binary sequence segmentation step. It turns out that many of their drawbacks come from this segmentation step. HTM uses

fixed thresholds for gap and length. DMM gives the best explanation in terms of *one* segment. When collinear segments are present, this is not necessarily the perceptually best interpretation. We propose to keep the DMM methodology but to search for a more structured event, the *multisegment*, that is to say a set of collinear and disjoint segments. As in the DMM approach, the best multisegment will be the least expected in the *a contrario* model. We shall see that this more sophisticated event results in a better interpretation of the image in terms of straight segments, see Fig. 1 bottom-right.

The dependence of the results of the proposed method on its parameters is very weak. In practice we fixed them once for the hundreds of images we tested.

The method we propose here can also be seen as a unified variational formulation: For each straight line and each potential multisegment interpretation, a functional gives a measure of meaningfulness. This measure allows to validate or not a detection and, at the same time, to rank them and select the best one.

The paper is organized as follows: Section 2 presents the DMM segment detector. Our algorithm is presented in Section 3. Finally, section 4 shows some results.

2. MEANINGFUL SEGMENTS

In [1], Desolneux *et al.* presented an algorithm for straight segment detection. The two key points of their approach are the use of gradient orientation and a new framework to deal with parameter setting.

The first step of DMM is illustrated in Fig. 2. The gradient of the input image is computed and only its orientation is kept. In Fig. 2 this information is codified by dash angles. Given a segment, one counts the number of *aligned points*, *i.e.*, points having the gradient orthogonal to the segment up to a certain precision θ. All potential straight segments on the image must be tested; those that satisfy a threshold criterion based on their length l and their number of aligned points k, are kept as detections.
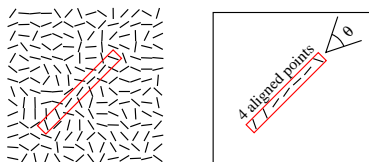

Fig. 2. Left: One segment shown over the level-line orientation field (orthogonal to the gradient orientation field). Right: The number of aligned points up to an angular tolerance θ is counted for each segment. The segment shown receives 4 votes among 7.

The detection is posed as a hypothesis testing problem. In the classical framework statistical models for the background and for the objects to be detected are needed. In DMM they proposed an *a contrario* approach: Only a statistical model

for the background is needed; the objects are detected as outliers.

For segment detection, a suitable background model is one in which all gradient angles are independent and uniformly distributed. It can be shown that this is the case for a Gaussian white noise image. Formally, an image X from the background model H_0 is a random image (defined on the grid $\Gamma = [1, N] \times [1, M] \subset \mathbb{Z}^2$) such that: (a) $\forall m \in \Gamma$, $\mathrm{Angle}(\nabla X(m))$ is uniformly distributed over $[0, 2\pi]$; (b) The family $\{\mathrm{Angle}(\nabla X(m))\}_{m \in \Gamma}$ is composed of independent random variables.

There are as many tests T_s as there are potential segments s in the image. On a $N \times N$ image, there are N^4 potential oriented segments, starting and ending on a point of the grid Γ. Each test relies on the statistics $k(s, x)$ which is the number of aligned points in segment s and image x. The detection is as follows: Reject H_0 if $k(s, x) \geqslant k_s$, accept H_0 otherwise. For this test, non-H_0 is also denoted H_s. Desolneux *et al.* proposed to set k_s in order to control the number of false detections. False detections are those that arise by chance on the unstructured background model H_0.

They define the Number of False Alarms of a segment $s \in \mathcal{S}$ and an image x, as

$$\mathrm{NFA}(s, x) = \#\mathcal{S} \cdot \mathbb{P}_{H_0}[k(s, X) \geqslant k(s, x)],$$

where X is a random image on H_0 and $\#\mathcal{S}$ is the number of potential segments in the image. When there is no ambiguity about the image x, we will use the notation $k(s)$ and $\mathrm{NFA}(s)$. $\mathrm{NFA}(s)$ is $\#\mathcal{S}$ times the p-value of the test $k(s)$.

The smaller the $\mathrm{NFA}(s)$ the more meaningful s is, *i.e.*, the less likely it is to appear in an image drawn from the H_0 model. Rejecting H_0 if and only if $\mathrm{NFA}(s) \leqslant \varepsilon$ gives what Desolneux *et al.* call the ε-*meaningful segments*. Lets call $\mathrm{NFA}(\varepsilon) = \mathbb{E}_{H_0} \sum_{s \in \mathcal{S}} \mathbb{1}_{\mathrm{NFA}(s, X) \leqslant \varepsilon}$, *i.e.*, the expected number of ε-meaningful segments detected under the H_0 hypothesis[1]. It can be proved [7] that $\mathrm{NFA}(\varepsilon) \leqslant \varepsilon$. In this way, ε controls the average number of false detections.

The dependence of the method on ε is very weak. Actually k_s is translated of $\sqrt{\log \varepsilon}$ whenever $\varepsilon \neq 1$. In practice we fix $\varepsilon = 1$ once for all. This corresponds to accept, on average, one false detection per image on the background model.

Computations can be done explicitly. If the angle tolerance θ is set to the value $\theta = 2\pi p$, the probability that a given point have the gradient aligned with a segment is p. As the gradient is independent at different image points, $k(s)$ follows a binomial law of parameters $l(s)$ and p. On an $N \times N$ image, one has: $\mathrm{NFA}(s) = N^4 \cdot B(l(s), k(s), p)$, where $B(l, k, p)$ stands for the binomial tail, that is, the probability for a binomial of parameters l and p to be larger than k.

Fig. 1 upper-right shows the 1-meaningful segments found on the image. All the segments that we perceive are among

---

[1]Note that $\mathrm{NFA}(\varepsilon)$ is not the same as $\mathrm{NFA}(s)$. $\mathrm{NFA}(\varepsilon)$ is not attached to any segment, but to the method itself.

them. But there are a lot more. Whenever a segment has a central part with many p-aligned points, it is systematically detected as meaningful because its aligned central part still makes it very unlikely to appear by chance. A similar argument explains why redundant parallel and slightly slanted detections occur near a good one.

In order to get rid of these redundant detections an exclusion principle was proposed [7]. The value of the NFA is used as a measure of quality of the segment to select the best ones. The segment with the lowest NFA owns its points and prevents other segments from using them. Then the NFA is recomputed for all remaining segments and the procedure is repeated until there are no more meaningful segments. Fig. 1 bottom-left shows the results.

3. MULTISEGMENTS

On Fig. 1 one can see the inaccuracy of segment extremities detected by DMM. The horizontal lines of the windows are detected as a long segment instead of three smaller ones. This problem arised from the way the exclusion principle was used to select the segments. Fig. 3 shows a numerical example of the problem.
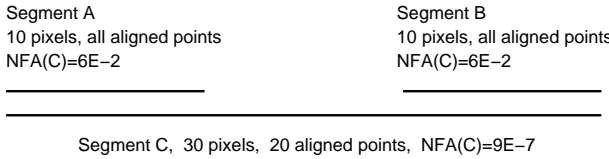
Segment A	Segment B
10 pixels, all aligned points	10 pixels, all aligned points
NFA(C)=6E−2	NFA(C)=6E−2

Segment C, 30 pixels, 20 aligned points, NFA(C)=9E−7

Fig. 3. The problem of selecting segments by exclusion principle. Segments A and B have 10 aligned points, each. In an 512 by 512 image its NFA is $6 \cdot 10^{-2}$. Segment C includes segments A and B and a gap of 10 non-aligned points. Its NFA is $9 \cdot 10^{-7}$. Segment C has the lowest NFA and will be selected by the exclusion principle applied to segments.

In [8] Delon *et al.* addressed a similar problem for histogram modes detection and explained why big modes can occlude little ones when using a similar detection principle. In [9] Cao *et al.* tackled this issue and showed that the proper way to decide whether or not a cluster C should be refined into two disjoint sub-clusters C_1 and C_2 is not only to compare the meaningfulness (measured in terms of NFA) of C_1 (*resp.* C_2) against the meaningfulness of C but also to look at the meaningfulness of the group formed by C_1 and C_2 against the meaningfulness of C. Our work is inspired by their method.

The general idea is to use a more global criterion to select straight segments. Instead of the one to one straight segment comparison, used in [1], we propose to compare different interpretations for a whole line in terms of sets of segments. This interpretation as a sequence of non-overlapping segments on a line is what we call a *multisegment*. It remains to measure the quality of a multisegment, its NFA, and then select the best one.

Given a line L, an n-multisegment with support in L is an n-tuple $(s_1, \ldots, s_n)$ of n *disjoint* segments s_i *contained* in L. The set of all n-multisegments with support in L is $\mathcal{M}(n, L)$. Given an image x and an n-multisegment $(s_1, \ldots, s_n)$, let us define the vector $k(s_1, \ldots, s_n, x) \in \mathbb{R}^n$ such that each component is the number of p-aligned pixels in the corresponding segment. Usually $k(s_1, \ldots, s_n, x)$ is simply denoted by $k(s_1, \ldots, s_n)$ when there is no ambiguity about the image x. Formally, $k(s_1, \ldots, s_n) = (k(s_1), \ldots, k(s_n))$.

An expression for the multisegment $\mathrm{NFA}(s_1, \ldots, s_n)$ can be found by the same steps as in the single segment case. The *a contrario* framework still holds: H_0 is chosen the same way as before, induced by Gaussian white noise images.

In the case of a single segment s, the test had the form $B(l(s), k(s), p) \leqslant \alpha$ with α equal to $\frac{\varepsilon}{\#S}$. The segments of a multisegment are not overlapping, which guarantees independent events. The test associated to a multisegment has the form $\prod_{i=1}^{n} B(l(s_i), k(s_i), p) \leqslant \alpha$. Now we define:

Definition 1. *For an n-multisegment $(s_1, \ldots, s_n)$ in $\mathcal{M}(n, L)$ (L is the support line) and an image x, its $\mathrm{NFA}(s_1, \ldots, s_n, x)$ is defined as*

$$\#\mathcal{L} \binom{l(L)}{2n} B(l(s_1), k(s_1), p) \prod_{i=2}^{n} (l(s_i)+1) B(l(s_i), k(s_i), p)$$

where $\#\mathcal{L}$ stands for the total number of lines in x, and $l(L)$ is the length of L. $\mathrm{NFA}(s_1, \ldots, s_n, x)$ will be abbreviated by $\mathrm{NFA}(s_1, \ldots, s_n)$ when there is no ambiguity about x.

We call ε-*meaningful multisegment* a multisegment such that $\mathrm{NFA}(s_1, \ldots, s_n) < \varepsilon$. The following proposition shows that the previous definition is coherent with the theory.

Proposition 1. *For all $n > 0$,*

$$\mathrm{NFA}(\varepsilon) \stackrel{\mathrm{def}}{=} \mathbb{E}_{H_0}[\textit{num. of } \varepsilon\textit{-meaningful } n\textit{-multisegments}] \leqslant \varepsilon$$

Let us use this new NFA to analyze the numerical example of Fig. 3. NFA now depends on $\#\mathcal{L}$ in a multiplicative way. One must know $\#\mathcal{L}$ in order to decide whether or not a given multisegment is ε-meaningful but does not require this information in order to compare two multisegments. There are 4 interesting multisegments to explain Fig. 3: just segment A, just segment B, just segment C, and the 2-multisegment formed by A and B, denoted (A,B). Lets fix $\#\mathcal{L} = 1$. The multisegment NFA values are $\mathrm{NFA}(A) = \mathrm{NFA}(B) = 1.6 \cdot 10^{-10}$, $\mathrm{NFA}(C) = 2.3 \cdot 10^{-14}$ and $\mathrm{NFA}(A, B) = 10 \cdot 10^{-19}$. This last configuration gives the best global interpretation.

The core of the algorithm just takes a binary sequence of points (aligned/not-aligned) in one line and gives the best interpretation in terms of multisegments.

A priori one has to test all potential multisegments on the sequence to select the best interpretation. This number is huge. The number of tests can be reduced by considering the runs (see Fig.4(a) for the notion of run) of all aligned

points of the sequence. It can be shown that only the multi-segments with segments that start and end with a run are interesting. Moreover, a dynamic programming algorithm can provide the result with complexity $O(r^3)$, if r is the number of runs on the line. The details will be published elsewhere. The whole algorithm tests all lines on the image. For each line it keeps the best multisegment interpretation. Then the exclusion principle is applied to avoid redundancy.

4. RESULTS

In white noise images, the number of ε-meaningful detections is controlled, being usually much smaller than ε.

Fig. 4 shows the performance on two synthetic binary sequences. Note that the multisegment approach gives the correct interpretations in both cases. The multisegment interpretation captures the structure when it is present (b), and explains the data as a noisy line when no structure is present (f). The two threshold algorithm used in HTM, (c) and (g), fails to give the right interpretation of both situations without parameter tuning. The DMM approach, (d) and (h), fails to get the small segments (d) due to the too local nature of their exclusion principle.
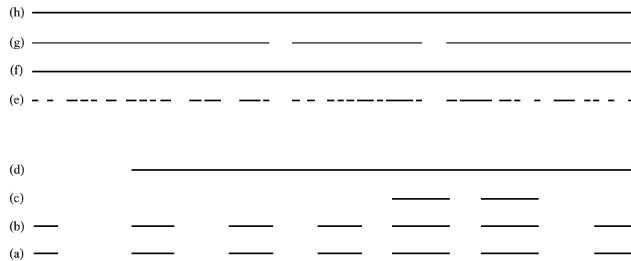

Fig. 4. Segmentation of two binary sequences. (a) A sequence of length 128 with 64 aligned points grouped into 7 runs; and its interpretations with multisegments (b), HTM (c) and by DMM (d). (e) Binary sequence generated by randomly drawing 64 out of 128 points with a uniform law over all possible such subsets; and its interpretation with multisegments (f), HTM (g) and DMM (h).

Fig. 5(a) shows an image of a building. The segments found by HTM are shown on (b). For this experiments we have used the HTM implementation of the XHoughtool package, freely available on the Internet, without tuning any parameters. One can see many false positives. Some of them can be corrected fixing the thresholds. HTM ignores the gradient orientation of the points, this produces several false detections. The DMM algorithm gives essentially good detections (c). Some segments on the image are accidently aligned. On those cases, DMM found a large segment instead of the smaller aligned ones. The multisegments result is shown on (d). This approach obtains the structure of aligned segments.
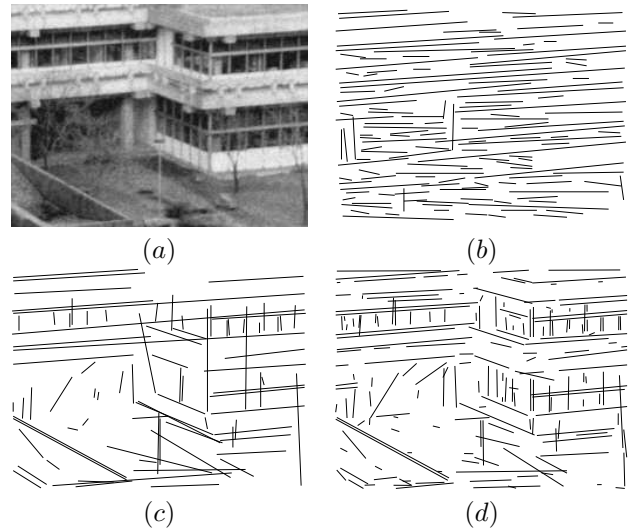
More experiments can be found online at `http://www.cmla.ens-cachan.fr/Utilisateurs/grompone/multi.htm`.

Fig. 5. A comparison of the segments detected using HTM, DMM and multisegments. (a): Image of a building. (b): Segments found by HTM. One can see many false positives. (c): Segments found by DMM. Note that some segments that are accidentally aligned on the image, are found as a large segment. (d): Segments found by the multisegment algorithm.

5. REFERENCES

[1] A. Desolneux, L. Moisan, and J.M. Morel, "Meaningful alignments," *IJCV*, vol. 40, no. 1, pp. 7–23, 2000.

[2] C.X. Ji and Z.P. Zhang, "Stereo match based on linear feature," in *ICPR88*, 1988, pp. 875–878.

[3] S. Mahadevan and D.P. Casasent, "Detection of triple junction parameters in microscope images," *SPIE*, pp. 204–214, 2001.

[4] P. Fränti, E.I. Ageenko, H. Kälviäinen, and S. Kukkonen, "Compression of line drawing images using hough transform for exploiting global dependencies," in *JCIS 1998*, 1998.

[5] O. Faugeras, R. Deriche, H. Mathieu, N.J. Ayache, and G. Randall, "The depth and motion analysis machine," *PRAI*, vol. 6, pp. 353–385, 1992.

[6] D.H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *PR*, vol. 13, no. 2, pp. 111–122, 1981.

[7] A. Desolneux, L. Moisan, and J.-M. Morel, "From gestalt theory to image analysis: A probabilistic approach," *Preprint MAP5 Num. 2006-6*, 2006.

[8] J. Delon, A. Desolneux, J.L. Lisani, and A.B. Petro, "A non-parametric approach for histogram segmentation," *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 253–261, 2007.

[9] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur, "A unified framework for detecting groups and application to shape recognition," *To appear in JMCV*, 2006.