

Using Segmentation Hierarchies for Matching and Recognition in Computer Vision *

Roland Glantz⁺, Marcello Pelillo⁺, and Walter G. Kropatsch[#]

⁺ Dipartimento di Informatica

Università Ca' Foscari di Venezia

Via Torino 155, 30172 Mestre (VE), Italy

e-mail:{glantz,pelillo}@dsi.unive.it

[#] Pattern Recognition and Image Processing Group 183/2

Institute for Computer Aided Automation, Vienna University of Technology

Favoritenstr. 9, A-1040 Vienna, Austria

e-mail:krw@prip.tuwien.ac.at

Abstract

When matching regions from “similar” images, one typically has the problem of missing counterparts due to local or even global variations of segmentation fineness. Matching segmentation hierarchies, however, not only increases the chances of finding counterparts, but also allows us to exploit the manifold constraints coming from the topological relations between any two regions in a hierarchy. To define the topological relations we represent a plane image \mathcal{I} by a plane attributed graph G and derive a finite topology \mathcal{O} from G . In particular, segmenting \mathcal{I} corresponds to taking a topological minor of G which, in turn, is equivalent to coarsening \mathcal{O} . Moreover, each finite topology involved is a coarsening of the standard topology on \mathbb{R}^2 . Then, we construct a weighted association graph G_A , the nodes of which represent potential matches and the edges of which indicate topological consistency with respect to \mathcal{O} . Specifically, a maximal weight clique of G_A corresponds to a topologically consistent mapping with maximal total similarity. To find “heavy” cliques, we extend a greedy pivoting-based heuristic to the weighted case. Experiments on pairs of stereo images, on a video sequence of a cluttered outdoor scene, and on a sequence of panoramic images demonstrate the effectiveness of our method.

1 Introduction

Vision tasks such as detection, recognition, and tracking usually involve segmentation. Although, in general, the segmentation method must be chosen

*This work is supported by the Austrian Science Foundation (FWF) under grant P14445-MAT and by MURST under grant MM09308497.

according to the application (segmentation itself is an ill-defined problem), the following situation is quite common.

- The segmentation method allows for various levels of fineness.
- For selected levels of fineness the corresponding segmentations form a hierarchy in the sense that a region from a coarser segmentation is the union of regions from a finer segmentation.
- The optimal level of fineness, if any, is a local property.
- One-to-one correspondences between regions and real world objects are rare. Often a region merely contains or is contained in a region corresponding to a real world object.

Hence, it is often a good idea to consider segmentation hierarchies instead of single segmentations determined by a fixed level of fineness.

For recognition tasks, another reason to employ hierarchies is that the objects to be recognized are often also hierarchical. In the following, our perspective is purely two-dimensional, i.e. we do not address problems (like occlusion) coming from the fact that three-dimensional objects in a three-dimensional world are represented by two-dimensional regions of two-dimensional images. As a consequence, the hierarchical relations of the objects and sub-objects must also hold for the corresponding regions. Recognizing the same hierarchical object in two segmentation hierarchies thus means to find a one-to-one hierarchy-preserving mapping between regions of one hierarchy and regions of the other hierarchy. Besides preservation of the hierarchy, it is natural to require that topological relations as the (non-)neighborhood relation and the (non-)enclosure relation are also preserved. For an example see Figure 1.

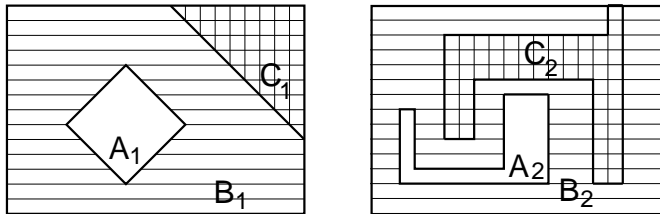


Figure 1: The mapping $A_1 \rightarrow A_2$, $B_1 \rightarrow B_2$, $C_1 \rightarrow C_2$ is topologically consistent.

In this paper we deal with ten topological relations between regions in a hierarchy and we require that a matching between hierarchies preserves these relations. Specifically, when matching region R_1 and region S_1 from the first hierarchy to regions R_2 and S_2 from the second hierarchy, the relation between R_1 and S_1 must be the same as the relation between R_2 and S_2 . We construct an association graph G_A whose nodes represent potential matches and whose edges indicate topological consistency. Let $\sigma(\cdot, \cdot)$ denote a similarity measure

between regions and let the weight of a node in the association graph be given by the σ -value of the corresponding potential match. Then, a matching between two hierarchies that preserves all topological relations, and that has a maximal total σ , corresponds to a maximal weight clique in G_A .

To find "heavy" cliques, we employ an extension of the *pivoting-based heuristic* PBH [9] to the weighted case. PBH follows a multi-start strategy in which each new start consists of a small clique. According to a look-ahead rule that evaluates the conditions for further growth, the cliques are then enlarged iteratively. When matching hierarchies via weighted association graphs, however, we can restrict the multi-start strategy to "heavy" initial cliques. Moreover, it is natural to let each initial clique contain the (potential) match of the two apexes and the the (potential) match of the background regions. Experimentally, we found that a single start suffices and that the formation of the cliques then proceeds top-down with respect to the hierarchical ordering of the regions.

The paper is organized as follows. In Section 2 we present the new framework for segmentation hierarchies. The special case of segmentation in terms of watersheds is addressed in Section 3. The definitions of the topological relations are given in Section 4 and the association graph is defined in Section 5. Here, we also present the extension of PBH to the weighted case. Experimental results are given in Section 6.

2 Graph-based Segmentation

Commonly, the smallest entities to be considered in a segmentation method are pixels. For example, in most methods based on the intuitive idea of watersheds [13] the pixels serve both as elements of the watersheds and as elements of the catchment basins. However, the natural elements to separate (two-dimensional) regions are (one-dimensional) curves. Analogously, the natural elements to separate (one-dimensional) curves are (zero-dimensional) points.

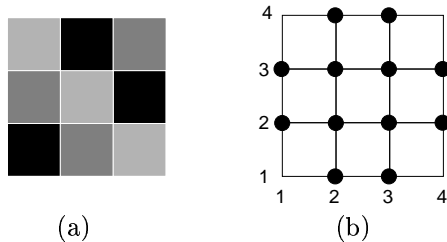


Figure 2: (a) Pixel image. (b) The corresponding plane graph has ten regions, one of which is unbounded.

The plan for formulating a segmentation concept respecting dimensionality is as follows. In Section 2.1 we define a plane image by means of a plane attributed graph G (see Figure 2a) and in Section 2.2 we show how G can be

constructed from a pixel image. In Section 2.3 we derive a cellular complex and a variant of the star topology from G . Finally, in Section 2.4, merges of regions and merges of boundaries are characterized in terms of two-dimensional, respectively one-dimensional, coarsening operations on \mathcal{O} .

2.1 Plane Graphs and Plane Images

Throughout the paper a *graph* $G = (V, E, \iota)$ is given by a finite set V of elements called vertices, a finite set E of elements called edges with $E \cap V = \emptyset$ and an incidence relation ι which associates with each edge $e \in E$ a subset of V with one or two elements. The vertices in $\iota(e)$ are called the end vertices of e .

Note that the definition includes graphs with self-loops (i.e. edges with only one end vertex) and multiple edges (i.e. several edges with identical sets of end vertices). A graph is called *simple*, if it has neither self-loops nor multiple edges.

The sequence $W = (v_0, e_1, v_1, e_2, v_2, \dots, e_k, v_k)$ with $v_i \in V$, $e_i \in E$ is called a *walk* in G , if $v_{i-1}, v_i \in \iota(e_i)$ for all $1 \leq i, \leq k$. The vertices v_0 and v_k are called the *start vertex* and the *end vertex* of W , respectively. A walk is called a *circuit*, if the start vertex and the end vertex form the only pair of identical vertices.

In the following, we restrict ourselves to a special class of plane graphs, i.e. plane graphs defined in terms of *arcs* and *closed polygons*. Arcs and closed polygons are concatenations of finitely many straight line segments in \mathbb{R}^2 . While an arc is homeomorphic to the closed unit interval $[0, 1]$, a closed polygon is homeomorphic to the unit circle in \mathbb{R}^2 [3]. On one hand, the restriction to this special class of plane graphs allows us to adopt the approach towards the definition of plane graphs chosen in [3]. On the other hand, the special class is general enough to deal with pixel-based images, Voronoi- and Delaunay-diagrams. In this paper a *plane* graph is a graph $G = (V, E, \iota)$ such that

- $V \subset \mathbb{R}^2$.
- $e_1 \cap e_2 = \emptyset$ for all $e_1 \neq e_2 \in E$.
- For all $e \in E$ the set $e \cup \iota(e)$ is either an arc or a closed polygon. The set $e \cup \iota(e)$ is a closed polygon, if and only if e is a self-loop.
- For each $e \in E$ such that the set $e \cup \iota(e)$ is an arc and for each homeomorphism $h_e : [0, 1] \mapsto e \cup \iota(e)$ it holds that $h_e\{0, 1\} = \iota(e)$.

In contrast to [3] the end vertices of an edge do not belong to the edge. Thus, G partitions \mathbb{R}^2 into points from V , piecewise linear elements from E , and *regions*, i.e. the connected components of $\mathbb{R}^2 \setminus (V \cup E)$. The unique unbounded region is called *background region* and the set of all regions is denoted by \bar{V} . If g is a mapping from \bar{V} to \mathbb{R}_0^+ , the triple (G, \bar{V}, g) is called *plane image* (with gray values $g(\cdot)$). Note that the “gray values” may also reflect geometric region properties.

2.2 From Pixels to Plane Graphs

For the experiments in this paper we convert classical, pixel-based gray level images into plane images, the bounded regions of which correspond to the pixels. We choose the plane image with the minimal set V of vertices. In particular, the corners of the image need not be represented by vertices.

The corners of the pixels are interpreted as points in \mathbb{R}^2 with integer coordinates (i, j) , $1 \leq i \leq M$, $1 \leq j \leq N$. Points that are corners of at least two pixels form the vertex set V of G (see Figure 2). For any two points $p \neq q \in V$ let

$$L(p, q) := \{p + \lambda(q - p) : 0 < \lambda < 1\} \quad (1)$$

denote the relatively open straight line segment connecting p and q in \mathbb{R}^2 . Furthermore, let

$$\begin{aligned} e_{ul} &:= L((2, 1), (1, 1)) \cup \{(1, 1)\} \cup L((1, 1), (1, 2)) \\ e_{ur} &:= L((1, N - 1), (1, N)) \cup \{(1, N)\} \cup L((1, N), (2, N)) \\ e_{lr} &:= L((M - 1, N), (M, N)) \cup \{(M, N)\} \cup L((M, N), (M, N - 1)) \\ e_{ll} &:= L((M, 2), (M, 1)) \cup \{(M, 1)\} \cup L((M, 1), (M - 1, 1)) \end{aligned}$$

and set $E_c := \{e_{ul}, e_{ur}, e_{lr}, e_{ll}\}$. Defining the 4-neighborhood N_4 of $(i, j) \in V$ by

$$N_4(i, j) = \{(m, n) \in \{1, M\} \times \{1, N\} : |i - m| + |j - n| = 1\}, \quad (2)$$

the edges of G are given by $E := E_c \cup E_4$, where

$$E_4 := \{L(p, q) : p, q \in V, p \in N_4(q)\}. \quad (3)$$

The incidence relations of the edges from E_c are set as follows.

$$\begin{aligned} \iota(e_{ul}) &:= \{(2, 1), (1, 2)\}, \\ \iota(e_{ur}) &:= \{(1, N - 1), (2, N)\}, \\ \iota(e_{lr}) &:= \{(M - 1, N), (M, N - 1)\}, \\ \iota(e_{ll}) &:= \{(M, 2), (M - 1, 1)\}. \end{aligned}$$

The definition of G is completed by setting

$$\iota(L(p, q)) := \{p, q\} \quad \forall L(p, q) \in E_4. \quad (4)$$

For each bounded region \bar{v} of G , i.e. relatively open square of side length one, $g(\bar{v})$ is set to the gray value of the corresponding pixel. The gray value of the background region is set to a special value g_b depending on the application. The segmentation method proposed in the next section does not depend on g_b .

Our concept of graph-based representations easily extends to surfaces with topologies different from that of \mathbb{R}^2 . For example, the cylindrical topology of panoramic images suggests to construct a graph G which is embedded on a cylinder as illustrated in Figure 3. Thus, each pixel corresponds to exactly one region of G on the mantle of a cylinder in \mathbb{R}^3 . However, we also consider the bottom and top disk of the cylinder as regions of G . The latter are called background regions of G .

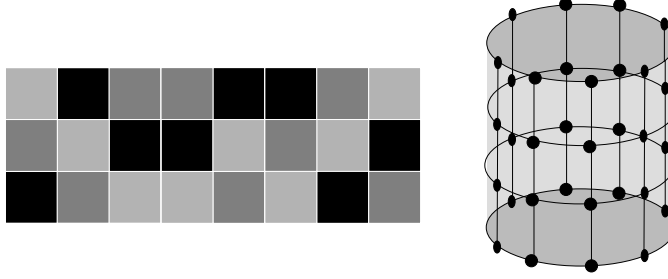


Figure 3: A panoramic image (left) and the corresponding cylindrical graph (right) with 26 regions.

2.3 A Variant of the Star Topology

Intuitively speaking, an edge from E always separates at most two elements from \bar{V} . Formally, the elements separated by e are the unique (and possibly identical) elements $\bar{v}, \bar{w} \in \bar{V}$ such that $\bar{v} \cup e \cup \bar{w}$ is an open subset of \mathbb{R}^2 . This defines a mapping $\bar{\tau}(\cdot)$ from E to the one- or two-element subsets of \bar{V} and thus a graph $\bar{G} = (\bar{V}, E, \bar{\tau})$. The graph \bar{G} is the dual of the graph G [3]. Note that the edge sets of G and \bar{G} are identical. Thus, a (one-dimensional) element of E serves to relate a pair of zero-dimensional elements from V and a pair of two-dimensional elements from \bar{V} . In terms of cellular complexes [6] the elements of V , E , and \bar{V} play the roles of zero-, one-, and two-dimensional cells, respectively. Formally, set

$$\mathcal{C} := V \cup E \cup \bar{V} \quad \text{and} \quad \dim(c) := \begin{cases} 0 & \text{if } c \in V, \\ 1 & \text{if } c \in E, \\ 2 & \text{if } c \in \bar{V}. \end{cases}$$

Note that \mathcal{C} contains finitely many subsets of \mathbb{R}^2 . Here and in the following it is important to distinguish between a collection of sets and the union of the sets from the collection. The union of sets from \mathcal{C} is \mathbb{R}^2 .

Following [1], the *star* of a cell c is a set of cells containing c and the higher-dimensional cells adjacent to c . Formally,

$$\text{star}(c) := \begin{cases} \{c\} & \text{if } c \in \bar{V}, \\ \{c\} \cup \bar{\tau}(c) & \text{if } c \in E, \\ \bigcup \{\text{star}(e) : c \in \bar{\tau}(e)\} & \text{if } c \in V. \end{cases}$$

The collection of open sets defining the so-called *star topology* is given by

$$\mathcal{O}_{\text{star}} := \{C \subset \mathcal{C} : \text{star}(c) \subset C \quad \forall c \in C\}. \quad (5)$$

In contrast to [6], we adhere to the property of the cells being subsets of \mathbb{R}^2 . Thus, the topology \mathcal{O} defined in Equation 8 will be a coarsening of the standard topology on \mathbb{R}^2 .

The union $S(c)$ of all cells contained in $star(c)$, i.e.

$$S(c) := \{p \in \mathbb{R}^2 : p \in c' \text{ for some } c' \in star(c)\} \quad (6)$$

is an open subset of \mathbb{R}^2 , where "open" means "open with respect to the standard topology on \mathbb{R}^2 ". Moreover, there exists a one-to-one correspondence between subsets of \mathcal{C} and the elements of

$$\mathcal{B} := \{B \subset \mathbb{R}^2 : B = \bigcup_{c \in C} c \text{ for some } C \subset \mathcal{C}\}. \quad (7)$$

Finally, \mathcal{O}_{star} corresponds to

$$\mathcal{O} := \{B \in \mathcal{B} : S(c) \subset B \quad \forall c \subset B, c \in \mathcal{C}\}. \quad (8)$$

Note that the elements of \mathcal{O} are subsets of \mathbb{R}^2 , whereas the elements of \mathcal{O}_{star} are subsets of \mathcal{C} . Clearly, every (finite) intersection and every union of elements from \mathcal{O} is contained in \mathcal{O} . Moreover, $\emptyset, \mathbb{R}^2 \in \mathcal{O}$. Consequently, \mathcal{O} is a coarsening of the standard topology on \mathbb{R}^2 .

2.4 Two- and One-dimensional Coarsening

Two-dimensional coarsening: Removing an edge e with $\bar{t}(e) = \{\bar{v}\} \cup \{\bar{w}\}$ (possibly $\bar{v} = \bar{w}$) from G , the regions \bar{v} and \bar{w} are replaced by the new region $S(e) = \bar{v} \cup e \cup \bar{w}$ (see Figure 4). Formally, the collection of *open sets* after the merge is

$$\mathcal{O}' = \{B \in \mathcal{O} : B \cap S(e) \in \{\emptyset, S(e)\}\}. \quad (9)$$

One-dimensional coarsening: If there exists a vertex $u \in V$ such that

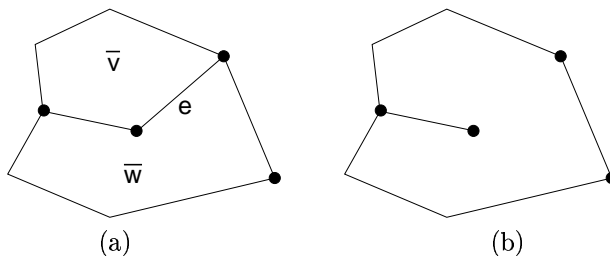


Figure 4: (a) Before two-dimensional coarsening around e . (b) After two-dimensional coarsening. All open sets are (not necessarily proper) supersets of $S(e) = \bar{v} \cup e \cup \bar{w}$, or they are disjoint from $S(e)$.

u is the end vertex of exactly two edges $e_1 \neq e_2$ of G , e_1 and e_2 may be concatenated without modifying the regions of G (see Figure 5). Formally, setting $e^* := e_1 \cup \{u\} \cup e_2$, a new graph $G^* = (V^*, E^*, \iota^*)$ is defined by

- $V^* := V \setminus \{u\}$,

- $E^* := E \cup \{e^*\} \setminus \{e_1, e_2\}$,
- $\iota^*(e') := \iota(e) \quad \forall e' \neq e^*$, and
- $\iota^*(e^*) := \iota(e_1) \cup \iota(e_2) \setminus \{u\}$.

Since $V^* \cup E^* = V \cup E$, it holds that G and G^* have the same regions. The collection \mathcal{O}^* of open sets, as defined by G^* , takes the form

$$\mathcal{O}^* = \{B \in \mathcal{O} : B \cap e^* \in \{\emptyset, e^*\}\}. \quad (10)$$

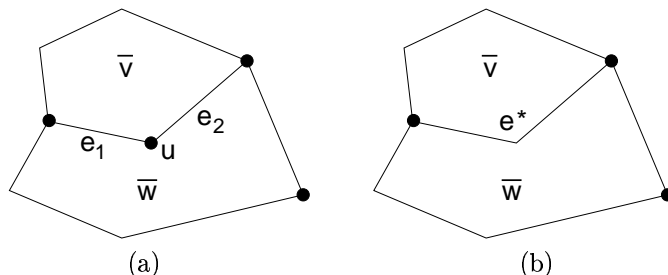


Figure 5: (a) Before one-dimensional coarsening around u . (b) After one-dimensional coarsening. All open sets are supersets of $e^* = e_1 \cup \{u\} \cup e_2$ or they are disjoint from e^* . The regions, in particular \bar{v} and \bar{w} , are not affected. One-dimensional coarsening around u is the reverse operation of subdividing e^* at u .

2.5 Segmentation Hierarchies

Note that one-dimensional coarsening is the reverse operation of subdividing an edge, i.e. placing a new vertex u on an edge e^* and thus subdividing e^* into two edges, each of which has u as an end vertex (see Figure 5). Thus, a sequence of two- and one-dimensional coarsening operations on a plane graph G yields a plane graph G_M such that G_M has a subdivision which is a subgraph of G . In other words, G_M is a topological minor of G [3]. Conversely, any (possibly disconnected) topological minor of G can be obtained by a sequence of one- and two-dimensional coarsening operations. Thus, we may define segmentation hierarchies in terms of topological minors.

Definition 2.1 (Segmentation hierarchy) *Let $\mathcal{I}_{i=0}^m = (G_i, \bar{V}_i, g_i)_{i=0}^m$ be a sequence of plane images such that G_{i-1} is a topological minor of G_i , $1 \leq i \leq m$. Then, the sequence $\mathcal{I}_{i=0}^m$ is called a segmentation hierarchy.*

The removal of e from G corresponds to the contraction of e in \bar{G} and vice versa (see Figure 6). The most general framework for the duality between

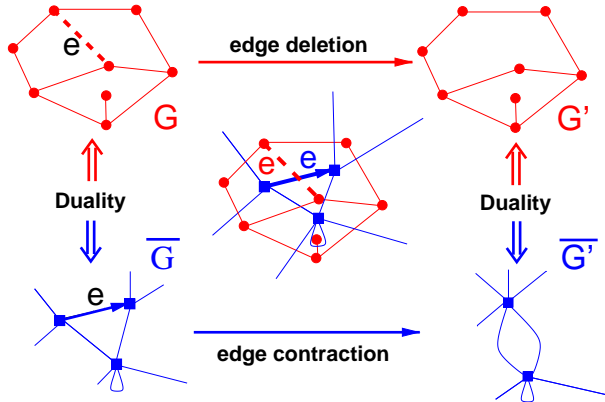


Figure 6: Duality of deletion and contraction. The deletion of e from G corresponds to the contraction of e in \overline{G} .

deletion and contraction is provided by the theory of matroids [12]. Technically, the coarsening operations can be done by the iteratively parallel method *dual graph contraction* [7]. In particular, it suffices to specify the conditions for edge contractions in G and in \overline{G} . For the experiments in this paper we adapted a C^{++} software called *dgc-tool* which was developed by the PRIP-group at the Vienna University of Technology.

3 Morphological Segmentation

Morphological segmentation methods rely on the intuitive idea of flooding a topographic surface in order to find the watersheds and to determine the catchment basins [10]. The topographic surface, in turn, is often derived from the original image by means of an edge filter [11]. In contrast to classical approaches [13], our graph-based approach allows to explicitly represent the one-dimensional region borders and the zero-dimensional region corners. Thus, we avoid the confusion of dimensions in many classical approaches arising from the fact that the watershed lines (a line should have dimension one) are in fact composed of (two-dimensional) pixels.

Let (G, \overline{V}, g) be a plane image. According to the concept outlined in Section 2, the closed 1D-watersheds of (G, \overline{V}, g) will be given by the edges of a topological minor G_M of G . In particular, any edge of G_M will be contained in some circuit.

Our graph-based concept allows us to employ edge filters, whose responses refer to (one-dimensional) edges instead of (two-dimensional) pixels. Let E' denote the set of edges, the star of which does not contain a background region (inner edges). Then, an edge filter is a mapping $f : E' \mapsto \mathbb{R}_0^+$. The following is independent of f , as long as f is a strictly increasing function of the absolute

gray value difference

$$f_1(e) := |g(\bar{v}) - g(\bar{w})|, \text{ where } \iota(e) = \{\bar{v}\} \cup \{\bar{w}\}. \quad (11)$$

The definition of the closed 1D-watersheds is motivated by the aim to guarantee a fixed reduction factor larger than one with respect to the non-background regions. Formally this means that there exists $r > 1$ such that the number of non-background regions in G is at least r times the number of non-background regions in G_M . The reduction factor must be independent of G . Thus, the topological minor obtained after at most $\log_r(|\bar{V}|)$ steps (after each step $f(\cdot)$ is updated) will have but one non-background region. A reduction factor larger than one not only guarantees to arrive at segmentations with a small number of regions, as is indispensable for the methods introduced in the next sections. It also guarantees that the whole segmentation hierarchy can be computed in $O(\log(|\bar{V}|))$ parallel steps [5].

Besides forming a circuit, the edges of a closed 1D-watershed are required to fulfill a local criterion in terms of the responses $f(\cdot)$ (see Figure 7). Recall

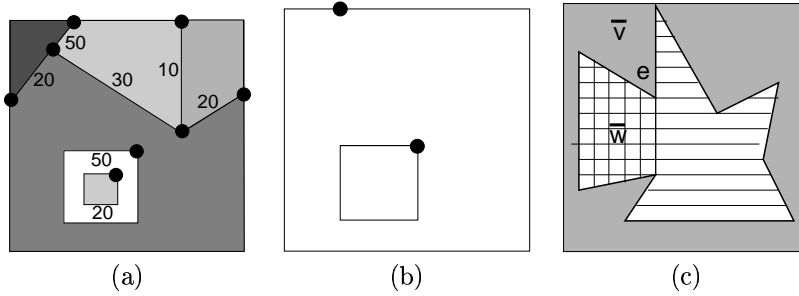


Figure 7: (a) Image with responses of an edge filter indicating absolute differences of mean gray values. (b) The closed 1D-watersheds of (a). (c) Illustration to the definition of enclosure.

that for an edge e we have that $star(e) = \{e\} \cup \{\bar{v}\} \cup \{\bar{w}\}$, where \bar{v} and \bar{w} are (possibly identical) elements of \bar{V} . The set $\delta(\bar{v})$ of edges bounding \bar{v} is defined by

$$\delta(\bar{v}) := \{e \in E : star(e) \cap \{\bar{v}\} = \{\bar{v}\}\}. \quad (12)$$

Note that $\delta(\bar{v})$ may be disconnected (see Figure 7). In the following we will formulate a local criterion for an inner edge *not* to be on a 1D-watershed. A closed 1D-watershed will then be defined as a circuit, the edges of which are maximal concatenations of edges that do *not* fulfill the criterion. In particular, an edge $e \in E'$ with (possibly identical) end vertices \bar{v} and \bar{w} in \bar{G} is *not* on a 1D-watershed, if the response on e is minimal with respect to all responses on edges from $\delta(\bar{v})$ or with respect to all responses on edges from $\delta(\bar{w})$. Formally, the minimal response of f in $\delta(\bar{v})$ is denoted by $min_f(\delta(\bar{v}))$ and the local criterion takes the form

$$f(e) = min_f(\delta(\bar{v})) \quad \vee \quad f(e) = min_f(\delta(\bar{w})). \quad (13)$$

Definition 3.1 (G_M , closed 1D-watershed of (G, \overline{V}, g) w.r.t. f) Let E_f be the set of edges from G that do not fulfill Equation 13 and let G_f be the plane subgraph of G that is induced by E_f . The closed 1D-watersheds of (G, \overline{V}, g) with respect to f are the edges of the unique topological minor $G_M = (V_M, E_M, \iota_M)$ of G such that the following holds.

1. For each edge e of each circuit of G_f there exists an edge $e^* \in E_M$ with $e \subset e^*$.
2. No vertex in V_M has a degree smaller than three.

It is natural to set the gray value g_1 of a region arising from a merge of the regions $\overline{v}_1, \dots, \overline{v}_k$ of G to the size-weighted mean of the gray values $g(\overline{v}_i)$, $1 \leq i \leq k$. Another application of the edge filter f , this time on $(G_M, \overline{V}_M, g_1)$, brings us back to the initial situation. We can now compute a topological minor of G_M , the edges of which are the closed 1D-watersheds of $(G_M, \overline{V}_M, g_1)$ with respect to f , and so on.

When matching segmentation hierarchies we are interested in hierarchies with regions that “grow slowly” as one ascends to higher levels. Indeed, this increases the chances of finding counterparts when the hierarchies are matched. Moreover, we wish to prevent merging via small regions. Specifically, let $\overline{v}_1, \overline{v}_2$, and \overline{v}_3 be three regions of G such that $\overline{v}_1, \overline{v}_3$ are large, \overline{v}_2 is small, and the gray value of \overline{v}_2 lies in between those of \overline{v}_1 and \overline{v}_3 . Then, it may happen that \overline{v}_1 and \overline{v}_3 are merged via \overline{v}_2 .

Both objectives are attained at the same time by introducing *directions* of contractions in \overline{G} and

1. setting the attribute of a directed edge to the product of the f -value of the corresponding undirected edge and the size of the region at the source of e ,
2. contracting an edge only if it points from a smaller to a larger region.
3. never contract two edges with the same source.

The result of this procedure is a segmentation hierarchy as defined at the end of Section 2. For an example see Figure 10.

4 Topological Relations and Consistent Pairs

The plan of the section is as follows. A subset relation, a neighborhood relation, an enclosure relation, and combinations thereof (see Figure 8) are defined for pairs of regions from possibly different topological minors in a segmentation hierarchy. Then, we specify what it means that a pair of regions from one hierarchy is topologically consistent with a pair from another hierarchy.

In the following, let $(G_i, \overline{V}_i, g_i)_{i=0}^m$ be a segmentation hierarchy with $G_i = (V_i, E_i, \iota_i)$ for all i . Furthermore, let $\overline{v} \in \overline{V}_i$ for some i and let $\overline{w} \neq \overline{v}$, $\overline{w} \in \overline{V}_j$ for some j . Besides $\overline{v} \subset \overline{w}$ or $\overline{v} \supset \overline{w}$ the regions \overline{v} and \overline{w} potentially fulfill the relations defined below.

- The regions \bar{v} and \bar{w} are said to be neighbors: $\bar{v} \sim \bar{w}$, if there exists an edge $e_0 \in E_0$ such that $S(e_0) \cap \bar{v} \neq \emptyset$ and $S(e_0) \cap \bar{w} \neq \emptyset$.
- The region \bar{v} is said to enclose \bar{w} : $\bar{v} \triangleright \bar{w}$, if there exists a closed polygon $e \in E_i$ for some i such that \bar{w} is contained in the interior of e and \bar{v} is contained in the exterior of e (Figure 7d).
- The region \bar{v} is said to be apart from the region \bar{w} : $\bar{v} \mid \bar{w}$, if

$$(\bar{v} \not\subset \bar{w}) \wedge (\bar{v} \not\supset \bar{w}) \wedge (\bar{v} \not\sim \bar{w}) \wedge (\bar{v} \not\triangleleft \bar{w}). \quad (14)$$

From $\bar{v} \neq \bar{w}$ it follows that the five relations $v \subset w$, $v \supset w$, $v \triangleright w$, $v \triangleleft w$, and $v \mid w$ exclude each other and cover all possibilities. Moreover, each of the five relations may occur together with the neighborhood relation and together with the complement of the neighborhood relation. Thus, there are ten combinations of the (non-) neighborhood relation with the other five relations and these ten relations cover all possibilities (see Figure 8). In the following, a combination of the (non-) neighborhood relation with one of the other five relations is denoted by the symbol of the (non-) neighborhood relation followed by the symbol for the other relation. Examples of the topological relations are given in Table 1 and Table 2.

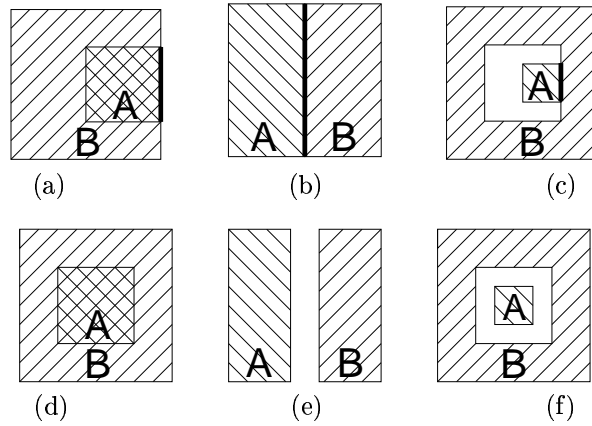


Figure 8: The 10 topological relations. The thick lines indicate the edges defining the neighborhood relations. (a) $A \sim \subset B$. (b) $A \sim \mid B$. (c) $A \sim \triangleleft B$, $B \sim \triangleright A$. (d) $A \not\subset B$, $B \not\supset A$. (e) $A \not\sim \mid B$. (f) $A \not\triangleleft B$, $B \not\triangleright A$.

Definition 4.1 (Topological consistency)

Let $(G^1_i, \bar{V}_i^1, g^1_i)_{i=0}^{k_1}$ and $(G^2_i, \bar{V}_i^2, g^2_i)_{i=0}^{k_2}$ be two segmentation hierarchies. Furthermore, let $\bar{v}^1 \in \bar{V}_i^1$ for some i , $\bar{w}^1 \in \bar{V}_j^1$ for some j , $\bar{v}^2 \in \bar{V}_k^2$ for some k , and $\bar{w}^2 \in \bar{V}_l^2$ for some l . The pair (\bar{v}^1, \bar{v}^2) is said to be topologically consistent

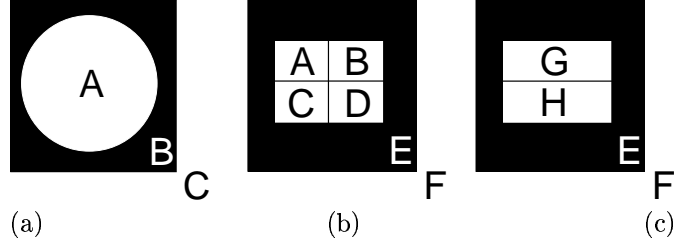


Figure 9: (a) Merging A and B first results in the topological relations of Table 1. All merging operations between (b) and (c) (in two single or one parallel step) yield the same topological relations. See Table 2.

Table 1: Topological relations from the hierarchy on Figure 9a. Merging A with B results in D .

	A	B	C	D	\mathbb{R}^2
A	=	$\sim\triangleleft$	$\not\triangleleft$	$\not\subset$	$\not\subset$
B	$\sim\triangleright$	=	$\sim\triangleleft$	$\sim\subset$	$\not\subset$
C	$\not\triangleright$	$\sim\triangleright$	=	$\sim\triangleright$	$\not\subset$
D	$\not\supset$	$\sim\supset$	$\sim\triangleleft$	=	$\not\subset$
\mathbb{R}^2	$\not\supset$	$\not\supset$	$\not\supset$	$\not\supset$	=

with the pair (\bar{w}^1, \bar{w}^2) , if the topological relation between \bar{v}^1 and \bar{w}^1 is the same as the topological relation between \bar{v}^2 and \bar{w}^2 .

5 Consistent Mappings and Association Graphs

When matching regions from two segmentation hierarchies we require that the topological relations of the mapped regions are preserved. In other words, we are looking for consistent mappings defined as follows.

Table 2: Topological relations from the hierarchy on Figures 9b.

	A	B	C	D	E	F	G	H
A	=	\sim	\sim	$\not\sim$	$\sim\triangleleft$	$\not\triangleleft$	$\sim\subset$	\sim
B	\sim	=	$\not\sim$	\sim	$\sim\triangleleft$	$\not\triangleleft$	$\sim\subset$	\sim
C	\sim	$\not\sim$	=	\sim	$\sim\triangleleft$	$\not\triangleleft$	\sim	$\sim\subset$
D	$\not\sim$	\sim	\sim	=	$\sim\triangleleft$	$\not\triangleleft$	\sim	$\sim\subset$
E	$\sim\triangleright$	$\sim\triangleright$	$\sim\triangleright$	$\sim\triangleright$	=	$\sim\triangleleft$	$\sim\triangleright$	$\sim\triangleright$
F	$\not\triangleright$	$\not\triangleright$	$\not\triangleright$	$\not\triangleright$	$\sim\triangleright$	=	$\not\triangleright$	$\not\triangleright$
G	$\sim\supset$	$\sim\supset$	\sim	\sim	$\sim\triangleleft$	$\not\triangleleft$	=	\sim
H	\sim	\sim	$\sim\supset$	$\sim\supset$	$\sim\triangleleft$	$\not\triangleleft$	\sim	=

Definition 5.1 ($(R^1, R^2, \text{Consistent})\mathcal{H}_1 - \mathcal{H}_2$ -Mapping M on U^1) Let $\mathcal{H}_1 := \mathcal{I}_{i=0}^{k_1} = (G^1_i, \bar{V}^1_i, g^1_i)_{i=0}^{k_1}$ and $\mathcal{H}_2 := \mathcal{J}_{i=0}^{k_2} = (G^2_i, \bar{V}^2_i, g^2_i)_{i=0}^{k_2}$ be two segmentation hierarchies, and let

$$R^j := \{\bar{V}^j_i : 0 \leq i \leq k_j\}. \quad (15)$$

A $\mathcal{H}_1 - \mathcal{H}_2$ -mapping M on $U^1 \subset R^1$ is a one-to-one mapping M from U^1 to a subset of R^2 . It is called *consistent*, if for any $\bar{v}^1, \bar{w}^1 \in U^1$ the pair (\bar{v}^1, \bar{w}^1) is topologically consistent with the pair $(M(\bar{v}^1), M(\bar{w}^1))$.

Let $\sigma : R^1 \times R^2 \leftrightarrow \mathbb{R}^+$ be a similarity measure between regions. It extends to a similarity measure Σ of a consistent mapping M on U^1 by

$$\Sigma(M) = \sum_{\bar{v}^1 \in U^1} \sigma(\bar{v}^1, M(\bar{v}^1)). \quad (16)$$

A consistent $\mathcal{H}_1 - \mathcal{H}_2$ mapping M on U^1 is said to be a *maximal* similarity mapping, if there exists no consistent $\mathcal{H}_1 - \mathcal{H}_2$ mapping on a proper superset of U^1 . M is said to be a *maximum* similarity mapping, if no $\mathcal{H}_1 - \mathcal{H}_2$ -mapping has a higher $\Sigma(M)$ -value than that of M .

Let G be a simple graph (see Section 2.1) with vertex set V and edge set E . A subset V_c of V is called a *clique* of G , if for all $u \neq v \in V_c$ there exists $e \in E$ with $\iota(e) = \{u, v\}$. Let the vertices of G be equipped with weights $\omega : V \rightarrow \mathbb{R}^+$. Then, a clique V_c is said to have *maximal weight*, if no proper superset of V_c is a clique of G . A clique V_c of G is said to have *maximum weight*, if no clique of G has a higher total weight than that of V_c .

In the following we will define a simple graph, the cliques of which correspond to consistent mappings.

Definition 5.2 (**Topological association graph G_A of \mathcal{H}_1 and \mathcal{H}_2**) The topological association graph of the segmentation hierarchies $\mathcal{H}_1 := (G^1_i, \bar{V}^1_i, g^1_i)_{i=0}^{k_1}$ and $\mathcal{H}_2 := (G^2_i, \bar{V}^2_i, g^2_i)_{i=0}^{k_2}$ is the graph $G_A = (V_A, E_A, \iota_A)$ defined by

- $V_A = (\bigcup_{i=1}^{k_1} \bar{V}^1_i) \times (\bigcup_{i=1}^{k_2} \bar{V}^2_i)$,
- $E_A = \{\{v, w\} : v \neq w \in V_A, v \text{ is topologically consistent with } w, \text{ and}$
- $\iota_A(e) = e \quad \forall e \in E_A$.

Theorem 5.3 Let $\sigma(\cdot, \cdot)$ be a similarity measure between regions, let M be a $\mathcal{H}_1 - \mathcal{H}_2$ -mapping on U^1 , and let $G_A = (V_A, E_A, \iota_A)$ be the topological association graph of \mathcal{H}_1 and \mathcal{H}_2 . Furthermore, let the weight of a vertex $(v^1, v^2) \in V_A$ be given by $\sigma(v^1, v^2)$. Then, M is consistent, if and only if the set $V_c := \{(v^1, M(v^1)) : v^1 \in U^1\}$ is a clique of G_A . In this case, M is a maximal (maximum) similarity measure, if and only if V_c is a maximal (maximum) clique of G_A .

The above theorem follows directly from the construction of G_A .

To find "heavy" cliques in G_A , we extend the pivoting-based heuristic (PBH) in [9] to the weighted case. Originally, PBH was used to solve a linear complementarity formulation [2] of a standard quadratic program which, in turn, is equivalent to the maximum clique problem.

We chose PBH, because it gave good results on related graph matching problems [4]. Moreover, it is particularly suited to hierarchical matching, since there is a natural starting point for the clique enlargement (see below).

PBH follows a multi-start strategy in which each new start consists of a small clique. According to a look-ahead rule that evaluates the conditions for further growth, the cliques are then enlarged iteratively. For the unweighted case Locatelli *et al.*[8] gave the following combinatorial interpretation of the look-ahead rule in [9].

For enlarging the current clique, always take a candidate whose degree is maximal in the subgraph induced by all candidates.

To arrive at high values of $\Sigma(M)$, we extend this rule. If $G_A = (V_A, E_A, \iota_A)$ is the topological association graph, $C \subset V$ denotes the set of candidates to enlarge the current clique, $N_C(c)$ denotes the neighborhood of a vertex c in the subgraph of G_A induced by C , i.e.

$$N_C(c) = C \cap \bigcup (\iota(e) : c \in \iota(e)), \quad (17)$$

then the extended look-ahead rule says the following.

For enlarging the current clique, always take a candidate c such that

$$\sum_{c' \in N_C(c)} \sigma(c') \text{ is maximal.} \quad (18)$$

The hierarchies suggest a good starting point for the formation of the clique, i.e. the potential matches of the apexes and the background regions. Note that in case of panoramic images we have two pairs of background regions. Formally, the initial 2-element clique in case of plane images takes the form

$$C_0 := \{(\bar{a}_1, \bar{a}_2), (\bar{b}_1, \bar{b}_2)\}, \quad (19)$$

where the a_i [b_i] stand for the apexes [background regions] of the i -th hierarchy. Thus, the original multi-start strategy of PBH turns into a single start strategy.

6 Results

The aim of the experiments is to roughly define the field in which the requirement of topological consistency in conjunction with ad-hoc similarity measures is sufficient to yield plausible matches. In Section 6.1 we apply our method to

two pairs of stereo images. While the experiment on the first pair shows that our method is capable of finding plausible matches without using the epipolar constraint, the second experiment demonstrates the limits of our approach. In Section 6.2 we use our method to match frames from a video sequence of a cluttered outdoor scene. Again, two experiments are performed, one on two close frames and one on two distant frames. These experiments demonstrate that, as expected, our method is strong when the topological relations between the regions are preserved, and that it fails wherever occlusion modifies the topological relations. Finally, the experiments in Section 6.3 show that our method is robust even to severe distortions between pairs of panoramic images.

All images were scaled to approximately 15.000 pixels and the generation of the hierarchy was always calculated as specified in Section 3. Moreover, to reduce memory requirements, we restricted each hierarchy to its upper 14 levels. The weight of a node in the association graph corresponding to the potential match (R_1, R_2) was always set to

$$\sigma(R_1, R_2) = ((1 - dx)(1 - dy)(1 - dg)(1 - ds))^\alpha, \quad (20)$$

where dx and dy stand for the normalized absolute deviation of the barycenters in x and in y , dg stands for the normalized absolute difference of the mean gray values,

$$ds = \frac{\min(\text{area}(R_1), \text{area}(R_2))}{\max(\text{area}(R_1), \text{area}(R_2))}, \quad (21)$$

and $\alpha = 0.3$ is an empirical value. Typically, the topological association graph is very dense. To save memory, we neglect an edge e between potential matches (R_1, R_2) and (S_1, S_2) , if the relative location of R_1 with respect to S_1 deviates significantly from the relative location of R_2 with respect to S_2 . Formally, e is neglected, if

$$\|b(R_1) - b(S_1) - (b(R_2) - b(S_2))\|_2 > DEV, \quad (22)$$

where $b(R) \in \mathbb{R}^2$ stands for the barycenter of region R and DEV is a threshold (set to 22 in the experiments on the highly distorted panoramic images, and to 7, otherwise).

To evaluate the advantages of fully hierarchical matching over flat matching between the base levels only, we performed both kinds of experiments. In all experiments the total similarity from fully hierarchical matching was about twice of that from flat matching.

6.1 Results on Stereo Images

For the first experiment we chose the stereo pair “arch” (Figure 10a and 11a) from the *CMU/VASC Image Database* at “<http://vasc.ri.cmu.edu/idb/html/stereo/>”, because the scene has a simple structure. Note, however, that the arch is segmented differently in the left and the right image. In particular, the left pillar of the arch in the right image does not fuse with the rest of the arch (Figure 11). Nevertheless (part of) all salient arc components could be matched

(Figure 12). The second experiment was performed on the stereo pair “cart-alt” (Figure 13). The images stem from the *CMU/VASC Image Database* at “<http://vasc.ri.cmu.edu/idb/html/cart-alt/>”. The result of the matching was rather poor, i.e. the matched regions are isolated, small, and corresponding regions often have a low similarity. Comparing the levels of the two hierarchies (see for example the bottom row of Figure 13) the problem seems to be that there are indeed few corresponding regions in the two hierarchies.

6.2 Results on a Video Sequence

For the experiments in this section we extracted the second, the fifth, and the tenth frame from a translating video sequence with a stabilized background (see Figures 14, 15 and 17). It is named “gard_7_9_A” and can be found at “www.lans.ece.utexas.edu/~strehl/res2/gard.html”. In [14] it is explained how the stabilized background was generated. The reasons for our choice were the following.

- Due to the stabilized background, a pair of corresponding regions not affected by the moving foreground, i.e. by the tree, must have similar locations. Thus, we can check whether the numerous hierarchical and topological relations between the regions are compatible with the location constraint.
- Visually, the regions of the flower garden are shaped irregularly and the boundaries are rather fuzzy. Hence, it is not surprising that the frames 2 and 5 are segmented differently. For example, compare Figure 14f with Figure 15f. We want to see how our method copes with this difficult situation, i.e. how it selects regions from various levels of the hierarchies to compensate for the differences between corresponding levels.
- The moving foreground disturbs the hierarchical and topological relations of its neighbors. Hence, we expect that the matching of the regions “near” the moving foreground is worse than for the other regions. If so, this would be another indication that the matches found by our method correspond to relations in the real world.

The correspondences from matching the second with the fifth frame are depicted in Figure 16. The manifold neighborhood relations between matched regions and the good agreement of the mean gray values make the result a consistent and reliable one. Note also that the poorly matched central part of the images coincides with the part influenced by the moving foreground, i.e. the tree. Indeed, since our method relies on corresponding topological relations, we cannot expect good matches wherever occlusion occurs.

Matching the distant second and tenth frame (Figure 17) the neighborhood relations of the matched regions are still numerous. However, the shape and the size of corresponding regions vary considerably. Since the agreement between the mean gray values of corresponding regions is still good, we assume that the

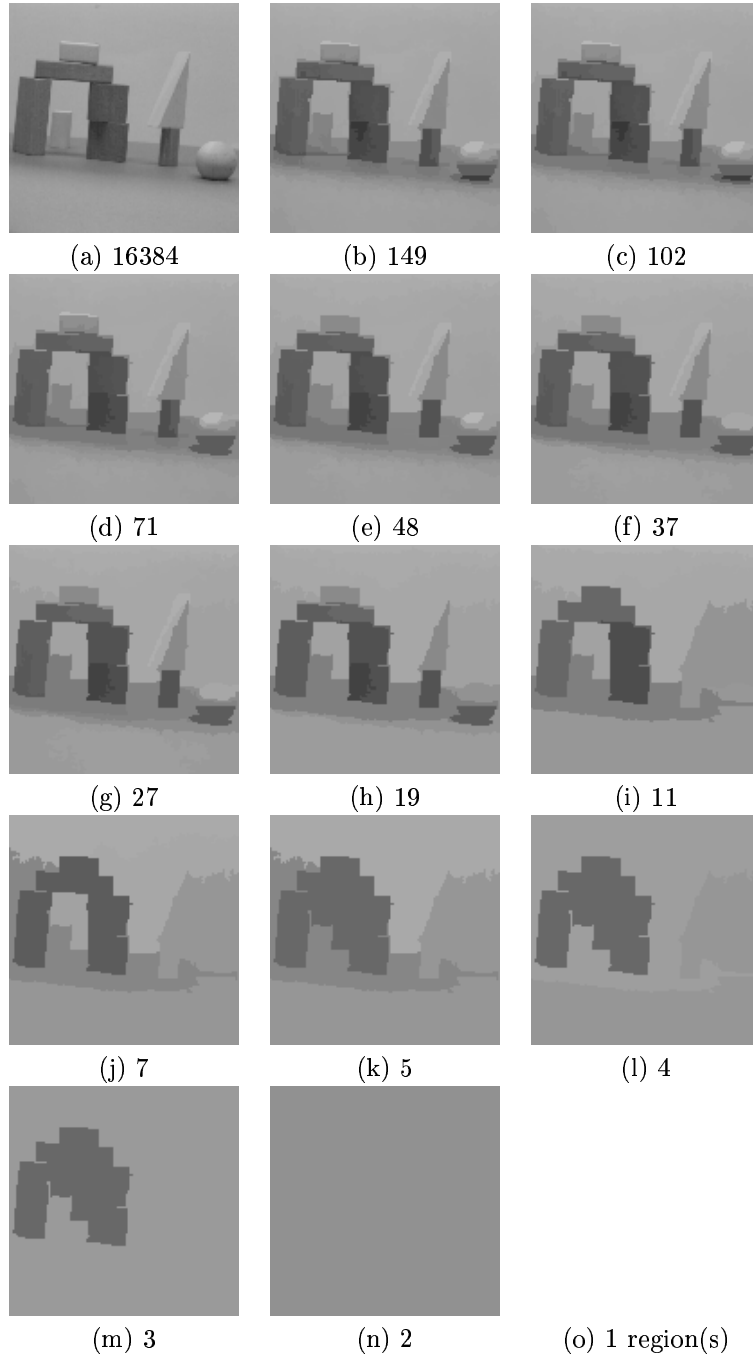


Figure 10: Segmentation hierarchy on the left image of the arch pair. (a) Original image. (b-o) Levels 14, 13, ..., 1.

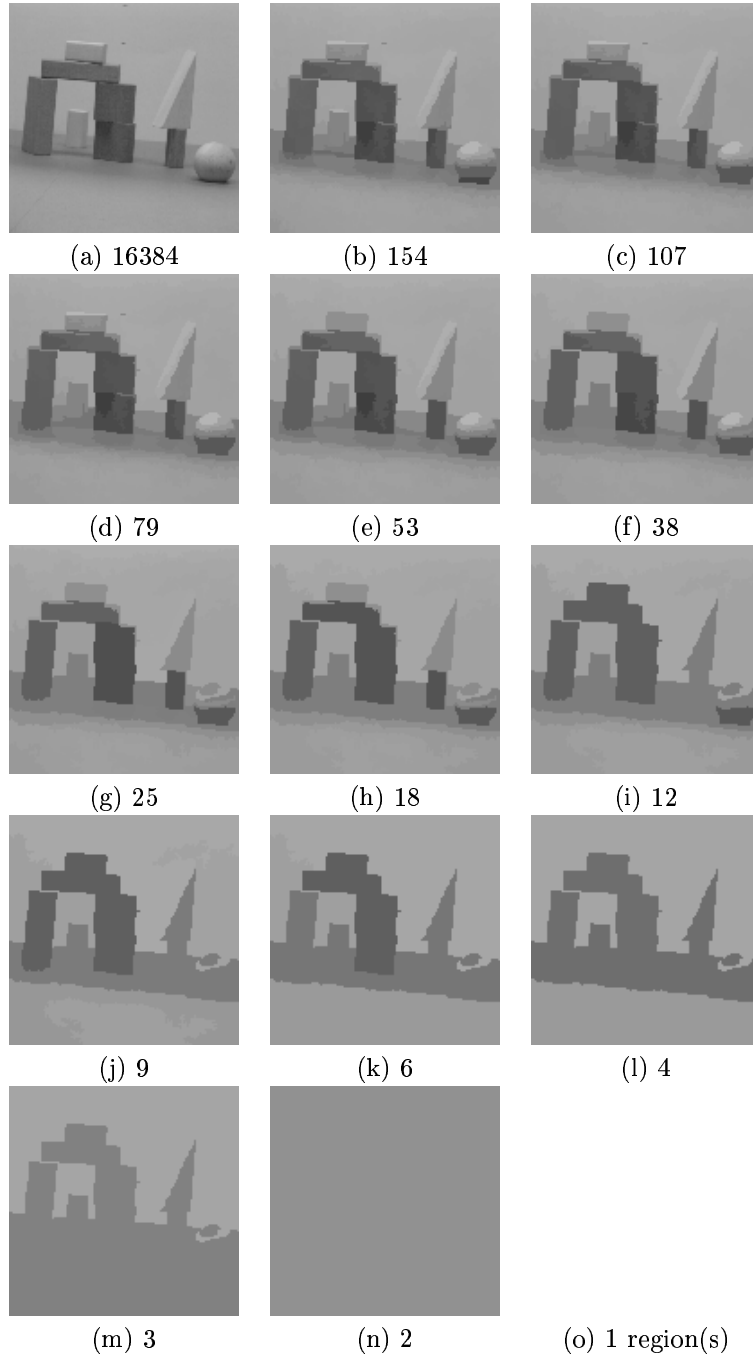


Figure 11: Segmentation hierarchy on the right image of the arch pair. (a) Original image. (b-o) Levels 14, 13, ..., 1.

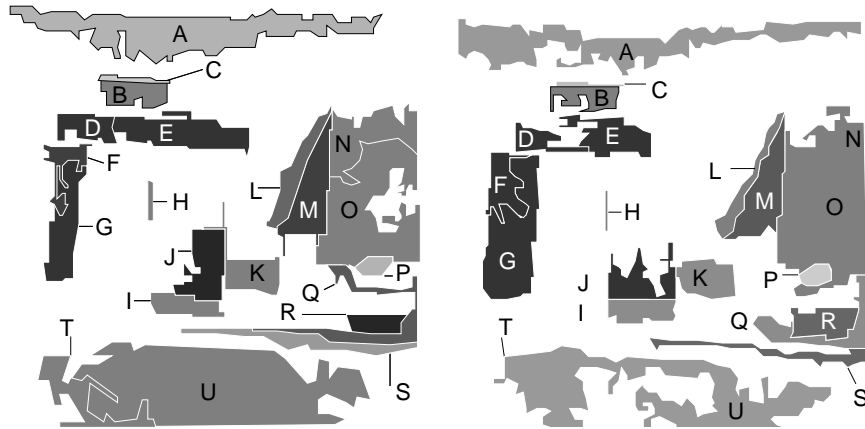


Figure 12: Matching the segmentation hierarchy shown in Figures 10b-o against the segmentation hierarchy shown in Figures 11b-o. Correspondences with respect to the fully hierarchical match are given by the letters.

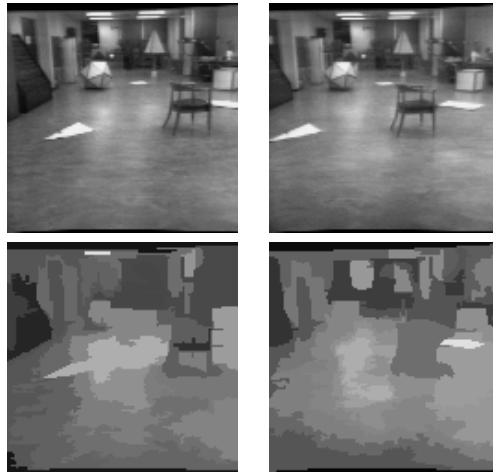


Figure 13: Problematic stereo pair “cart-alt”. The left [right] column refers to the left [right] image. The upper [lower] row depicts segmentations with 28 regions each.

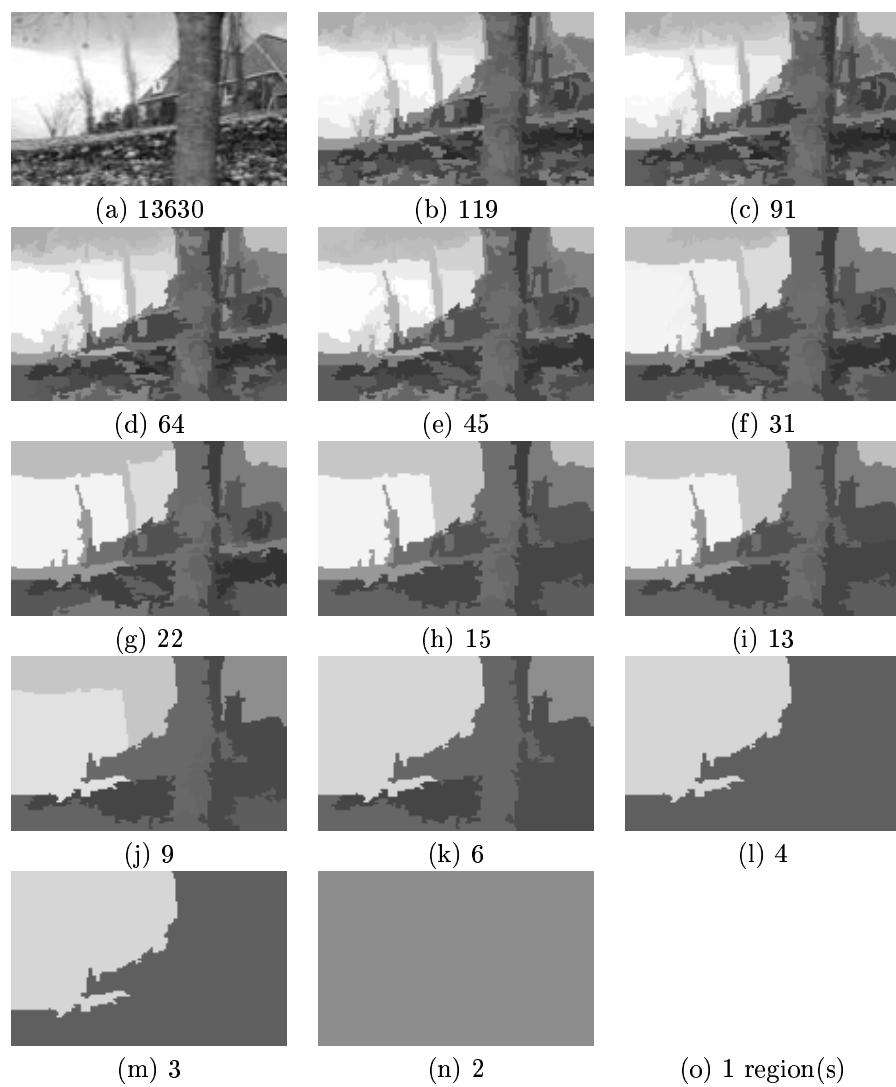


Figure 14: Segmentation hierarchy on the second frame of the `gard_7_9_A` sequence. (a) Original image. (b-o) Levels 14, 13, ..., 1.

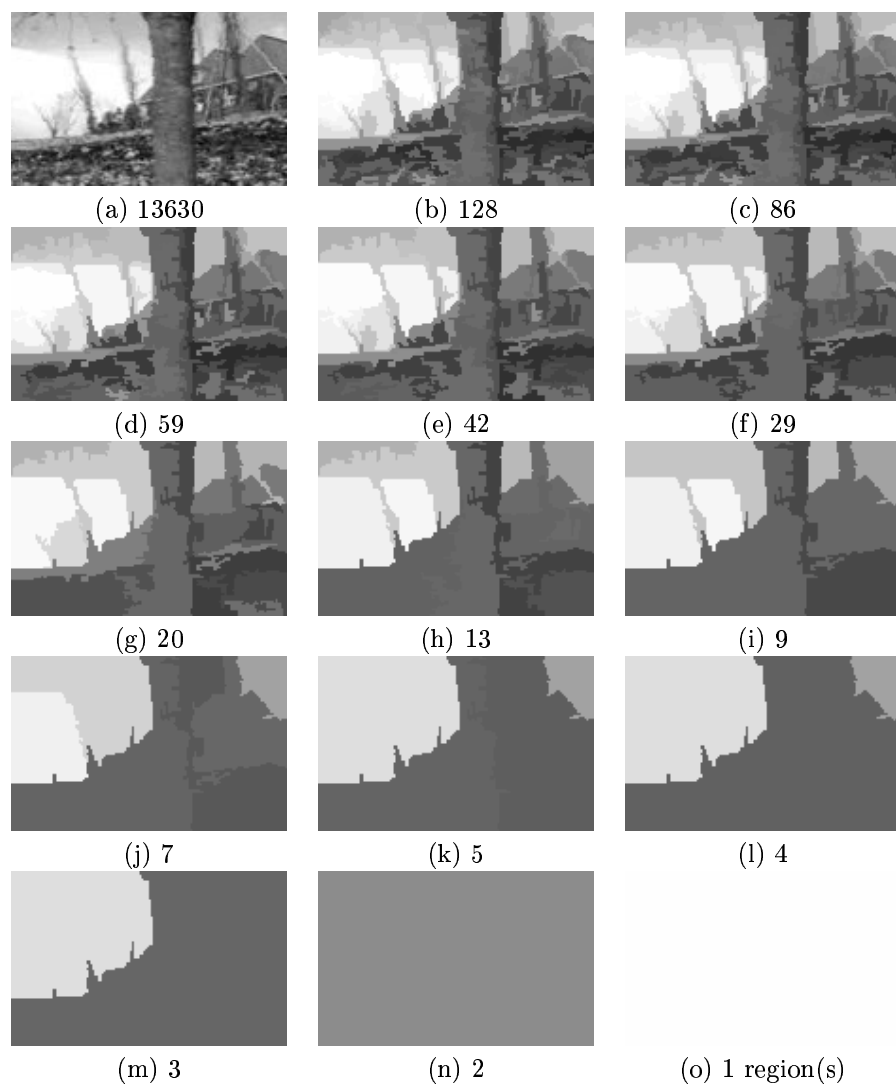


Figure 15: Segmentation hierarchy on the 5th frame of the gard_7_9_A sequence. (a) Original image. (b-o) Levels 14, 13, ..., 1.

matches are mostly correct and the variations of the shape and the size are due to “real” distortions.

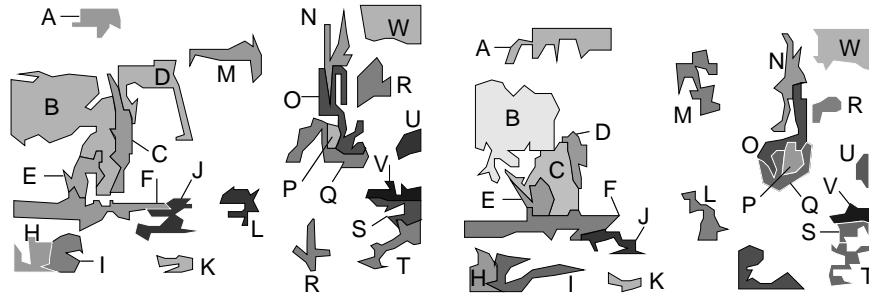


Figure 16: Matching the segmentation hierarchy shown in Figures 14b-o against the segmentation hierarchy shown in Figures 15b-o. Correspondences with respect to the fully hierarchical match are given by the letters

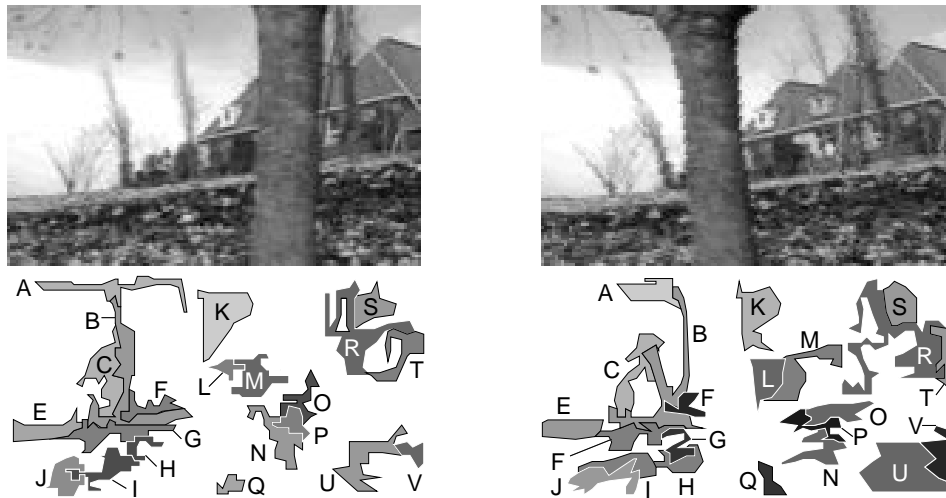


Figure 17: Matching the two distant frames 2 and 10 from the gard.7.9.A sequence. The left [right] column refers to frame 2 [10]. The upper and the lower row shows the original images and the matched regions (fully hierarchical match), respectively.

6.3 Results on Panoramic Images

For the experiments in this section we used a sequence of panoramic images provided by the *Cognitive Vision Group* at the *Computer Vision Laboratory*, University of Ljubljana and the *Center for Machine Perception*, CTU, Prague (<http://lrv.fri.uni-lj.si/matjazj/backyard/testings/>). The sequence simulates a path of a mobile robot through a lab. We focussed on the close pair *cmppath.23*, *cmppath.25* and the more distant pair *cmppath.25*, *cmppath.30*. (Figures 18 19 20).

Clearly, the matching results of the close pair are better, since the matched regions (in both images) cover more area, exhibit more neighborhood relations, and corresponding regions agree well with respect to size, shape and mean gray value. However, also many correspondences between the distant pair seem to agree with correspondences in the real world.

7 Conclusions and Outlook

We showed how to build a weighted association graph, the maximal weight cliques of which correspond to topologically consistent one-to-one mappings with maximal total region similarity between two segmentation hierarchies. Topological relations between any two regions from the same hierarchy were defined by means of representing the hierarchy by a stack of topological minors. In particular, taking a topological minor corresponds to coarsening a variant of the star topology. We applied the new method to matching stereo images, matching frames from a video sequence of a cluttered outdoor scene, and matching frames from a sequence of panoramic images. In all cases, requiring topological consistency over the whole hierarchy and using ad-hoc similarity measures was sufficient to arrive at plausible matches. Moreover, using segmentation hierarchies instead of single segmentations turned out to improve the results considerably. Thus, our concept to formulate and exploit topological consistency in hierarchies is widely applicable. Furthermore, by using weighted association graphs, our concept allows to incorporate further constraints as the epipolar constraint or prior knowledge about the kind of distortions.

References

- [1] E. Ahronovitz, J.-P. Aubert, and C. Fiorio. The star-topology: a topology for image analysis. In *Cinquième Colloque DGCI*, pages 107–116. LLAIC1, Université d’Auvergne, ISBN 2-87663-040-0, September 1995.
- [2] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Accademic Press, Boston, MA, 1992.
- [3] R. Diestel. *Graph Theory*. Springer, New York, 1997.

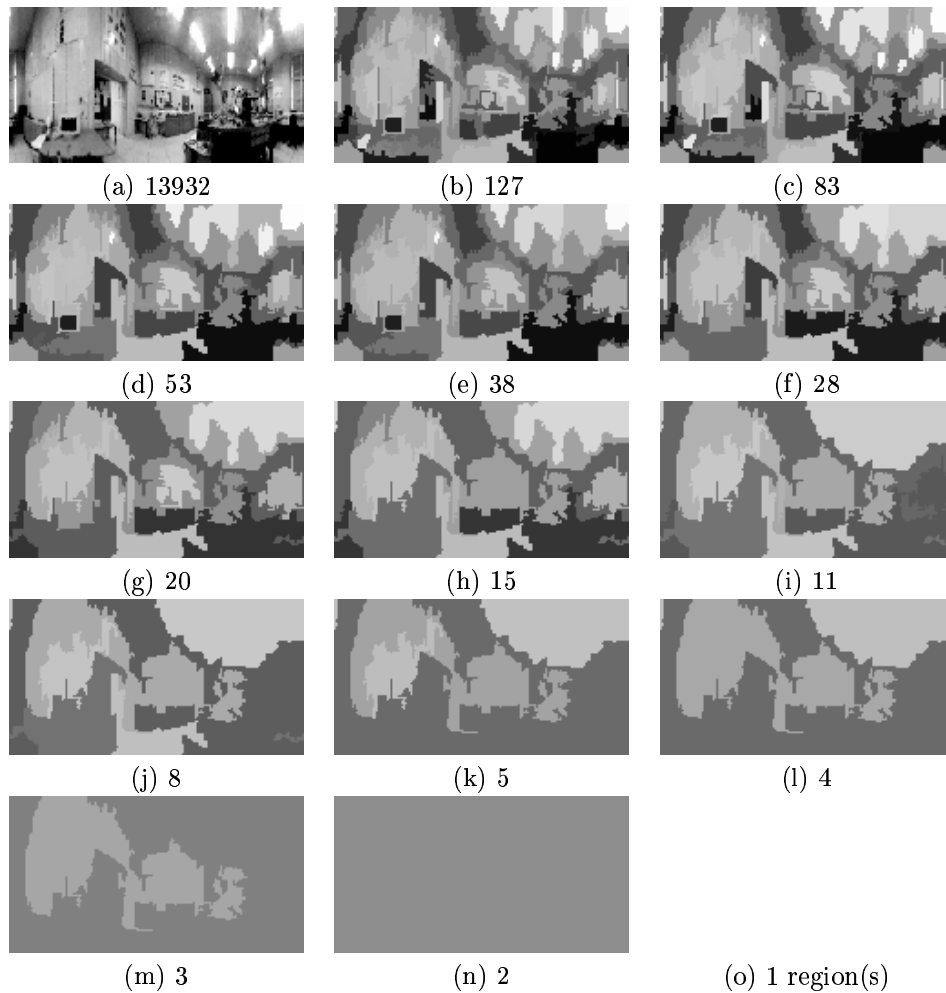


Figure 18: Segmentation hierarchy on *cmppath.23*. (a) Original image. (b-o) Levels 14, 13, ..., 1.

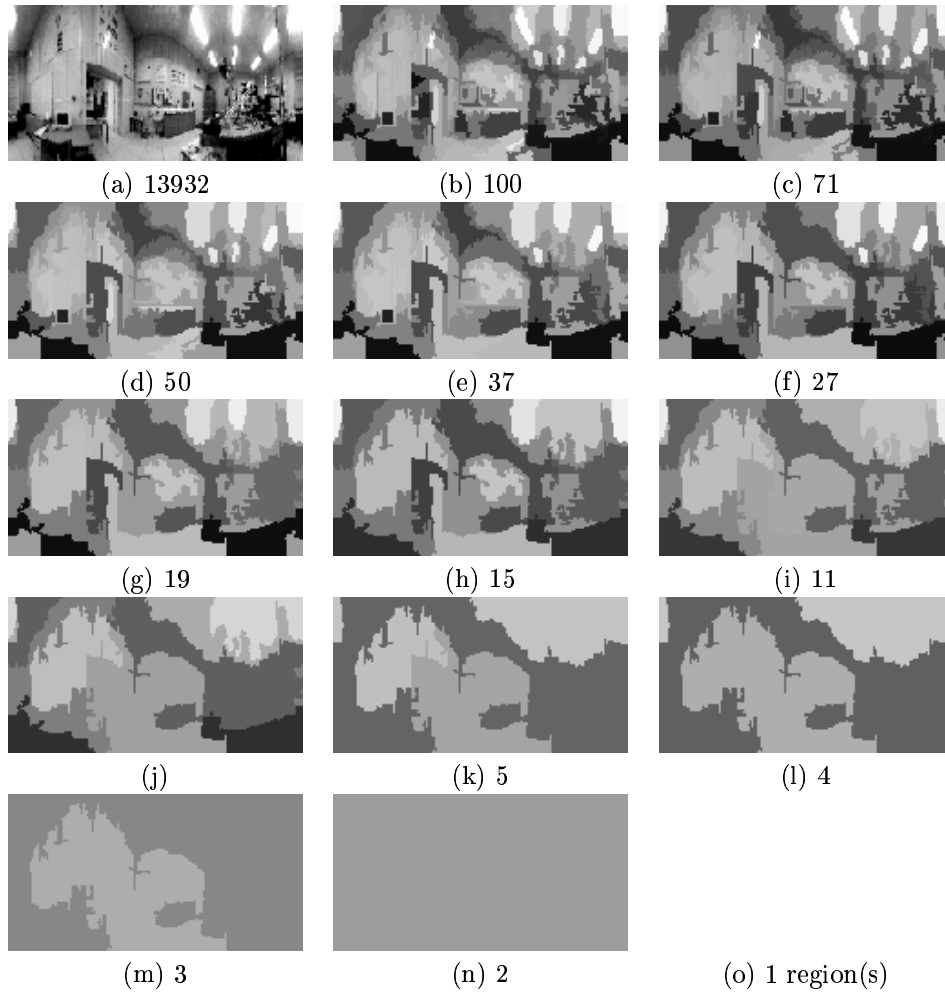


Figure 19: Segmentation hierarchy on *cmppath.25*. (a) Original image. (b-o) Levels 14, 13, ..., 1.

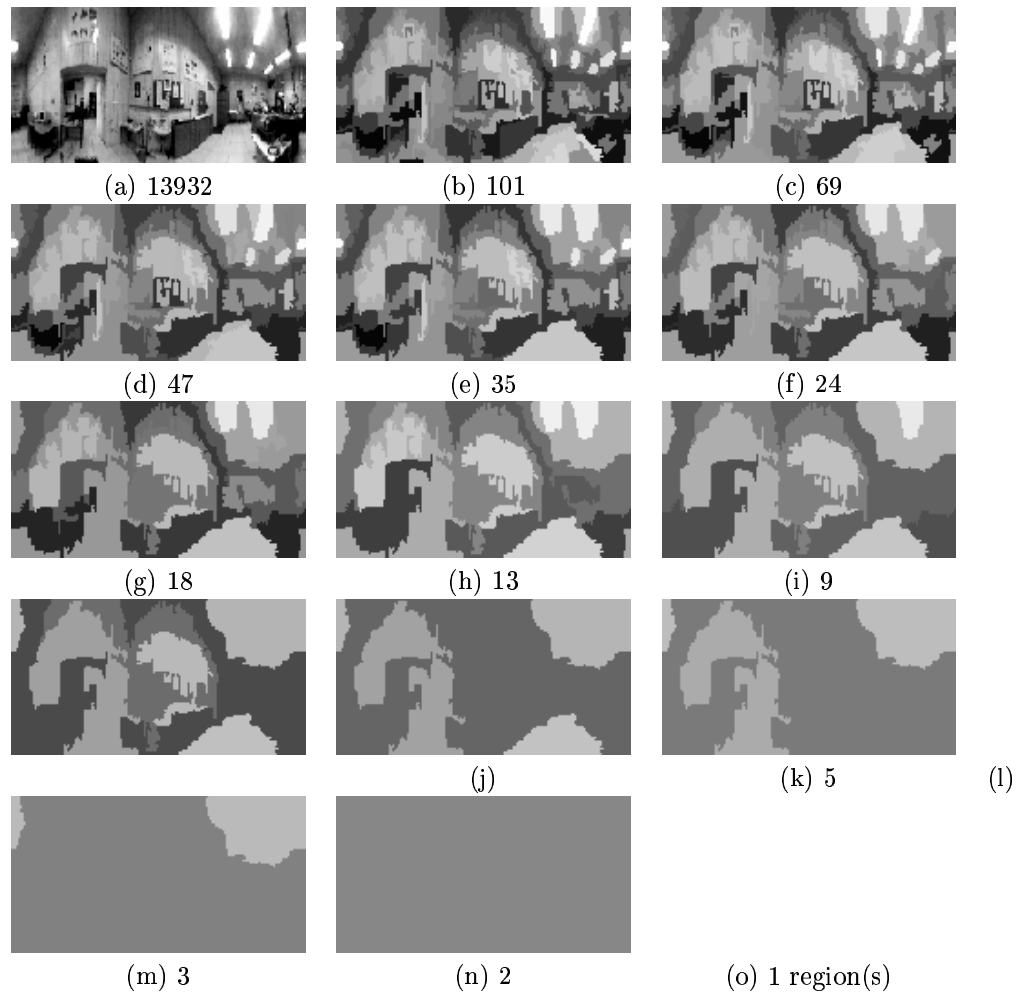


Figure 20: Segmentation hierarchy on *cmppath.30*. (a) Original image. (b-o) Levels 14, 13, ..., 1.

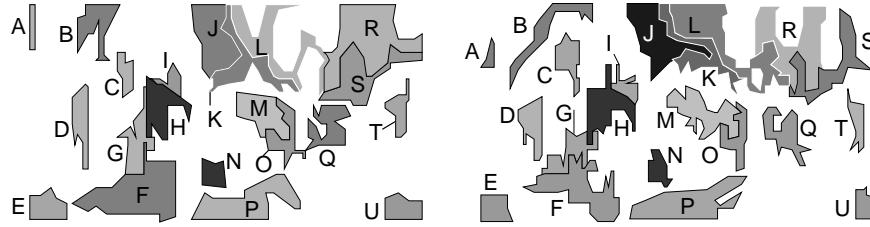


Figure 21: Matching the segmentation hierarchy shown in Figures 18b-o against the segmentation hierarchy shown in Figures 19b-o. Correspondences with respect to the fully hierarchical match are given by the letters.

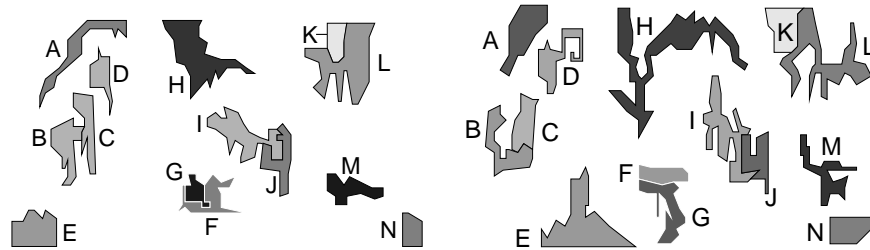


Figure 22: Matching the segmentation hierarchy shown in Figures 19b-o against the segmentation hierarchy shown in Figures 20b-o. Correspondences with respect to the fully hierarchical match are given by the letters.

- [4] P. Fossier, R. Glantz, M. Locatelli, and M. Pelillo. Swap strategies for graph matching. Accepted. In *Proc. of 12th Int. Conf. Image Anal. Process.*, 2003.
- [5] Y. Haxhimusa, R. Glantz, M. Saib, and W. G. Kropatsch. Logarithmic tapering graph pyramid. In L. V. Gool, editor, *Proceeding of the DAGM Symposium 2002*, volume 1876 of *Lecture Notes in Computer Science*, pages 117–124, Zürich, Switzerland, September 2002. Springer.
- [6] V. A. Kovalevsky. Finite topology as applied to image analysis. *Computer Vision, Graphics, and Image Processing*, Vol. 46:pp.141–161, 1989.
- [7] W. G. Kropatsch. Building Irregular Pyramids by Dual Graph Contraction. *IEE-Proc. Vision, Image and Signal Proc.*, 142(6):366 – 374, 1995.
- [8] M. Locatelli, I. M. Bomze, and M. Pelillo. Swaps, diversification, and the combinatorics of pivoting for the maximum weight clique. Technical Report CS-2002-12, Dipartimento di Informatica, Università Ca' Foscari di Venezia, 2002.
- [9] A. Massaro, M. Pelillo, and I. M. Bomze. A complementary pivoting approach to the maximum weight clique problem. *SIAM J. Optim.*, 12(4):928–948, 2002.
- [10] A. Meijsler and J. Roerdink. A Disjoint Set Algorithm for the Watershed Transform. In *Proc. of EUSIPCO'98, IX European Signal Processing Conference*, pages 1665 – 1668, Rhodes, Greece, 1998.
- [11] F. Meyer. Graph based morphological segmentation. In W. G. Kropatsch and J.-M. Jolion, editors, *2nd IAPR-TC-15 Workshop on Graph-based Representation*, pages 51–60. OCG-Schriftenreihe, Band 126, Österreichische Computer Gesellschaft, 1999.
- [12] J. Oxley. *Matroid theory*. Oxford University Press, New York, USA, 1992.
- [13] J. B. Roerdink and A. Meijster. The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. *Fundamenta Informatica*, 41:187–228, 2000.
- [14] A. Strehl and J. K. Aggarwal. A new bayesian relaxation framework for the estimation and segmentation of multiple motions. In *Proc. IEEE Southwest Symp. on Image Anal. and Interpret. (SSIAI 2000)*, pages 21–25. IEEE-Computer Society Press, 2000.