Grouping and Segmentation

Quantitative Analysis of Grouping Processes

Arnon Amir and Michael Lindenbaum

Computer Science Department, Technion, Haifa 32000, ISRAEL
arnon,mic@cs.technion.ac.il

Abstract. This paper presents a quantitative approach to grouping. A generic grouping method, which may be applied to many domains, is given, and an analysis of its expected grouping quality is done. The grouping method is divided into two parts: Constructing a graph representation of the geometric relations in the data set, and then finding the "best" partition of the graph into groups. Both stages are implemented using known statistical tools such as Wald's SPRT algorithm and the *Maximum Likelihood* criterion. The accompanying quantitative analysis shows some relations between the data quality, the reliability of the grouping cues and the computational efforts, to the expected grouping quality. To our best knowledge, such an analysis of a grouping process is given here for the first time. The synthesis of specific grouping algorithms is demonstrated for three different grouping tasks and domains. Experimental results show the ability of this generic approach to provide successful algorithm in specific domains.
Keywords : Grouping Analysis, Perceptual Grouping, Performance Prediction, Generic Grouping Algorithm, Graph Clustering, Maximum Likelihood, Wald's SPRT.

1 Introduction

This paper presents a quantitative approach to grouping, which contains a generic grouping method, and focuses on analyzing the relation between the information available to the grouping process and the corresponding grouping quality.

The proposed method separates between two components of the grouping method: the *grouping cues* that are used and the *grouping mechanism* that combines them into a partition of the data set. Our grouping process is based on a special graph representation, in which the vertices are the observed data elements (edges,pixels, etc.) and the arcs contain the grouping information and are estimated by cues. (Others, e.g. [8, 4, 11], have used graphs for grouping algorithms, but we use it differently here.) The hypothesized grouping is a partition of the graph which maximize a functional over all the possible partitions. In contrast to most other grouping methods, which depend on the domain in which the grouping is done, this grouping mechanism is domain independent.

Good cues are essential for successful grouping, but finding them is not our aim here. Instead we consider the cues as given, model them as random variables, and quantify their reliability using the properties of the corresponding distribution. Moreover, we suggest a general method, called the *cue enhancement*

procedure, for improving the reliability of grouping cues, and show an interesting tradeoff between the computational efforts and the achievable reliability of the enhanced cue.

Unlike other grouping methods, the proposed method provides, for the first time, some relations between the quality of the available data, the computational effort invested, and the grouping performance, quantified by several measures.

2 The Grouping Task and its Graph Representation

The *grouping task* is a partitioning problem. Let $S = \{v_1, v_2, \ldots, v_N\}$ be the set of data elements, which may consist, for example, of the boundary points in an image. S is naturally divided into several *groups* (disjoint subsets) so that all data elements in the same group belong to the same object, lie on the same smooth curve, or associated with each other in some other manner. $S = S_0 \cup S_1 \cup S_2 \cup \ldots \cup S_M$. In the context of the grouping task the data set is given but its partition is unknown and should be inferred from indirect information given in the form of grouping cues [1].

Grouping cues are the building blocks of the grouping process and shall be treated as the only source of information available for this task. The grouping cues are domain-dependent and may be regarded as scalar functions $C(A)$ defined over subsets $A \subset S$ of the data feature set. Such cue functions should be discriminative, and should also be invariant to change of the viewing transformation and robust to noise [6]. At this stage we consider only bi-feature cues, defined over data subsets including two elements ($|A| = 2$). Bi-feature cues may be either the cues used by most common grouping processes, or the result of the *cue enhancement procedure*, which accumulates statistical information by using multi-feature cues, and is described in Sec 5. From now on we shall use the notation $C(e)$ for the bi-feature cues, where $e = (u, v)$, $u, v \in S$ is also the arc connects the nodes u, v in the following graph representation.

A *reliability measure* for grouping cues, which is domain-independent, is specified as follows: Consider the cue function to be a random variable, the distribution of which depends on whether the two data features belong to the same group or not. For binary cues, which provides only negative or positive answers, this dependency is simply quantified by two error probabilities: ϵ_{miss} is the probability that the cue $C(A)$ indicates a wrong negative answer, and ϵ_{fa} is the probability that the cue indicates a wrong positive answer (false alarm). If both $\epsilon_{miss} = 0$ and $\epsilon_{fa} = 0$, then $C(A)$ is an ideal cue. This characterization can sometimes be calculated using analytical models (e.g. [6]), and can always be approximated using Monte-Carlo experimentations.

Both the unknown partition into groups, and the data available from the cues, are represented using graphs. The nodes of all the graphs are the observed data elements, $V = S$, but the arcs may take different meanings, as explained

[1] We should also mention, that according to another grouping concept the hypothesized groups are not necessarily disjoint. We believe that at least some of the tools developed here are useful for the other approaches.

in Figure 1. The unknown partition, which is to be determined, is represented by the *target graph*, $G_t = (V, E_t)$, composed of several disconnected complete subgraphs (cliques). Every clique represents a different object (or group). A graph with this characterization is called a *clique graph* and the class of such graphs is denoted $\mathcal{G}_c$. We shall denote by $E_c(V_j)$ the arcs of a complete subgraph (clique) $V_j \subseteq V$. Knowing that $G_t \in \mathcal{G}_c$, the grouping algorithm should provide a *hypothesis graph*, $G_h = (V, E_h) \in \mathcal{G}_c$, which should be as close as possible to G_t.

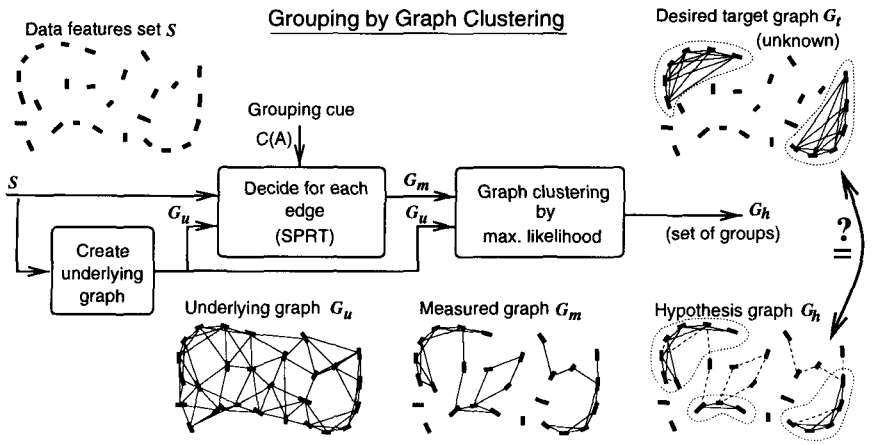


Fig. 1. The proposed grouping process: The image is a set of data features (edgels in this illustration) every one of which is represented by a node of a graph. The designer should decide about a cue and about the set of feature-pairs to be evaluated using this cue. This set of feature-pairs is specified by the arcs of the *underlying graph* G_u. The first step of the algorithm is to use *grouping cues* to decide, for every feature pair in $G_u = (V, E_u)$, if both data features belong to the same group. These decisions are represented by the a *measured graph* $G_m = (V, E_m)$: every arc corresponds to a positive decision (hence $E_m \subseteq E_u$). The known reliability of these decisions is used in the second and last step to find a maximum likelihood partitioning of the graph, which is represented by the hypothesized (clique) graph G_h. A main issue considered in this paper is the relation between this hypothesis G_h and the ground truth *target graph*, G_t, which is unknown.

3 The Generic Grouping Algorithm

The algorithm consists of two main stages: cue evaluation for (many) feature pairs and maximum likelihood graph partitioning. Before these stages, two decisions should be made by the designer. The first is to choose an appropriate grouping cue. The second one is to choose the set of feature-pairs to be evaluated using this cue. This set of feature-pairs is specified by the arcs of the *underlying*

graph G_u. In principle, all feature pairs, corresponding to a *complete underlying graph*, $G_u = (V, E_c(V))$, should be evaluated using the cue, in order to extract the maximal information. Some cues are meaningful, however, only for near or adjacent data elements (e.g. while looking for smooth curves). In such cases, a locally connected underlying graph should be used. Another consideration which affects the choice of the underlying graph is the reliability of the grouping process and the computational effort invested in it. At this stage we assume that both the cue and the associated adequate "topology" are either given or chosen intuitively.

In the first stage of the grouping process, every feature pair, $e = (u, v)$, corresponds to an arc in $G_u = (V, E_u)$ is considered, and the cue function is used to decide whether the two data features belong to the same group, or not. The positive decisions are represented as the arcs of the *measured graph* $G_m = (V, E_m)$. G_m carries the information accumulated in the first stage to the second one.

Recall that every arc-decision made in the first stage is modeled as a binary random variable, the statistics of which depends on whether the two data features belong to the same group or whether they not. In the context of this paper, the cue decisions assumed to be independent and identically distributed, and are characterized by two error probabilities [2]:

$$\epsilon_{miss}(e) = Prob(e \in (E_t \cap E_u) \backslash E_m) \qquad \epsilon_{fa}(e) = Prob(e \in E_m \backslash (E_t \cap E_u)) \quad (1)$$

The likelihood of the measurement graph, G_m, for every candidate hypothesis $G = (V, E) \in \mathcal{G}_c$, is then given by

$$L\{G_m|G\} = \prod_{e \in E_u} L\{e|E\} \quad \text{where} \quad L\{e|E\} = \begin{cases} \epsilon_{miss} & \text{if } e \in E \backslash E_m \\ \epsilon_{fa} & \text{if } e \in E_m \backslash E \\ 1 - \epsilon_{miss} & \text{if } e \in E \cap E_m \\ 1 - \epsilon_{fa} & \text{if } e \notin E \cup E_m \ . \end{cases} \quad (2)$$

We propose now to use the maximum likelihood principle, and to hypothesize the most likely clique graph

$$G_h = \arg \max_{G \in \mathcal{G}_c} L\{G_m|G\}. \quad (3)$$

The maximum likelihood criterion specifies the (not necessarily unique) grouping result, G_h, but is not a constructive algorithm. We therefore address the theoretical aspect and the practical side separately.

From the theoretical point of view, we shall now assume that the hypothesis which maximizes the likelihood may be found, and address our main question: **"what is the relation between the result G_h, and the unknown target graph G_t?"** This question is interesting because it is concerned with predicting the grouping performance. If we can show that these two graphs are close in some sense, then it means that algorithms which use the maximum likelihood principle

[2] This definition of $(\epsilon_{miss}, \epsilon_{fa})$, in terms of the graph notation, is identical to the previous one, which refers to the cue reliability.

have predictable expected behavior and that even if we can't know G_t, the grouping hypothesis G_h they produces is close enough to the true partitioning. This question is considered in the next section.

From the practical point of view, one should ask if this optimization problem can be solved in a reasonable time. Some people use simulated annealing to solve similar problems [4]. Others use heuristic algorithms [9]. We developed a heuristic algorithm which is based on finding seeds of the groups, which are (almost) cliques in G_m. Then it makes iterative modifications, using a greedy policy, until a (local) maximum of the likelihood function is obtained. In our experiments this algorithm performs nicely. More details can be found in [2].

4 Analysis of The Grouping Quality

This section quantifies some aspects of the similarity between the unknown scene grouping (represented by G_t), and the hypothesized maximum-likelihood grouping (represented by G_h). We provide a fundamental claim, and two of its results. The fundamental claim provides a necessary condition, satisfied by any partition selected according to the maximum likelihood principle. Consider two nodes-disjoint subsets V_i, V_j of the graph $G = (V, E)$, and denote their cut by $J(V_i, V_j) = \{e = (u, v) | u \in V_i,\ v \in V_j\}$. Let $l_u(V_i, V_j) = |J(V_i, V_j) \cap E_u|$ denote the cut width relative to the underlying graph. Similarly, let $l_m(V_i, V_j) = |J(V_i, V_j) \cap E_m|$ denote the cut width relative to the measurement graph ($l_m \leq l_u$). Then,

Claim 1. necessary condition: *Let* $G_h = (V, E_h)$, $V = \{V_1 \cup V_2 \cup \ldots\}$, $E_h = \{E_c(V_1) \cup E_c(V_2) \cup \ldots\}$ *be the maximum likelihood hypothesis (satisfying eq. (3)), and let* $\alpha = \left(1 + \frac{\log(\epsilon_{fa}/(1-\epsilon_{miss}))}{\log(\epsilon_{miss}/(1-\epsilon_{fa}))}\right)^{-1}$ *Then,*
1. For any bisection of any group $V_i = V_i' \cup V_i''$ $(V_i' \cap V_i'' = \emptyset)$,

$$l_m(V_i', V_i'') \geq \alpha l_u(V_i', V_i'').$$

2. For any two groups V_i, V_j, $i \neq j$,

$$l_m(V_i, V_j) \leq \alpha l_u(V_i, V_j).$$

Proof. For proving the first part, consider the likelihood ratio between two hypotheses: One is G_h and the other, denoted $\tilde{G}_h$, is constructed from G_h by separating V_i into two different groups, V_i' and V_i''. Denote $l_m = l_m(V_i', V_i'')$, $l_u = l_u(V_i', V_i'')$. Then

$$\frac{L\{G_m|G_h\}}{L\{G_m|\tilde{G}_h\}} = \prod_{e \in J(V_i', V_i'') \cap E_u} \frac{Pr\{e|E_h\}}{Pr\{e|\tilde{E}_h\}} = \left(\frac{1-\epsilon_{miss}}{\epsilon_{fa}}\right)^{l_m} \left(\frac{\epsilon_{miss}}{1-\epsilon_{fa}}\right)^{l_u - l_m}.$$

(arcs of $E_u \setminus J(V_i', V_i'')$ do not affect that ratio, and therefore are not counted). This likelihood ratio is an increasing function of l_m and is larger than 1, for $l_m \geq \alpha l_u$. Therefore, if the claim is not satisfied, then $\tilde{G}_h$ is more likely then G_h which contradicts the assumption that (3) holds. The second part of the claim is proved in a similar manner.

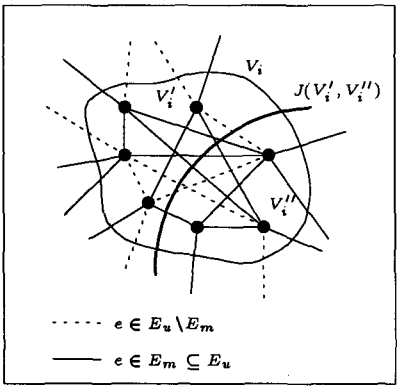


Fig. 2. The cut involved in splitting a group into two (proof 1).

Qualitatively, the claim shows that a maximum likelihood grouping must satisfy local conditions between many pairs of feature subsets. It further implies that a grouping error, either in the form of adding an alien data feature to a group (denoted *addition error*) or deleting its member (denoted *deletion error*), requires a substantial number of false alarms, or misses, respectively. The parameter α, specifying the fraction of cut edges required to merge two subsets reflects the expected error types ; if $\epsilon_{fa} = \epsilon_{miss}$, then $\alpha = 0.5$, while if $\epsilon_{fa} > \epsilon_{miss}$ then $\alpha > 0.5$. This claim implies that choosing a sufficiently dense underlying graph can significantly improve the grouping performance, and compensate for unreliable cues.

Two cases were considered in [2]: A complete underlying graph, and a locally connected underlying graph. A complete underlying graph provides the maximal information. Therefore, it may lead to excellent grouping accuracy. The next claim, given here as an example, bounds the probability of getting k' addition errors or more:

Claim 2. *Let G_u be a complete graph. Let S_i and V^* denote a true group and a maximum likelihood hypothesized group containing at least k nodes of S_i. Then, the probability that V^* contains k' nodes or more which are alien to S_i, is at most*

$$p_{k'-aliens} \leq \sum_{j=k'}^{N-k} \binom{N-k}{j} \sum_{i=k_{min}(j)}^{kj} \binom{kj}{i} \epsilon_{fa}^{i} (1-\epsilon_{fa})^{kj-i} \qquad (k_{min}(j) = \lceil \alpha k j \rceil) \tag{4}$$

(for proof see [2]). This upper bound is plotted in Figure 3(Right). Other results in [2] provides the bound for the probability of k-deletion errors, the expected number of such errors, and more. These results simply state that if the original group S_i is big enough and $\epsilon_{miss}, \epsilon_{fa}$ are small enough, it is very likely that the maximum likelihood partition will include one group for each object, containing most of S_i, and very few aliens. Experimental results for these two grouping error types are given in Figure 6 (d,e) and discussed in Section 6.

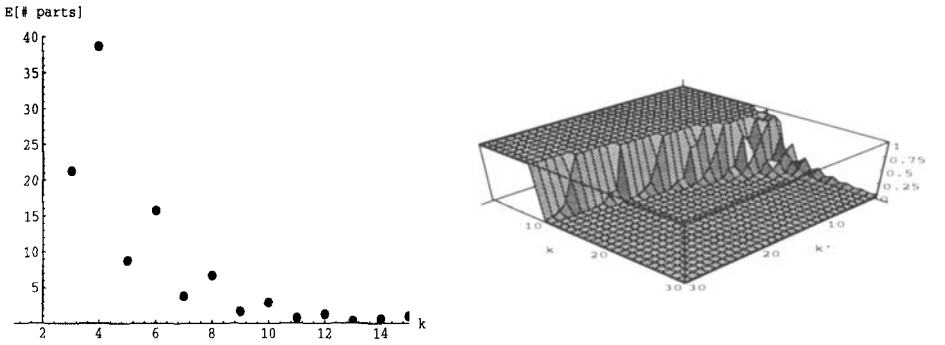


Fig. 3. Two predictions of the analysis: Left: A k-connected curve-like group (e.g. smooth curve) is likely to break into a number of sub-groups. The graph shows an upper bound on the expected number of sub-groups versus the minimal cut size in the group, k. Here the group size (length) is 400 elements, $\epsilon_{miss} = 0.14$ and $\epsilon_{fa} = 0.1$ (typical values for images like Figure 8 (a)). It shows how increasing connectivity quickly reduces the false division of this type of groups. Right: Upper bound on the probability for adding any k' alien data features to a group of size k, using a complete underlying graph (claim 2). This probability is negligible for $k > 15$ (Here $\epsilon_{miss} = \epsilon_{fa} = 0.2$).

Another prediction, given in Figure 3 (Left), corresponds to a second case, where the underlying graph is not a complete graph, but a locally-connected one where every node is connected to the closest k data features. This is useful, for example, for the common curve-like groups, in which all data features are ordered along some curve (e.g the smooth curve experiment). The grouping process may divide a long curve into a number of fragments. An upper bound on the expected number of these parts is shown in Figure 3(Left).

5 The Cue Enhancement Procedure (CEP)

The performance of the grouping algorithm depends very much on the reliability of the cues available to it. This section shows how the reliability of a grouping cue can be significantly improved by using statistical evidence accumulation techniques. This method is not restricted only to our grouping algorithm, and can be used also in other grouping algorithms. Two of the three given grouping examples (co-linearity and smoothness) use this procedure.

The CEP considers one pair of data features, $e = (u, v)$, $e \in E_u$, at a time, and uses some of the other data features in order to decide whether or not this pair is *consistent* (belongs to the same true group). Its result serves as a very reliable *binary-cue*, $C(e)$, as defined in Section 2. The key to CEP is the use of *Multi-feature cues*, $C(A)$, associated with three data features or more. The important observation here is that if the pair, $e = (u, v)$, is not consistent, then

any larger subset A, $u, v \in A$ is not consistent as well. Therefore, the (multi-feature) cue, $C(A)$, which test for the consistency of A, carries some statistical information on the consistency of e. To evaluate $C(e)$, the CEP draws several random data subsets, $A_1, A_2, \ldots$, of size $k > 2$, which contain the pair e, and find $c_j = C(A_j), \quad j = 1, 2, \ldots$. $C(A)$ is a deterministic function of A, but c_j may be considered as an instance of a random variable, c, as some part of A is randomly selected.

The statistics of c depends on the data pair e, and in particular, on its consistency. A conclusive reliable decision on the consistency of e is determined adaptively and efficiently by a well-known method for statistical evidence integration: Wald's *Sequential Probability Ratio Test (SPRT)* algorithm [10].

The distribution of c depends on an unknown binary parameter (the consistency of e), which takes the value of ω_0 (false) or ω_1 (true). The SPRT quantifies the evidence obtained from each trial by the log-likelihood ratio function of its result $h(c) = \ln \frac{p_1(c)}{p_0(c)}$ where $p_i = Pr\{c|\omega_i\}$ $i = 0, 1$ are the probability functions of the two different populations and c is the value assigned to the random variable in this trial (see Figure 5). When several trials are taken, the log-likelihood function of the composite event $\bar{c} = (c_1, c_2, \ldots, c_n)$ should be considered. If, however, the trials are independent then this composite log-likelihood function becomes $\sigma_n = \sum_{j=1}^{n} h(c_j)$. The sum σ_n serves as the statistics by which the decision is made. The SPRT-based cue enhancement procedure is summarized in Figure 5. The upper and lower limits, $a > 0 > b$, depend only on the required

For every feature pair $e = (u, v)$ in the underlying graph:

1. **Set the evidence accumulator, σ, and the trials counter, n, to 0.**
2. **Randomly choose $k - 2$ data features $x_3, \ldots, x_k \in S \backslash \{u, v\}$**
3. **Calculate $c = C(\{u, v, x_3, \ldots, x_k\})$.**
4. **Update the evidence accumulator $\sigma = \sigma + \log \frac{P_1(c)}{P_0(c)}$.**
5. **if $\sigma \geq a$ or if $n \geq n_0$ and $\sigma > 0$, output: (u, v) is consistent.**
   if $\sigma \leq b$ or if $n \geq n_0$ and $\sigma < 0$, output: (u, v) is inconsistent.
   else, repeat (2)-(5)

Fig. 4. The SPRT-based Cue Enhancement Procedure (CEP)

cue reliability $\epsilon_{miss}, \epsilon_{fa}$ (defined in eq. 1), which are specified by the user, and do not depend on the distribution of the random variable c. We calculate a, b using a practical approximation, proposed by Wald [10], which is very accurate when $\epsilon_{miss}, \epsilon_{fa}$ are small: $a = \log(\frac{1-\epsilon_{miss}}{\epsilon_{fa}})$ $\quad b = \log(\epsilon_{miss}(1 - \epsilon_{fa}))$. The derivation of $P_1(c), P_0(c)$, which depends on $\epsilon_{miss}, \epsilon_{fa}$ defined before, and on some combinatorial considerations, is given in [2].

The basic SPRT algorithm terminates with probability one and is optimal in the sense that it requires a minimal expected number of tests to obtain the

required decision error [10]. This expected number of tests is given by:

$$E\{n|\omega_0\} = [a\epsilon_{fa}+b(1-\epsilon_{fa})]/\eta_0 \qquad E\{n|\omega_1\} = [a(1-\epsilon_{miss})+b\epsilon_{miss}]/\eta_1 \tag{5}$$

where η_0, η_1 are the conditional expected amounts of evidence from a single test: $\eta_i = E\{h(c)|\omega_i\} \quad i = 0, 1$. The maximal allowed trials number, n_0, is set to be few times larger than $E\{n\}$.

Claim 3. *Given that*

(a) *The statistics of the cue values evaluated over all data subsets containing a consistent (inconsistent) arc is approximately the same, and*

(b) *The cues extracted from two random subsets including the same feature pair are independent identically distributed random variables,*

then the CEP can identify the consistency of the feature pair e within any specified error tolerance, ϵ_{miss}, ϵ_{fa}, irrespective of the reliability of the basic cue, $C(A)$.

Arbitrarily high performance is practically impossible because it requires a large number of trials leading to a contradiction of the independence assumption. Therefore, the reliability of the basic cue, $P_0(c), P_1(c)$, is important to achieve a lower expected-number of trials, $E\{n\}$. Indeed, our experiments show that the SPRT significantly improves the cue reliability but that the achievable error rate is not arbitrarily small (see Section 6). In the co-linearity experiments, $C(A)$ is a binary cue that depends on a threshold. A threshold cause a tradeoff between the miss to the false alarm ratios of the cue. For any given required reliability of the CEP , $(\epsilon_{miss}, \epsilon_{fa})$, we use eq. (5) to find the optimal threshold level, which minimize $E\{n\}$ (See Figure 5).

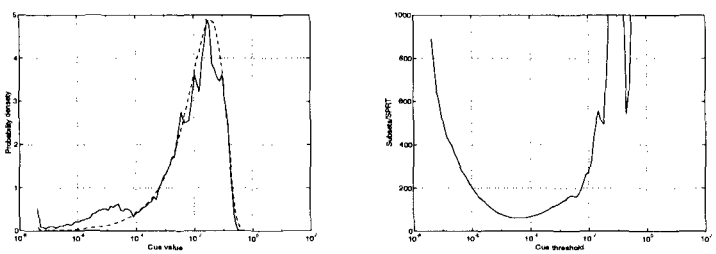


Fig. 5. Left: The two distributions of the co-linearity multi-feature cue, $p_0(A)$ (dashed), and $p_1(A)$ (solid). Although these two are very similar, their populations can be distinguished with less than 5% error (see Figure 6(d,e,f)). Right: The expected number of trials needed for the cue enhancement procedure as a function of the selected cue threshold. The optimal cue threshold correspond to the minima of this curve, as shown in Figure 5 (Right).

For a constant specified reliability $(\epsilon_{miss}, \epsilon_{fa})$, the expected run-time of the cue enhancement procedure is constant. The expected total run-time for evaluating all the arcs of the *underlying graph*, G_u, is, therefore, linear in the number of arcs.

6 Simulation and experimentation

This section presents three different grouping applications, implemented in three different domains, as instances of the generic grouping algorithm described above (see Table 1). The aim of these examples is to show that useful grouping algorithms may be obtained as instances of the generic approach and to examine the performance predictions against experimental results. (For all the technical details, more results and other examples please refer to [2].) The first example

Table 1. The three instances of the generic grouping algorithm

	The 1st example (Co-linear points)	The 2nd example (Smooth Curves)	The 3rd example (Motion Segm)
data elements	points in R^2	edgels	patches of Affine optical flow
grouping cues	co-linearity	co-circularity and proximity	consistency with Affine motion
Cue's extent	global	local	global
Enhanced cue	subsets of 3 points	subsets of 3 edgels	–
underlying graph	complete graph	locally connected graph	a complete graph
grouping mechanism:	maximum likelihood graph clustering (same program)		

is of grouping points by co-linearity cues. Given a set of points in R^2 (or in R^n), the algorithm should partition the data into co-linear groups (and one background set). To remove any doubt, we do not intend to propose this example as an efficient (or even reasonable) method for detecting co-linear clusters (Hough transform or RANSAC, for example, are better methods). We have chosen this example because it is a characteristic example of grouping tasks associated with globally valid cues (and complete underlying graphs). Moreover, it provides a convenient way for measuring grouping performance, the quantification and prediction of which is our main interest here.

We consider synthetic random images containing randomly drawn points (e.g Figure 6 (a)). A typical grouping result is shown in Figure 6 (b,c). Few of the quantitative results show the effect of the cue reliability on the overall grouping quality (Figure 6(d,e)) and on the CEP computational time (Figure 6(f)). Regardless the choice of $(\epsilon_{miss}, \epsilon_{fa})$, all the 5 lines were always detected as the 5 largest groups in our experiments. The selection of $(\epsilon_{miss}, \epsilon_{fa})$ does affects, however, the overall grouping quality. This is measured by counting the addition errors and the deletion errors, as shown in Figures 6 (d), and (e), respectively.

The second example is the grouping of edgels by smoothness and proximity. Starting from an image of edgels, (data feature = edge location + gradient direction), the algorithm should group edgels which lie on the same smooth curve. This is a very useful grouping task, considered by many researchers (see, e.g [3, 12, 4, 7]). We test this procedure both on synthetic and real images, and the results are very good in both cases (see Figure 8 and Figure 9).

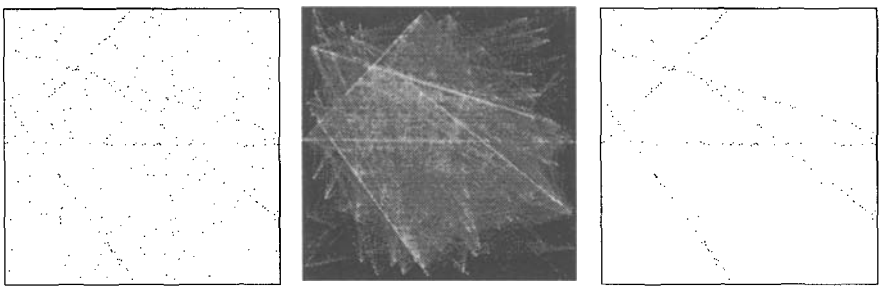
The third grouping algorithm is based on common motion. The data features are pixel blocks, which should be grouped together if their motion obeys the same rule, that is if the given optical flow over them is consistent with one Affine motion model [5, 1]. Technically, every pixel block is represented by its location and six parameters of the local Affine motion model (calculated using Least Squares). No cue enhancement is used here, and the cue is not very reliable: typical error probabilities are $\epsilon_{miss} = 0.35$ and $\epsilon_{fa} = 0.2$. Still, the results are comparable to those obtained by a domain specific algorithm [1]. The final clustering result is shown in Figure 7 (Right).

7 Discussion

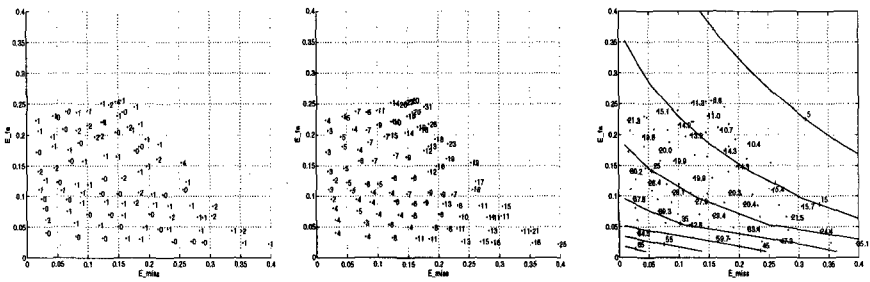
The goal of this work is to provide a theoretical framework and a generic algorithm that may be applied to various domains and that have predictable performance. The proposed algorithm relies on established statistical techniques such as sequential testing and maximum likelihood, which are well known. However, this paper is distinctive from previous approaches because it provides, for the first time, an analysis of the use of these principles, which relates the expected grouping quality to the cue reliability, the connectivity used, and in some cases the computational effort invested. We did not limit ourselves to the theoretical study: three grouping applications, in different domains, are implemented as instances of the generic grouping algorithm. Although we made an argument against visually judging the merits of vision algorithm, we would like to indicate here that our results are similar to those obtained by domain specific methods (e.g. [7, 4] for smoothness based grouping). Note that G_m may also be used to create a saliency map, by specifying the saliency of a feature as the degree of the corresponding node in G_m (e.g Figure 8(d),9(d)). This is also comparable with other's results (e.g. [3]). Its suitability for figure-ground discrimination is now under study.

From our analysis and experimentation it is apparent that higher connectivity of the objects in G_u can enhance the grouping quality. Therefore, the selection of cues should consider, in addition to their reliability, also their spatial extent. Another consideration is the use of multi-feature cues, and the cue enhancement possibility (CEP).

Our analysis of the computational complexity is not complete. We still do not have complexity results for the second stage, of finding the maximum likelihood partition. This task is known to be difficult, and for now we use a heuristic algorithm, which gave good results in our experiments. Another research direction is to use our methodology in the context of a different grouping notion, different than partitioning, by which the hypothesized groups are not necessarily disjoint.

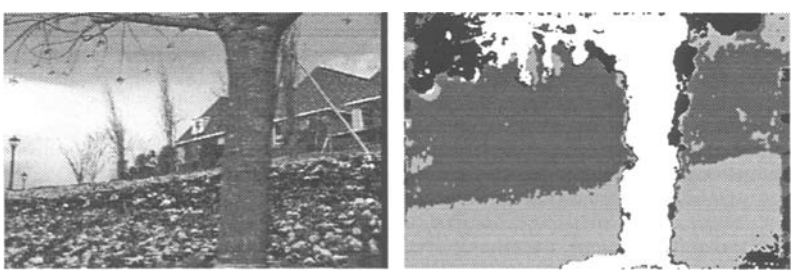


a. Data: A set of points. b. Measured graph G_m c. All 5 detected groups



d. Deleted points (*deletion error*) from all 5 lines. e. Added points (*addition error*) to all 5 lines. f. CEP average time $E\{n\}$.

Fig. 6. Grouping of co-linear points, and its analysis. The data, (a), is associated with five lines, contains 30 points in the vicinity of each of them, and 150 "noise" points. The grouping result is near-optimal, and is close to the predictions for a complete underlying graph. Quantitative results show how the resulting grouping quality depends on the cue reliability (d)(e). Every point represents a complete grouping process and is labeled by the total addition/deletion errors. The average number of trials needed to achieve this enhanced cue reliability by the CEP , $E\{n\}$, is given in (f) near every point, and is compared to the predicted value, given by the labeled curves.



a. One image of a sequence. b. Final image segmentation.

Fig. 7. Image segmentation into regions consistent with the same Affine motion parameters. Grouping is done on the optical-flow image (see Sec. 6). A post-processing stage use the obtained grouping to calculate an Affine motion model for every group, and to classify each pixel to the nearest model (The same post-processing used in [1]. Black pixels were not classified). The underlying graph is a complete graph of about 600 nodes (180,000 arcs), and the runtime is about 5 minutes.

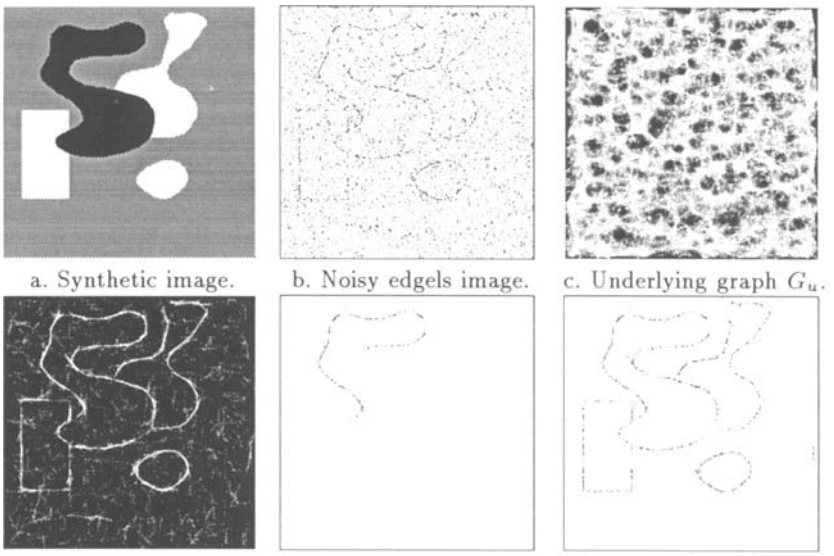


Fig. 8. Grouping of smooth curves in a synthetic image. Edges and gradient where found using image a. The underlying graph, G_u, consists of 5,000 elements (edgels), and 110,000 arcs. The processing time is 3 min on a Super-Spark CPU.

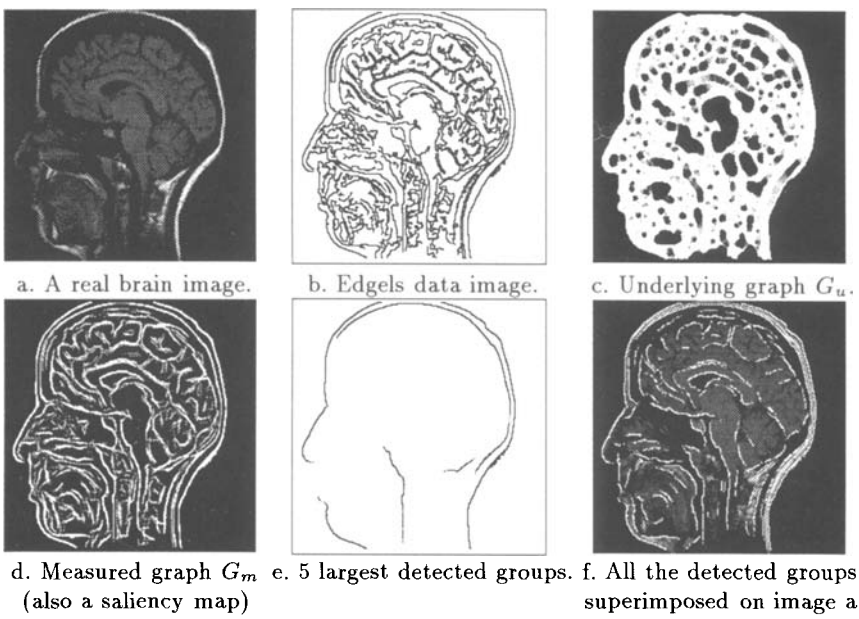


Fig. 9. Grouping of smooth curves in a brain image. The underlying graph, G_u, consists of 10,400 edgels and 230,000 arcs. The processing time is about 10 minutes on a Super-Spark CPU.

Acknowledgments:

We would like to thank John Wang for providing us the Flowers garden optical flow data, and the anonymous referees for their helpful comments. This research was supported by the fund for the promotion of research at the Technion.

References

1. ADELSON, E. H., AND WANG, J. Y. A. Representing moving images with layers. Tech. Rep. 279, M.I.T, May 1994.
2. AMIR, A., AND LINDENBAUM, M. The construction and analysis of a generic grouping algorithm. Tech. Rep. CIS-9418, Technion, Israel, Nov. 1994.
3. GUY, G., AND MEDIONI, G. Perceptual grouping using global saliency-enhancing operators. In *ICPR-92* (1992), vol. I, pp. 99–103.
4. HERAULT, L., AND HORAUD, R. Figure-ground discrimination: A combinatorial optimization approach. *PAMI 15*, 9 (Sep 1993), 899–914.
5. JACOBS, D. W., AND CHENNUBHOTLA, C. Finding structurally consistent motion correspondences. In *ICPR-94, Jerusalem* (1994), pp. 650–653.
6. LOWE, D. G. *Perceptual Organization and Visual Recognition.* Kluwer Academic Pub., 1985.
7. SHA'ASHUA, A., AND ULLMAN, S. Grouping contours by iterated pairing network. *Neural Information Processing Systems (NIPS) 3* (1990).
8. SHAPIRO, L. G., AND HARALICK, R. M. Decomposition of two-dimentional shapes by graph-theoretic clustering. *PAMI 1*, 1 (Jan. 1979), 10–20.
9. VOSSELMAN, G. *Relational Matching*, third ed. Lect. Notes in CS. Springer, 1992.
10. WALD, A. *Sequencial Analysis*, third ed. Wiley Publications in Statistics. 1947(1952).
11. WU, Z., AND LEAHY, R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *PAMI 15*, 11 (Nov 1993), 1001–1113.
12. ZISSERMAN, A., MUNDY, J., FORSYTH, D., AND LIU, J. Class-based grouping in perspective images. In *ICCV-95, MIT* (1995), pp. 183–188.