From line segments to more organized Gestalts

Boshra Rajaei[1,2], Rafael Grompone von Gioi [1], Jean-Michel Morel[1]
[1] *CMLA, ENS Cachan, France*
[2] *Sadjad University of Technology, Mashhad, Iran*

In this paper, we reconsider the early computer vision bottom-up program, according to which higher level features (geometric structures) in an image could be built up recursively from elementary features by simple grouping principles coming from Gestalt theory. Taking advantage of the (recent) advances in reliable line segment detectors, we propose three feature detectors that constitute one step up in this bottom up pyramid. For any digital image, our unsupervised algorithm computes three classic Gestalts from the set of predetected line segments: good continuations, nonlocal alignments, and bars. The methodology is based on a common stochastic *a contrario model* yielding three simple detection formulas, characterized by their number of false alarms. This detection algorithm is illustrated on several digital images.

arXiv:1603.05763v1 [cs.CV] 18 Mar 2016

INTRODUCTION

A basic task of the visual system is to group fragments of a scene into objects and to separate one object from others and from background. Perceptual grouping has inspired many researches in psychology in the past century. In 1920's, a systematic approach for investigating human perception was introduced under the name of Gestalt theory [1]. The Gestalt school proposed the existence of a short list of grouping laws governing visual perception. The Gestalt laws include, but are not limited to, similarity, proximity, connectedness and good continuation. Gestaltists argue that partial grouping laws recursively group image elements to form more organized Gestalts. But the grouping laws were qualitative and lacked a quantitative formalization.

Since its origins, computer vision has been interested in Gestalt laws and there have been several attempts to formalize aspects of the Gestalt program [2, 3]. We will concentrate on a particular approach [4, 5] which has led to the conception of several Gestalt detectors to organize meaningful geometric structures in a digital image. Based on the *non-accidentalness principle* [2, 6], an observed structure is considered meaningful when the relation between its parts is too regular to be the result of an accidental arrangement of independent parts. This is the rationale behind the *a contrario* model for determining potential Gestalts which is formulated in the next section.

The *a contrario* methodology was used in [7], inspired by the *good continuation* Gestalt principle, to propose an algorithm to detect alignments of points in a point pattern. The detected point alignments are further employed toward a nonparametric vanishing point estimating algorithm again based on the non-accidentalness principle [8]. Based on the same principle, an automatic line segment detector (LSD) is provided in [9, 10] with linear time complexity. The same approach is used in the ED-Lines straight line calculator [11] but with the difference that the latter algorithm starts from an edge drawing output while the former is based on image level lines. Similar ideas were also proposed to detect circles and ellipses [12, 13]. Level lines are employed to detect good continuations and image corners in [14]. Using edge pieces (edgelets) the authors of [15] adopt a featured *a contrario* scheme to extract the strong edges and eliminate irrelevant and textural ones. Extracting an object of interest from textural background or from outlier points is studied in [16, 17] where, for instance, fixed contrast spots are detected in mammographical images.

The *a contrario* framework was used for image segmentation in [18]. To obtain robust segments, the authors suggest a combination of the *a contrario* approach and of Monte-Carlo simulation. Also, to solve the hierarchical segmentation issue, the authors of [19] suggest two *a contrario* criteria for measuring region and merging meaningfulness based on homogeneity and boundary contrast.

In this paper, we start from the "partial Gestalts" constituted by all LSD detected segments, and explore if they can be used as basic elements for more elaborate Gestalt groups such as long straight lines, good continuations and parallel segments. Toward this aim, the *a contrario* model is adopted on the line segment distribution. It is used repeatedly for detecting the mentioned structures as non-accidental. We shall observe in most images that, with the exception of a few isolated segments, all line segments are classified precisely and hence the algorithm forms another level of segmentation pyramid toward a complete analysis and understanding of each image's structure. The main advantage of this approach is its low time complexity due to the exploitation of line segments as input structures which are far less numerous than the unstructured set of all image pixels.

THEORETICAL MODELING

The *a contrario* detection approach has a probabilistic basis. Consider an event of interest e. According to the non-accidentalness principle, e is meaningful if its

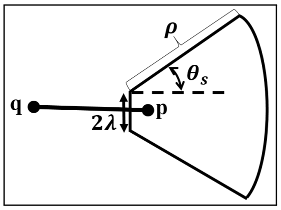

FIG. 1. Given the endpoint p of a line segment, this figure defines the domain $\mathcal{S}$ around p, in which the presence of successive segment endpoints is searched by the algorithm. This search space depends on three parameters, ρ, θ_s and λ.

expectation is low under the stochastic model H_0. The stochastic expectation of an event is called its number of false alarms (NFA) and is defined as

$$\mathrm{NFA}(e) = N_{test} P_{H_0}(e) \tag{1}$$

where N_{test} indicates the number of possible occurrences of e and $P_{H_0}(e)$ is the probability of e happening under H_0. A relatively small NFA implies a rare event e under the *a contrario* model and therefore, a meaningful one. An event e is called ϵ-meaningful if $\mathrm{NFA}(e) < \epsilon$. In this paper, three types of events are under focus: non-local alignments, good continuations and parallelism. For each of them, the *a contrario* model H_0 is adopted of stochastic line segments, with independent and uniformly distributed tips on the image domain. Here, we provide a formal definition of each event.

Definition 1. *A sequence of k line segments $l_i, l_{i+1}, \ldots, l_j$ form a potential good continuation $\wp^{k,d,\theta}$ if and only if $d = D(l_i, l_{i+1}, \ldots, l_j) < \rho$ and $\theta = \angle(l_i, l_{i+1}, \ldots, l_j) < \theta_s$ for predefined constant ρ and θ_s thresholds.*

In the above definition $D(\cdot)$ and $\angle(\cdot)$ denote respectively the maximum of the distances and the maximum of the angles between successive line segments in a sequence. The two threshold values ρ and θ_s limit the search space around each line tip for finding close line segments as represented in Figure 1. A small margin λ is allowed to deal with possible misalignments in LSD outputs. This special adopted sector is denoted by $\mathcal{S}$ in the sequel.

Definition 2. *A good continuation satisfying $\theta_s < 3°$ is called a non-local alignment event $\zeta^{k,d,\theta}$.*

Definition 3. *Two line segments or non-local alignments l_i and l_j are said to form a bar $\Im^{d,\theta}$ if and only if $d = D_m(l_i, l_j) < \rho$ and $\theta = \angle(l_i, l_j) \approx \pi$ for a predefined constant ρ threshold.*

Here, $D_m(\cdot)$ stands for the mutual distance operator which is calculated as the average distance between the respective tips of both segments.

Theorem 1. *Under the* a contrario *model that all segment tips are uniform i.i.d. spatial variables in the image domain, the number of false alarms (NFA) of a good continuation event $\wp^{k,d,\theta}$ according to Definition 1 is equal to*

$$NFA(\wp^{k,d,\theta}) = 2N \cdot 3^{k-1} \cdot \left((N-1) \frac{\theta d^2}{mn} \cdot \frac{\theta}{\pi} \right)^{k-1} \tag{2}$$

where N indicates the total number of line segments.

Proof. According to (1), the computation of the NFA consists of two parts: the number of tests and the probability of the geometric event under H_0. Here, we have $2N$ different choices for the first tip of $\wp^{k,d,\theta}$ (2 tips per line segment). Assuming that at each of the $k-1$ line segment extremities we test the 3 closest tips at most, we have 3 choices per joint and, therefore, $N_{test} = 2N3^{k-1}$.

For the probability term of $P_{H_0}(\wp^{k,d,\theta})$, let us first compute the probability that only two line segments be in good continuation as specified in Figure 1 and with parameters (d, θ). We shall then extend the computation to k segments. This probability consists of two terms: $\Pi_s \simeq \frac{\theta d^2}{mn}$ is the probability that one tip falls into $\mathcal{S}^{d,\theta,\lambda}$ around the other tip[20] and $\Pi_a \simeq \frac{\theta}{\pi}$ is the probability that a maximum angle of θ occurs between the two line segments. Therefore, having one endpoint in this area is the complement of the event where none of the $(N-1)$ other endpoints is there. Thus $P_{H_0}(\wp^{2,d,\theta}) = 1 - (1 - \Pi_s \Pi_a)^{N-1}$. Consequently, for $k-1$ junctions in a sequence of k segments we have $P_{H_0}(\wp^{k,d,\theta}) = (1 - (1 - \Pi_s \Pi_a)^{N-1})^{k-1} \simeq ((N-1)\Pi_s \Pi_a)^{k-1}$. ☐

Theorem 2. *Under the same a contrario assumption as in Theorem 1, the number of false alarms of a non-local alignment event, $\zeta^{k,d,\theta}$, according to Definition 2 is equal to*

$$NFA(\zeta^{k,d,\theta}) = 2N \cdot 3^{k-1} \left((N-1) \frac{2\lambda d}{mn} \cdot \frac{\theta}{\pi} \right)^{k-1}, \tag{3}$$

and the number of false alarms of parallel line (bar) event, $\Im^{d,\theta}$, according to Definition 3 is equal to

$$NFA(\Im^{d,\theta}) = 3N \cdot (N-1) \left(\frac{\pi d^2}{mn} \right)^2 \cdot \frac{\theta}{\pi}. \tag{4}$$

Proof. The proof of the above theorem is similar to that of Theorem 1. Since non-local alignments are special types of good continuation, their NFA calculation is straightforward. The only point is that for small $\theta < 3°$, λ is not negligible anymore. Indeed, for small angles $\mathcal{S}^{d,\theta,\lambda}$ turns into a rectangle with dimensions 2λ by d. Accordingly, by replacing $\Pi_s = \frac{2\lambda d}{mn}$ in the derivation of Theorem 1, we obtain (3).

Using similar reasoning for the case of parallel lines, since we only have two segments or straight lines, the N_{test} term reduces to $3N$. By substituting $\theta = \pi$ according to Definition 3, the overall NFA is formulated as (4). $\square$

GESTALT DETECTOR

As mentioned earlier, by employing line segments as initial groups it is possible to detect more organized and sophisticated Gestalts. In this paper, we exploit the LSD algorithm [9] to produce $l_1, l_2, \ldots, l_N$ initial line segments. The next step is finding all possible instances of our three Gestalts of interest in the form of sequences of line segments or *chains*[21].

Definition 4. *A good continuation (Definition 1) chain* $\wp^{k,d,\theta}$ *is meaningful if and only if* $NFA(\wp^{k,d,\theta}) < \epsilon$ *and there exists no subchain of* $\wp$ *with smaller NFA.*

Meaningful non-local alignments and bars follow the same definition. We assume $\epsilon = 1$ as a simple way to allow less than one false alarm on average for each Gestalt type per image [10]. Finally, by merging all meaningful Gestalts we obtain a drawing of the input image. The next section addresses the efficiency of the proposed approach applied on real-world images.

EXPERIMENTAL RESULTS

Man-made structures like buildings, furniture and natural essences present many good continuations in the form of line segments and curves or basic geometrical shapes like rectangles and parallelograms. To organize these structures using the proposed algorithm in this paper, we first applied LSD algorithm [10]. Afterwards, the two initial thresholds θ_s and ρ in Definitions 1 to 3 must be determined. These two parameters limit the number of chains that later are considered by the *a contrario* model. The parameter θ_s controls the smoothness of output Gestalts while ρ is a parameter proportional to the image size restricting the maximum acceptable distance between line segments of a candidate chain. In the sequel and to detect all kinds of parallelogram shapes, θ_s is set to $150°$ and ρ is experimentally fixed at $\min(10, \lceil 0.1 \times \max(m, n) \rceil)$ pixels by default.

Figure 2 shows the result of the joint application of all feature detectors on an image. In each output image, organized Gestalts are depicted with different colors and finally the residual line segments are shown in the last image. Note that the detected Gestalts might be grouped again into longer continuations using another level of *a contrario* model. What is remained from the image mostly only consists of isolated line segments that are not expected to belong to any organized structure. More results on more real-world or synthetic images are provided in Figure 3.

CONCLUSION

The three detectors proposed in this paper represent one step up in the Gestalt grouping pyramid. The good experimental point is that few line segments are left out unorganized, a requirement that was called “articulation without rest” in the Gestalt literature [22]. Clearly this step up must be completed by further bottom up grouping. For example, good continuation curves present gaps that must be completed with irregular contours. Other gaps in good continuations or alignments must be explained by T-junctions; bars and non-local alignments may be grouped again with the same good continuation and parallelism principles. These further steps are required to solve the figure-background problem by unsupervised algorithms.

ACKNOWLEDGMENT

Work partly founded by the European Research Council (advanced grant Twelve Labours $n°246961$).

[1] M. Wertheimer, Psychologische Forschung **4**, 301 (1923).
[2] D. Lowe, *Perceptual Organization and Visual Recognition* (Kluwer Academic Publishers, 1985).
[3] S. Sarkar and K. L. Boyer, IEEE Trans. Syst. Man Cybern. **23**, 382 (1993).
[4] A. Desolneux, L. Moisan, and J.-M. Morel, Int. J. Comput. Vision **40**, 7 (2000).
[5] A. Desolneux, L. Moisan, and J.-M. Morel, *From Gestalt Theory to Image Analysis* (Springer, 2008).
[6] I. Rock, *The Logic of Perception* (The MIT Press, 1983).
[7] J. Lezama, J.-M. Morel, G. Randall, and R. Grompone von Gioi, TPAMI **37**, 499 (2015).
[8] J. Lezama, R. Grompone von Gioi, G. Randall, and J.-M. Morel, in *CVPR* (2014) pp. 509–515.
[9] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, IEEE Trans. Pattern Anal. Mach. Intell. **32**, 722 (2010).
[10] R. Grompone von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, IPOL (2012).
[11] C. Akinlar and C. Topal, Pattern Recognit. Lett. **32**, 1633 (2011).
[12] V. Pătrăucean, P. Gurdjos, and R. Grompone von Gioi, in *ECCV* (2012).
[13] C. Akinlar and C. Topal, Pattern Recognit. **46**, 725 (2013).
[14] F. Cao, Computing and Visualization in Science **7**, 3 (2004).
[15] N. Widynski and M. Mignotte, in *Proceedings of ICSIPA* (2011) pp. 421–426.

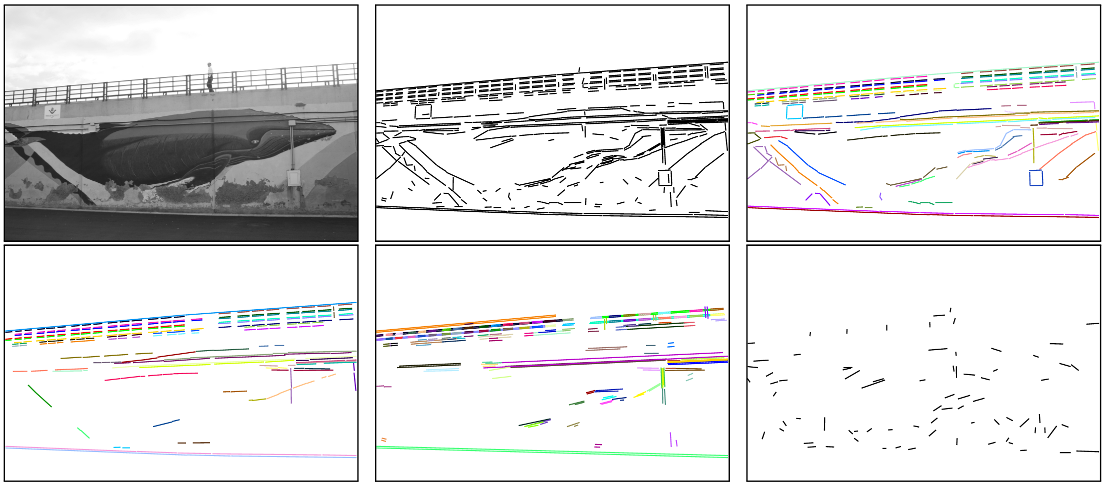

FIG. 2. Gestalt detector performance over Whale image – each color indicates one detected structure. Left to right and top to bottom: initial image, LSD line segments, all good continuations, non-local alignments, pairs of parallel segments (bars), and finally all segments that do not belong to any of the former structures. A segment can belong simultaneously to several of these higher order *partial Gestalts*. Zoom-in by a factor of 400% is recommended.

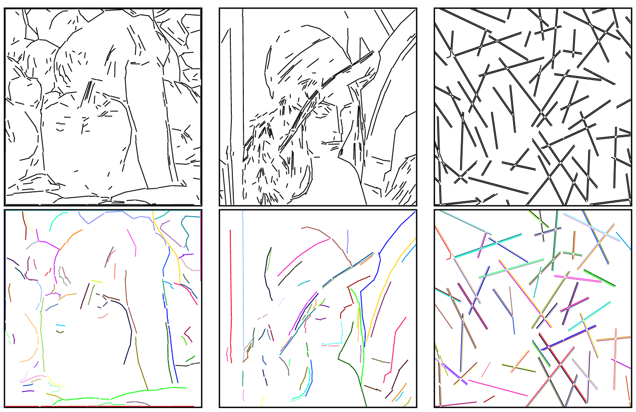

FIG. 3. Good continuation grouping – each color indicates one structure. First and second lines show LSD segments and good continuation segments, respectively, for (from left to right) Peppers, Lena and Random-Lines. In the Random-Line image as it is expected, no continuation is detected except over fragments of the same line. Zoom-in by a factor of 400% is recommended.

[16] B. Grosjean and L. Moisan, J. Math. Imaging Vision **33**, 313 (2009).

[17] D. Gerogiannis, C. Nikou, and A. Likas, IEEE Signal Processing Lett. **22**, 1638 (2015).

[18] N. Burrus, T. Bernard, and J.-M. Jolion, Pattern Recognit. **42**, 1520 (2009).

[19] J. Cardelino, V. Caselles, M. Bertalmío, and G. Randall, in *Proceedings of ICIP* (IEEE, 2009) pp. 4041–4044.

[20] Since λ is relatively small, we use this simple tight upper bound for Π_s. We also neglect the fact that the angular sector may sometimes fall outside the image domain, with area mn.

[21] A time efficient algorithm to prune the detection space can be tested in the online facility http://gestalt-detector.webs.com .

[22] G. Kanizsa, *Organization in vision: Essays on Gestalt perception* (Praeger New York:, 1979).