A Grouping Principle and Four Applications

Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel

Abstract—Wertheimer's theory suggests a general perception law according to which objects having a quality in common get perceptually grouped. The Helmholtz principle is a quantitative version of this general grouping law. It states that a grouping is perceptually "meaningful" if its number of occurrences would be very small in a random situation: Geometric structures are then characterized as large deviations from randomness. In two previous works, we have applied this principle to the detection of orientation alignments and boundaries in a digital image. In this paper, we show that the method is fully general and can be extended to a grouping by any quality. We treat as an illustration the alignments of objects, their grouping by color and by size, and the vicinity gestalt (clusters). Collaboration of the gestalt grouping laws and their pyramidal structure are illustrated in a case study.

Index Terms—Gestalt grouping laws, a contrario probabilistic model, binomial law, number of false alarms, histogram modes, clusters, alignments.

1 What Is a Partial Gestalt?

According to Gestalt theory, "grouping" is the main process in our visual perception [9], [16]. Whenever points (or previously formed visual objects) have one or several characteristics in common, they get grouped and form a new, larger visual object, a "Gestalt." Some of the main grouping characteristics are proximity (clustering), color constancy (connectedness), "good continuation" (differentiability of boundaries), alignment (presence of straight lines or objects of a same kind aligned), parallelism (between lines, oriented objects, etc.), similarity of shape (between objects), common orientation (between points or oriented objects), convexity (of boundaries, of a group), closure (for a curve), constant width, ... In addition, the grouping principle is recursive. For example, if points have been grouped into lines, then these lines may again be grouped according (e.g.,) to parallelism, etc. A square whose boundary has been drawn in black with a pencil on a white sheet is perceived by connectedness (the boundary is a black line), constant width (of the stroke), convexity and closure (of the black pencil stroke), parallelism (between opposite sides), orthogonality (between adjacent sides), finally, equidistance (of both pairs of opposite sides).

The square is a global gestalt, and the result of concurring geometric qualities that we shall call *partial gestalts*. Many Computer Vision methods attempt to compute the (very diverse in nature) partial gestalts. To take an instance, the snakes method [10] attempts to capture the closed smooth curves, a combination of the "closure" and "good continuation" gestalts. Some more recent works try to define grouping rules for combining local information into organized contours [8], [13]. In [2], we have treated alignments (straight edges) and in [3] general boundaries and edges. In this paper, we treat four more examples of partial gestalts, namely, the object alignments, the clusters, and quality grouping by color, orientation or size. In [1], a vanishing point detector is treated by a clever extension of the same method. As a first evidence of the recursive character of gestalt laws, we shall push one of the experiments to prove that the partial gestalt recursive building up can be led up to the third level (gestalts built by three successive partial gestalt grouping principles).

1.1 Helmholtz Principle

In [2], we outlined a computational method to decide whether a given partial gestalt (computed by any segmentation or grouping method) is reliable or not. We treated the detection of straight segments, as one of the most basic gestalts (see [16]). The method's main delivery are *absolute thresholds*, depending only on the image size, permitting to decide when a peak in the Hough transform is significant or not.

We applied a general perception law, the Helmholtz principle. This principle yields computational grouping thresholds associated with each gestalt quality. Assume that objects $O_1, \dots, O_n$ are present in an image. Assume that k of them, say $O_1, \dots, O_k$, have a common feature, say, same color, same orientation, etc. We are then facing the dilemma: Is this common feature happening by chance or is it significant and enough to group $O_1, \dots, O_k$? In order to answer this question, we make the following mental experiment: we assume *a contrario* that the considered quality has been randomly and uniformly distributed on all objects, i.e., $O_1, \dots, O_n$. Notice that this quality may be spatial (like position, orientation). Then, we (mentally) assume that the observed position of objects in the image is a random realization of this uniform process. We finally ask the question: Is the observed repartition probable or not? The Helmholtz principle states that, if the expectation in the image of the observed configuration $O_1, \dots, O_k$ is very small, then the grouping of these object makes sense, is a Gestalt.

Definition 1 (ε-meaningful event) [2]. *We say that an event of type "such configuration of points has such property" is* ε-meaningful, *if the expectation of the number of occurrences of this event is less than* ε *under the uniform random assumption.*

As an example of generic computation we can do with this definition, let us assume that the probability that a given object O_i has the considered quality is equal to p. Then, under the independence assumption, the probability that at least k objects out of the observed n have this quality is

$$B(p,n,k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i},$$

i.e., the tail of the binomial distribution. In order to get an upper bound of the number of false alarms, i.e., the expectation of the geometric event happening by pure chance, we can simply multiply the above probability by the number of tests we perform on the image. Let us call N_T the number of tests. Then, in most cases, we shall consider in the next sections, a considered event will be defined as ε-meaningful if

$$N_T B(p,n,k) \leq \varepsilon.$$

We call in the following the left-hand member of this inequality the "number of false alarms" (NFA). When $\varepsilon \leq 1$, we talk about meaningful events. This seems to contradict the necessary notion of a parameter-less theory. Now, it does not since the ϵ-dependency of meaningfulness is in fact a $\log \epsilon$-dependency. We refer to [2] for a complete discussion of this definition.

The general method we have just outlined can be viewed as a systematization of Stewart's "MINPRAN" method [15]. It was also proposed in the early Lowe work [12], but, to the best of our knowledge, not systematically developed.

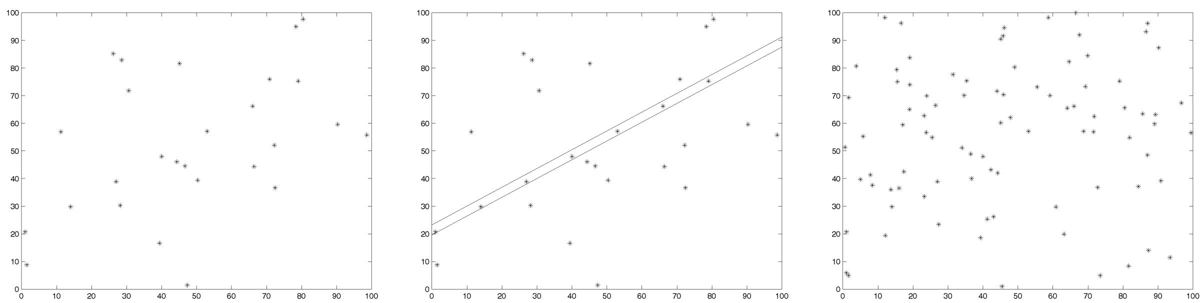
2 Object Alignments

The first partial gestalt we shall consider is a direct application of the above definition. We consider the case of objects whose barycenters are aligned. Assume that we observe M objects of a certain kind in an image. Our a contrario assumption for the application of

- *A. Desolneux is with UFR Math-Info, 45, rue des Saints-Pères, 75270 Paris cedex 06, France. E-mail: desolneux@math-info.univ-paris5.fr.*
- *L. Moisan and J.-M. Morel are with CMLA-UMR 8536, ENS Cachan, 61, avenue du Président Wilson, 94235 Cachan cedex, France. E-mail: {moisan, morel}@cmla.ens-cachan.fr.*

Manuscript received 15 Dec. 2001; revised 16 July 2002; accepted 20 Oct. 2002.
Recommended for acceptance by D. Jacobs and M. Lindenbaum.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117717.

0162-8828/03/$17.00 © 2003 IEEE Published by the IEEE Computer Society



(a) (b) (c)

Fig. 1. An illustration of Helmholtz principle: Noncasual alignments are automatically detected by Helmholtz principle as a large deviation from randomness. (a) Shows 20 uniformly randomly distributed dots and seven aligned added. (b) This meaningful (and seeable) alignment is detected as a large deviation. (c) Same alignment added to 80 random dots. The alignment is no more meaningful (and no more seeable). In order to be meaningful, it would need to contain at least 11 points.

Helmholtz principle is that the M barycenters (x_i, y_i) are i.i.d. uniformly on a domain Ω. A meaningful alignment of points must be a meaningful peak in the Hough Transform (see [11], [14] for a very similar approach). Now, the accuracy matter must be addressed. Points are supposed to be aligned if they all fall into a strip thin enough, in sufficient number. Let S be a strip of width a. Let $p(S)$ denote the prior probability for a point to fall in S, and let $k(S)$ denote the number of points (among the M) which are in S.

Definition 2. *A strip S is* ϵ-meaningful *if*

$$NFA(S) = N_s \cdot B(p(S), M, k(S)) \leq \epsilon,$$

where N_s is the number of considered strips.

2.1 The Number of Tests

We now have to discuss what the considered strips will be. In [2], we considered the relatively close problem of orientation alignments (straight edges) in a digital image. In that case, we tested *all* possible segments in the image, that is, about N^2 tests if N denotes the image number of pixels. A similar technique can be applied here to the strips and yields $N_s \simeq 2\pi(R/a)^2$, where R is the diameter of the image domain Ω and a the minimal width of a strip. One is led to sample all possible strip widths in a logarithmic scale and to sample accordingly the angles between tested strips in order to get a good covering of all directions. Thus, the number of strips N_s only depends on the size of the image and this yields a parameterless detection method. This is the first way to compute (and test) the possible strips.

2.2 Second Testing Method

Another way, which speeds up a lot the detection and makes it perceptually realistic, is to only consider strips whose endpoints are observed dots. In that case, we obtain

$$N_s = \alpha \frac{M(M-1)}{2},$$

where α denotes the number of considered widths (about 10) and $\frac{M(M-1)}{2}$ simply is the number of pairs of points. Both methods for computing N_s are valid, but they do not give the same result! Clearly, the first method is preferable in the case of a very dense set of points, assimilable to a texture, and the second method when the set of points is sparse. This second definition of N_s fits in the general Definition 2. Notice, however, the slight obvious change in the computation of $k(S)$. It denotes the number of dots that fell into the strip, with the exception, of course, of the two endpoints defining the strip.

At this point, we must answer an objection: Aren't we cheating and choosing the theory that gives the better result? We have two possible values for N_s and the smallest N_s will give the largest number of detections. When two testing methods are available, perception must obviously choose the one giving the smaller test number. Indeed, there is perceptual evidence that grouping processes may depend on density, and that different methods could be relevant for dense and for sparse patterns [17]. Hence, the second testing method we present here should be preferred for sparse distributions of points, whereas the initial model based on density would give a smaller number of tests when the number of points is large. This economy principle in the number of tests is being developed in recent works of Geman et al. [5], [6].

We compared both definitions of object alignments in the examples of Fig. 1. When we use the larger N_s corresponding to all widths (from 3 to 12 pixels) and all segments of the image, *we simply do not detect any alignment*. This is due to the testing overdose: by doing so, we have tested many times the same alignments, and have also tested many strips which contained no dots at all. The second definition of N_s happens to give a perceptually correct result. This result is displayed in Fig. 1b where we see the only detected strip. This same alignment is no more detectable on the right. The tested widths range from 2 to 16: strips thinner than 2 pixels are nonrealistic in natural (nonsynthetic) images and strips larger than 16 give no more the appearance of alignments in a 512×512 image.

3 Histogram Modes and Groups

As we mentioned in Section 1, points or objects having one or several features in common are grouped because they have this feature in common. Assume k objects $O_1, \ldots, O_k$, among a longer list $O_1, \ldots, O_n$, have some quality Q in common. Assume that this quality is actually measured as a real number. Then, our decision of whether the grouping of $O_1, \ldots, O_k$ is relevant must be based on the fact that the values $Q(O_1), \ldots, Q(O_k)$ make a *meaningful mode* of the histogram of P. Thus, the single quality grouping is led back to the question of an automatic, parameterless, histogram mode detector. This mode detector depends upon the kind of feature under consideration. We shall consider two paradigmatic cases, namely, the case of orientations, where the histogram can be assumed by Helmholtz principle to be flat, and the case of the objects sizes (areas) where the null assumption is that the size histogram is decreasing.

3.1 Meaningful Groups of Objects According to Their Orientation and to Their Gray Level

In the sequel, we quantize the possible orientations and gray levels in the usual way and we take the a contrario assumption that the M values of orientation (or gray level) are i.i.d. uniformly on $\{1, 2, \ldots, L\}$. Consider an interval $[a, b] \subset [1, L]$ and let $k(a, b)$ denote the number of objects with gestalt quality value in $[a, b]$. We define $p(a, b) = (b - a + 1)/L$ as the prior probability that an object's quality $Q(O)$ falls in $[a, b]$. With the same generic argument as in Section 1, we have Definition 3.

Definition 3. *An interval* $[a, b]$ *is* ϵ-meaningful *if*

$$NFA([a, b]) = N_i \cdot B(p(a, b), M, k(a, b)) \leq \epsilon,$$

where N_i *is the number of considered intervals* ($N_i = L(L+1)/2$). *An interval* $[a, b]$ *is said to be maximal meaningful if it is meaningful and if it does not contain, or is not contained in, a more meaningful interval.*

It can be proven in the same way as for orientation alignments [2], [4] that maximal meaningful intervals do not intersect. Thus, we get an operational definition of meaningful modes as disjoint subintervals of $[1, L]$.

3.2 Size of Objects

The preceding arguments are easily adapted to Helmholtz type assumptions on nonuniform histograms. A very generic way to group objects in an image is their similarity of size. Now, it would be a total nonsense to assume any uniform law on the objects sizes. There are several powerful arguments in favor of a statistical decreasing law for size. These arguments derive from perspective laws, or from the occlusion dead leaves model, or directly from statistical observations of natural images [7]. Our Helmholtz qualitative hypothesis is then: the prior distribution of the size of objects is **decreasing**.

Definition 4. *An interval* $[a, b]$ *is* ϵ-meaningful *(for the decreasing assumption) if*

$$NFA([a, b]) = N_i \cdot \max_{p \in \mathcal{D}} B(p(a, b), M, k(a, b)) \leq \epsilon,$$

where $\mathcal{P}_d$ *is the set of decreasing probability distributions* (p_i) *on* $\{1, 2, \dots, L\}$, *and* $p(a, b) = \sum_{i=a}^{b} p_i$.

In the same way as in the flat histogram assumption, one can define maximal meaningful intervals and prove that maximal meaningful intervals do not intersect.

4 Meaningful Groups or Clusters

4.1 Model

The cluster example is the seminal one in Gestalt theory where it is called "proximity" gestalt [9]. Assume that we see a set of dots on a white sheet and those dots happen to be grouped in one or several clusters, separated by desert regions. In order to characterize each cluster as an event with very low probability, we shall make all computations with the a contrario or *background* model that the dots have been uniformly distributed over the white sheet. This amounts to considering the dots as distributed over the sheet by a binomial process. We then call A the simply connected region, with area σ (the area of the sheet is normalized to 1, containing a given observed cluster of dots. Assume that we observe k points in A and $M - k$ outside. Then, the "cluster probability" of observing at least k points among the M inside A is given by $B(\sigma, M, k)$. It is easily checked by large deviations estimates that if k/M exceeds σ, this probability can become very small. Now, the event is not a generic event in that we have fixed a posteriori the domain A. The real a priori event we can define is "there is a simply connected domain A, with area σ, containing at least k points." The associated number of false alarms is the expected number of such domains A, that is, $N_{\mathcal{D}} B(\sigma, M, k)$, where $\mathcal{D}$ is the set of all possible domains A and $N_{\mathcal{D}}$ its cardinality.

In order to allow the number of false alarms to be small, we need to consider a small set $\mathcal{D}$ of admissible domains. To that aim, we have to *sample* the set of simply connected domains by encoding their boundaries as "low resolution" Jordan curves. We consider a low-resolution grid in the image, which for a sake of low complexity we take to be hexagonal, with mesh step m. The number of curves with length lm starting from a point and supported by the grid is bounded from above by 2^l. The overall number of low resolution curves with length lm is bounded by $N_m^2 2^l$, where $N_m = \frac{4}{3\sqrt{3}m^2}$ is the (approximate) number of mesh points lying on the sheet. Thus, we can consider several resolutions $m_1 < m_2 < \ldots < m_q$, for example in logarithmic scale, with m_1 larger than the pixel size and m_q lower than the image size, so that q is actually a small number. Our set of domains will be the set of all Jordan curves at all given resolutions, with discrete length—measured in the corresponding mesh—strictly smaller than a fixed length L (so that $\sum_{l<L} 2^l \leq 2^L$). Thus, the overall number of possible low-resolution curves (that is, $N_{\mathcal{D}}$) is bounded by $N^2 q 2^L$, where $N = N_{m_1}$. Notice that all numbers here are relatively small since the phenomenology excludes very intrincated clusters to be perceived. Thus, L is always taken to be smaller than, say, 30.

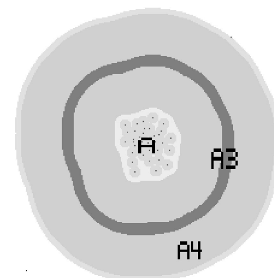


Fig. 2. The sets A_3 and A_4 associated to a cluster A.

It can also happen that a cluster is not overcrowded, but only fairly isolated from the other dots. To take this into account, we consider "thick" low-resolution curves, obtained by dilating the low-resolution curves defined above. The events we now look for include the fact that no point should fall inside the "thick" low-resolution curve defining the cluster domain A. If r is the number of allowed values for σ', the area of the "thick" curve, we can define meaningful clusters as follows.

Definition 5. *We say that a group of* k *dots (among* M*) is an* ε-meaningful *cluster if there exists an empty thick low-resolution curve (with discrete length* L *and area* σ'*) enclosing the* k *points in a domain with area* σ *such that*

$$NFA(\sigma, M, k, \sigma', L) := N^2 q r 2^L \sum_{i=k}^{M} \binom{M}{i} \sigma^i (1 - \sigma - \sigma')^{M-i} \leq \varepsilon. \quad (1)$$

4.2 Algorithm

Since the cluster detection algorithm is not obviously fast, we shall give some implementation details. Let P_i, $i = 1..M$ be the points observed. We assume that M is reasonably small, say $M \leq 1,000$. We write $d(P_i, P_j)$ for the usual Euclidean distance between P_i and P_j.

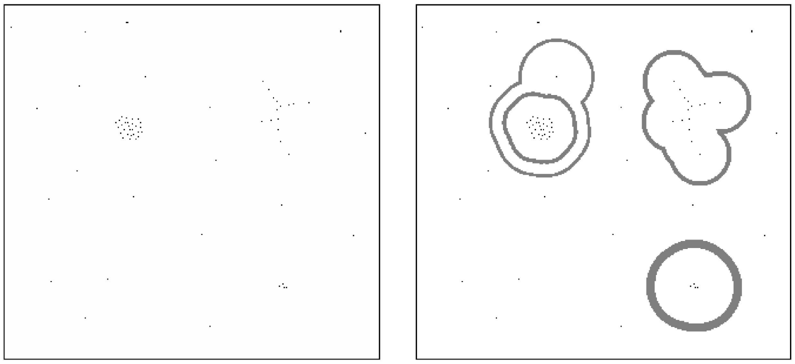
4.2.1 Computation of the Minimum Spanning Tree

$$\begin{aligned}&\text{initialization : each point } P_i \text{ is a tree}\\&\text{while there remains more than one tree}\\&\quad\text{find the 2 nearest trees and fuse them.}\end{aligned}$$

When we fuse two trees A and B, they become the two children of a new node to which we attach a value δ, the distance between A and B (that is, the minimum distance between an leaf of A and a leaf of B). The complexity of this step is $O(M^2 \log M)$ in the average, since we sort the distances $d(P_i, P_j)$ $(1 \leq i < j \leq M)$ once for all.

4.2.2 Computation of the Meaningfulness of Each Cluster

In the minimum spanning tree, each subtree associated to a root node A with value δ corresponds to a δ-cluster (named A_0) made of



(a) (b)

Fig. 3. (a) Clusters of dots and (b) their automatic detection: The thick (low resolution) curves indicate roughly the skeleton of the detected region that contains no dots. The cluster is meaningful when it contains enough points and is surrounded by a thick enough empty region.

the connected union of the disks with radius $\delta/2$ centered on the points encountered in the subtree. We compute the meaningfulness ($-\log_{10} NFA$) of each cluster with (1).

Now, the point is to estimate l (the length of the low-resolution curve), σ and σ'. For each cluster A_0, we can compute ρ, the distance of A_0 to the $\delta/2$-dilate of the remaining points. It is given by $\rho = \delta' - \delta$, where δ' is the value associated to the parent of A ($\delta' = +\infty$ if A is the root of the minimum spanning tree). If $\rho \neq 0$, we then compute, for $\alpha \in]0, 1[$ fixed,

$$A_1 = D_\rho(A_0), \quad A_2 = A_1 - E_\rho(A_1),$$
$$A_3 = E_{\rho(1-\alpha)/2}(A_2), \quad A_4 = D_{\rho(1-2\alpha)/2}(A_3),$$

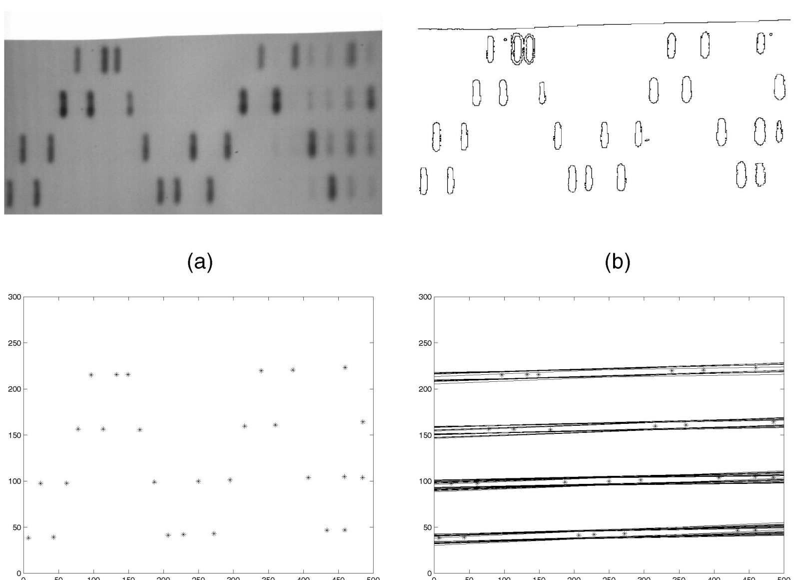
where E_r and D_r represent, respectively, the erosion and dilation operators associated to a disk with radius r (see Fig. 2). We recall that $A_0 = D_{\delta/2}(\cup_i\{P_i\})$, where the P_is are the points encountered in the subtree defined by the node A.

The domain A_3 is a "thick low-resolution curve" of width $\alpha\rho$, defined by the dilate of a low-resolution curve C' lying on the hexagonal mesh. As we do not know where C' should precisely lie in A_3, only the A_4 domain will count as "empty domain," and not $D_{\rho(1-\alpha)/2}(A_3)$. We then define

$$l = C \cdot \left\lceil \frac{\text{area}(\mathrm{A}_3)}{\alpha^2\rho^2} \right\rceil, \quad \sigma' = \text{area}(A_4), \quad \sigma = \text{area}(A_2),$$

where $\lceil\cdot\rceil$ represents the upper integer part, and C is a constant such that for any continuous curve with length l_0, there exists a discrete curve with length less than Cl_0 supported by the unit step hexagonal mesh . We conjecture that $C \leq 3/2$, and use this value in practice.

The areas mentioned can be computed using a bitmapped image with a convenient size. This computation is done for some quantized values of α, provided that the associated discrete length l satisfies $l \leq L$. In theory, we cannot choose exactly ρ but we should take the nearest smaller value among the resolutions m_i. In practice, this does not affect the computations much, since the number of resolutions chosen has very little effect on the NFA. An example of cluster detection is given in Fig. 3.



(a) (b) (c) (d)

Fig. 4. Gestalt grouping principles at work for building an "order 3" gestalt (alignment of blobs of the same size). (a) Original DNA image. (b) Maximal meaningful boundaries. (c) Barycenters of all meaningful regions whose area is inside the only maximal meaningful mode of the region areas histogram. (d) Meaningful alignments of these points.

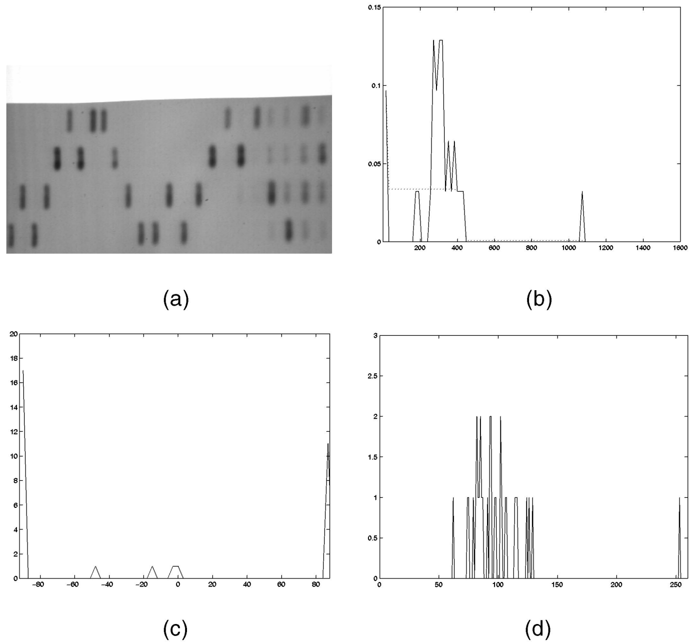


Fig. 5. Collaboration of gestalts: the objects tend to be grouped similarly by several different partial gestalts. From left to right (a) Original image. (b) Histogram of areas of the meaningful blobs. There is a unique maximal mode (256-416). The outliers are the double blob, the white background region and the three tiny blobs. (c) Histogram of orientations of the meaningful blobs (computed as the principal axis of each blob). There is a single maximal meaningful mode (interval). This mode is the interval 85-95. It contains 28 objects out of 32. The outliers are the white background region and three tiny spots. (d) Histogram of the mean gray levels inside each block. There is a single maximal mode containing 30 objects out of 32, in the gray-level interval 74-130. The outliers are the background white region and the darkest spot.

4.3 Maximal Clusters

Once we have computed the meaningfulness of each cluster, we can look for maximal meaningful clusters by selecting local maxima of the meaningfulness with respect to inclusion [2]. Precisely, we shall say that a cluster A is maximal if for any child (respectively, parent) B of A, one has $NFA(B) > NFA(A)$ (respectively, $NFA(B) \geq NFA(A)$). As usual, we have the property that two maximal meaningful clusters are either equal or have no common point.

5 Experiments

In Fig. 4, we show the application of several partial gestalt detectors to a same figure, organized according to the recursive principle we mentioned in the introduction. In Fig. 4b, we see the maximal meaningful boundaries obtained by the parameterless method described in [3]. These boundaries surround regions which we shall call "objects." Each object can be attributed several qualities, such as its barycenter, its average gray level, its orientation, etc. In Fig. 5, we show the histograms of areas, which has a single maximal mode, according to the definitions of Section 3.2. This mode corresponds to the seeable blobs and rules out the very large background regions and the three small spots detected as objects. We can proceed to look for subgroups in the group of blobs, according to other gestalt qualities. Alignments, in the sense of Section 2 can be again automatically detected. In Fig. 4c, we see the barycenters of all detected meaningful boundaries that belong to the same area histogram mode. On the right, the detected alignments are shown. We actually detect several slightly divergent strips because they all contain the same aligned points. This experiment has led to compute an "order 3" gestalt (boundary + size + alignment). As shown in Fig. 5, the final alignments would be the same if we had grouped the region by their gray level, or by their orientation. We face here one of the main challenges of Gestalt theory, namely: how to quantize the "collaboration" between different gestalt qualities.

6 Conclusion

We have shown that the automatic detection of gestalts, which we previously formalized in two applications, can be extended to several other cases. The derivation of quantitative thresholds is systematic and obeys a similar formalism in several very different cases. However, a specific discussion is required for each partial gestalt quality since each probabilistic a contrario model is specific to the partial gestalt. We have also to address sampling issues since each object space (such as lines or orientations or sizes) must be given a sampling rate and a reference histogram (see [1]), ibidem. The *collaboration* and the *recursive use* of the grouping principles have only been illustrated by hand and on a particular example. Thus, here are some quite open problems: 1) the general principles by which partial gestalts collaborate, 2) the hierarchy of gestalts and the solution of conflicts, and 3) the general principle by which a global final description is obtained. These principles exist as gestalt principles but, for the time being, do not have computational counterparts.

Acknowledgments

This work was partially supported by the US Office of Naval Research under grant N00014-97-1-0839, the *Centre National d'Etudes Spatiales*, the *Centre National de la Recherche Scientifique*, and the *Ministère de la Recherche et de la Technologie*. The authors would like to thank the Fondation des Treilles for having hosted them during the redaction of this paper.

References

[1] A. Almansa, A. Desolneux, and S. Vamech, "Vanishing Point Detection without Any A Priori Information," vol. 25, no. 4, pp. 502-507, Apr. 2003.

[2] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful Alignments," *Int'l J. Computer Vision,* vol. 40, no. 1, pp. 7-23, 2000.

[3] A. Desolneux, L. Moisan, and J.M. Morel, "Edge Detection by Helmholtz Principle," *J. Math. Imaging and Vision*, vol. 14, no. 3, pp. 271-284, 2001.

[4] A. Desolneux, L. Moisan, and J.-M. Morel, "Maximal Meaningful Events and Applications to Image Analysis," *Annals of Statistics*, pending publication, http://www. cmla. ens-cachan. fr/Cmla/Publications/2000.

[5] F. Fleuret and D. Geman, "Coarse-to-Fine Face Detection," *Int'l J. Computer Vision,* vol. 41, no. 1, pp. 85-107, 2001.
[6] D. Geman and B. Jedynak, "Model-Based Classification Trees," *IEEE Trans. Information Theory,* vol. 47, no. 3, 2001.
[7] Y. Gousseau, "The Size of Objects in Natural Images," PhD dissertation, Université Paris-Dauphine, 2000.
[8] G. Guy and G. Medioni, "Inferring Global Perceptual Contours from Local Features," *Int'l J. Computer Vision,* vol. 20, no. 1, pp. 113-133, 1996.
[9] G. Kanizsa, *Grammatica del Vedere,* Il Mulino, Bologna, 1980. Traduction française: *La grammaire du voir,* Diderot Editeur, Arts et Sciences, 1996.
[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Proc. First Int'l Computer Vision Conf.,* 1987.
[11] N. Kiryati, Y. Eldar, and A.M. Bruckstein, "A Probabilistic Hough Transform," *Pattern Recognition,* vol. 24, no. 4, pp. 303-316, 1991.
[12] D. Lowe, *Perceptual Organisation and Visual Recognition,* Kluwer Academic, 1985.
[13] A. Sha'Ashua and S. Ullman, "Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network," *Proc. Second Int'l Conf. Computer Vision,* pp. 321-327, 1988.
[14] D. Shaked, O. Yaron, and N. Kiryati, "Deriving Stopping Rules for the Probabilistic Hough Transform by Sequential Analysis," *Computer Vision and Image Understanding,* vol. 63, no. 3, pp. 512-526, 1996.
[15] C.V. Stewart, "MINPRAN: A New Robust Estimator for Computer Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, pp. 925-938, 1995.
[16] M. Wertheimer, "Untersuchungen zur Lehre der Gestalt," *Psychologische Forschung,* vol. 4, pp. 301-350, 1923.
[17] S.W. Zucker and C. David, "Points and End-Points: A Size-Spacing Constraint for Dot Grouping," *Perception,* vol. 17, pp. 229-247, 1988.

Path-Based Clustering for Grouping of Smooth Curves and Texture Segmentation

Bernd Fischer, *Student Member*, *IEEE*, and
Joachim M. Buhmann, *Member*, *IEEE*

Abstract—Perceptual Grouping organizes image parts in clusters based on psychophysically plausible similarity measures. We propose a novel grouping method in this paper, which stresses connectedness of image elements via mediating elements rather than favoring high mutual similarity. This grouping principle yields superior clustering results when objects are distributed on low-dimensional extended manifolds in a feature space, and not as local point clouds. In addition to extracting connected structures, objects are singled out as outliers when they are too far away from any cluster structure. The objective function for this perceptual organization principle is optimized by a fast agglomerative algorithm. We report on perceptual organization experiments where small edge elements are grouped to smooth curves. The generality of the method is emphasized by results from grouping textured images with texture gradients in an unsupervised fashion.

Index Terms—Clustering, perceptual grouping, texture segmentation, resampling.

1 Introduction

Image interpretation and recognition of image structure and of image context is one of the main goals of computer vision. The information loss between 3D objects and 2D images is compensated to some extent by perceptual organization rules in biological vision, which generates a holistic percept from local measurements. Perceptual organization helps to provide additional information about the 3D object and it extracts important information about a scene. This processing step reduces the size of the image data significantly. Among the central algorithmic procedures for perceptual organization are clustering principles like generalized k-means methods or clustering methods for proximity data [1], [2]. Features in images like short edge pieces or local textured image patches, are grouped in such a way that these objects are mutually very similar and might even be replaced by a prototypical representative.

This grouping principle, however, is not applicable in situations where local continuity and similarity of features is used to group them together, although they might be very different on a global scale. Image patches with a strong texture gradient or short edge pieces of smooth but moderately curved boundaries belong to this class of clustering problems. We propose in this paper, a new grouping approach referred to as *Path-Based Clustering* [3], which measures local homogeneity rather than global similarity of objects. The objects are small edge elements with a position and a direction, called edgels. The costs function favors groups of edgels which form smooth curves and separate those structures from noisy distractors which are interpreted as random fluctuations in the background.

First, Path-Based Clustering with automatic detection of outliers is mathematically described in Section 3. The new Path-Based Clustering method defines a connectedness criterion, which groups objects together if they are connected by a sequence of

• *The authors are with the Department of Computer Science III, Rheinische Friedrich-Wilhelms-Universität, Römerstr. 164, 53117 Bonn, Germany. E-mail:{fischerb, jb}@cs.uni-bonn.de.*

Manuscript received 22 Dec. 2001; revised 9 July 2002; accepted 22 Oct. 2002. Recommended for acceptance by D. Jacobs and M. Lindenbaum. For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117715.

0162-8828/03/$17.00 © 2003 IEEE Published by the IEEE Computer Society