

Content-based image retrieval by matching hierarchical attributed region adjacency graphs

Benedikt Fischer¹, Christian Thies, Mark O. Güld, Thomas M. Lehmann
Department of Medical Informatics
Aachen University of Technology (RWTH), Aachen, Germany

ABSTRACT

Content-based image retrieval requires a formal description of visual information. In medical applications, all relevant biological objects have to be represented by this description. Although color as the primary feature has proven successful in publicly available retrieval systems of general purpose, this description is not applicable to most medical images. Additionally, it has been shown that global features characterizing the whole image do not lead to acceptable results in the medical context or that they are only suitable for specific applications. For a general purpose content-based comparison of medical images, local, i.e. regional features that are collected on multiple scales must be used. A hierarchical attributed region adjacency graph (HARAG) provides such a representation and transfers image comparison to graph matching. However, building a HARAG from an image requires a restriction in size to be computationally feasible while at the same time all visually plausible information must be preserved. For this purpose, mechanisms for the reduction of the graph size are presented. Even with a reduced graph, the problem of graph matching remains NP-complete. In this paper, the Similarity Flooding approach and Hopfield-style neural networks are adapted from the graph matching community to the needs of HARAG comparison. Based on synthetic image material build from simple geometric objects, all visually similar regions were matched accordingly showing the framework's general applicability to content-based image retrieval of medical images.

Keywords: Picture archiving and communication systems (PACS), Content-based image retrieval (CBIR), Graph thinning, Graph matching, Artificial intelligence, Neural networks, Computer-aided diagnosis, Shape analysis, Multi-scale image representation.

1. INTRODUCTION

Methods for the content-based retrieval of images are gaining importance together with the ever-increasing size of digital image archives. General approaches often use color as the primary source of information for the retrieval. In the context of medical images, however, this information is most often not available. Consequently, existing medical retrieval systems are limited to a specific application [1,2,3,4] since they are based on context-specific features. The project "Image Retrieval in Medical Applications" (IRMA, <http://irma-project.org>) aims at a general retrieval framework for medical images covering all imaging modalities and applications. For this purpose, in addition to global features, local, i.e. regional, features are utilized for image segmentation, leading to a region adjacency graph (RAG). In order to detect all visually plausible regions of an image, a multiscale decomposition of the image is performed. Each region is described by local features, which are stored as attributes in the graph. The graph therefore constitutes a hierarchical attributed region adjacency graph (HARAG). The comparison of images necessary for image retrieval can now be viewed as the problem of matching the corresponding HARAGs.

Graph matching has long been the subject of research. As in the paper of SHAPIRO and HARALICK often the number of consistent subgraphs has been used as a measure for the structural similarity between inexact graphs [5]. Other concepts, such as the one described by ESHERA and FU [6] have frequently used the edit distance as distance measure. Still another approach, which avoids the edit distance, is to find the maximal common subgraph of two graphs such as in [7] or [8]. All of the methods mentioned above though simply consider the graph structure for matching while neglecting the information within the nodes. In the context of IRMA, the main information if a given node (region) corresponds to

¹ Corresponding author: Benedikt Fischer, Department of Medical Informatics, Aachen University of Technology, Pauwelsstr. 30, D-52057 Aachen, Germany, email: bfischer@mi.rwth-aachen.de, web: <http://www.irma-project.org>

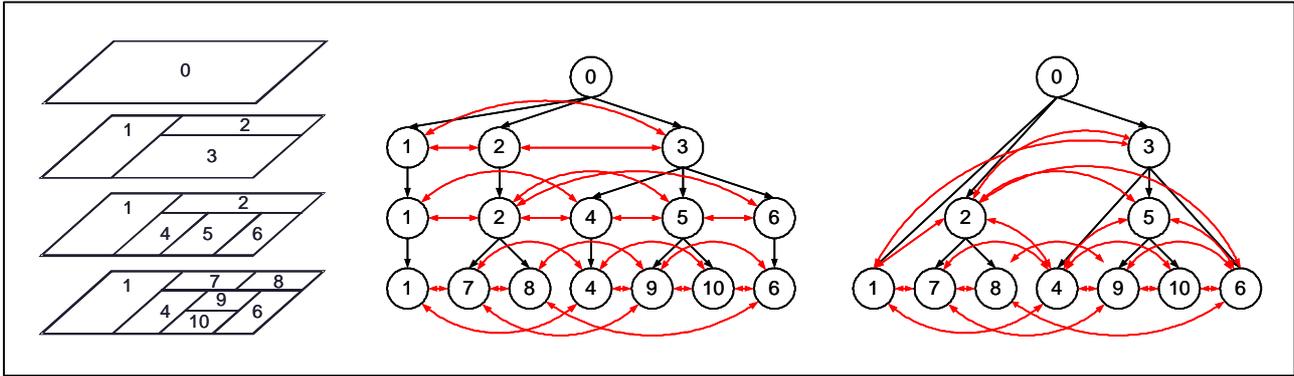


Figure 1: Multiple scales on an image (left) and corresponding hierarchical region adjacency graph (middle) and (right).

another one lies within the nodes itself. The edit distance could be adjusted to include distances between the blob features as well, yet does not seem appropriate simply by the sheer size of the graphs and the exponential number of possible edits.

Existing algorithms employing graph-matching for image retrieval are often carried out on manually segmented images, e.g. [9]. Hence, all objects relevant to the queries are directly available for the query. The objects cannot only be described by their feature vectors containing local information, but additionally by their spatial relationship used to create a graph [9,10]. For the purpose of a general framework for medical image retrieval, such as IRMA, however, no “correct” segmentation for all query contexts can be given as monoscale approaches are insufficient to extract all visually perceivable structures. Therefore a multiscale approach is needed which is able to meet the varying requirements of different query context. In this paper two strategies of graph matching are adapted to HARAG comparison and evaluated based on synthetical images. Beforehand, a strategy for graph thinning is introduced.

2. METHODS

2.1 Graph generation

Causal multiscale techniques allow the organization of image regions on different scales in a tree-like hierarchical structure. A well-explored approach builds such structures by a pixel-based region merging algorithm [11]. Here, all pixels are viewed as initial regions and two regions are merged with respect to their similarity in feature space. The merge process has to implement a transitive closure in order to avoid unbalanced clustering of adjacent regions. In addition, the non-linear filtering technique of Nagao and Matsuyama is applied as a pre-processing step to smooth homogeneous areas but simultaneously emphasize the contrast between adjacent regions [12]. Image regions are represented by a best-fitting ellipse (blob, [13]) and their mean feature vector. The hierarchical data structure originates from a protocol of the merging process. In addition to the hierarchy relation and the feature vector as node attributes, the adjacency relation between regions is stored in the graph yielding a HARAG. Figure 1 (middle) exemplifies such a HARAG generation.

The resulting tree has a worst-case size of $P \cdot (P+1)/2$ nodes if the underlying image consists of P pixels. Since for medical images sizes of 1024x1024 pixels and above are not uncommon, this constitutes far too many nodes to be handled efficiently by a matching algorithm. Hence, a reduction of the graph size is essential.

2.2 Graph thinning

In Figure 1, the first graph directly corresponds to the merging process depicted on the left. The nodes 1, 2, 4, and 6 are present multiple times in the graph, namely in each scale from their creation until the represented region is merged into a larger one. This does not only lead to redundancy, as the feature vectors are always stored with the node, but also complicates a lookup of a region’s adjacent regions as all scales have to be parsed. By only inserting a node once into the graph, the worst-case graph size for an image of P pixels can be reduced from $P \cdot (P+1)/2$ to $2 \cdot P$ nodes. On the other

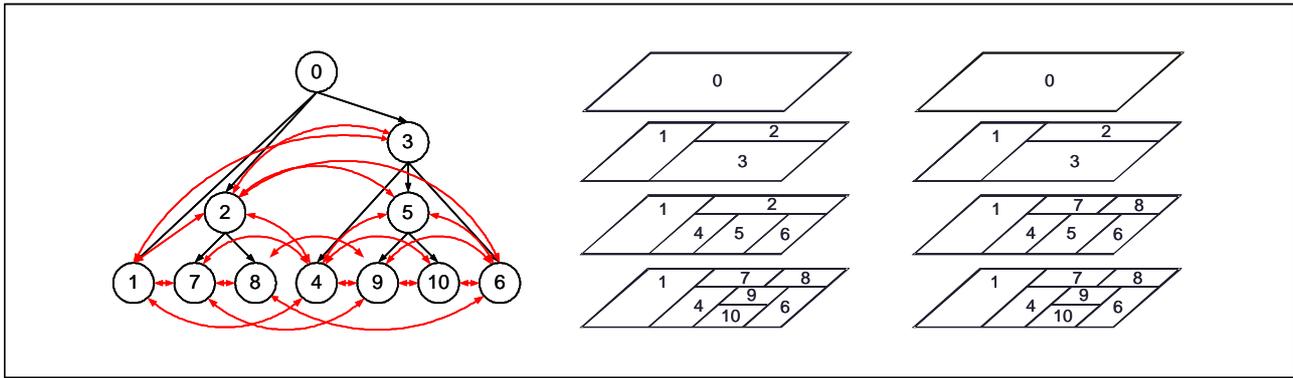


Figure 2: Ambiguous scale-information (middle, right) obtained from a graph of unique nodes (left).

hand, this may lead to ambiguities for the definition of the scales. Figure 2 illustrates this effect: Node 2 could have been created on the same scale as node 3 or as node 5. Therefore, the creation level of each node is stored as an additional node attribute.

Visually similar regions of separate images may be represented by nodes on different levels in their respective generated graphs. Therefore, a search for a likely matching node cannot be limited to the same level. If an algorithm matches nodes between two given graphs, a comparison of each node of the first graph with each node of the second graph leads to a runtime-complexity of $O(N \cdot M)$ when the graphs have N and M nodes, respectively. Obviously, the leaves, i.e. the regions consisting of only single pixels, can be pruned since a pixel-wise comparison of two images can be performed faster without graphs. Furthermore, in the first merging steps mostly small homogeneous regions are developed which do not contain important visual information. Therefore, these merging steps can be omitted in the graph structure and only the result of several of such merges is recorded as a single node. The number of omitted steps depends on the image class, which is determined in a prior processing step of the IRMA system.

2.3 Graph matching

After generating and thinning the graphs, the problem of matching nodes between the graphs has to be coped with. Mathematically, this problem constitutes the NP-hard search for an (inexact) subgraph-isomorphism. On account of the high resolution of today's medical images and considering that large databases with several thousand images have to be processed, this leaves heuristic solutions rather than exact algorithms as the only applicable approach.

Although graph matching has been common in various research fields for quite a while, the requirements for the application to image retrieval are somewhat different. In other research areas, the matching is computed solely by the analyzing the graph structure while disregarding any possible content in the nodes itself. Yet graphs for the representation of images such as the HARAGs used in this project store the information about the image regions itself in the attributes attached to each node. The graph edges provide supplementary information about the relations between the nodes, e.g. inclusion over different scales, adjacencies, relative positions, and several more. Depending on the user, the similarity of images relies primarily on the similarity of the depicted regions and only secondary on the relations between them or vice versa. Therefore, any graph matcher for image retrieval must be able to include information possibly contained in the nodes. For this purpose, two graph matching strategies which are able to exploit node contents and relational information have been adapted for their application to image retrieval:

1. Similarity Flooding [15],
2. Hopfield-style Neural Networks [16]

Both algorithms use an additional graph ("Pairwise Connectivity Graph", PCG, in [15], "Extended Compatibility Graph", ECG, in [16]) as data structure for the similarity computation. For simplicity, this graph will be referenced as Mapping Graph (MG) throughout the remainder of this text, if the general approach common to both algorithms is being referred to. Each node of the MG corresponds to one mapping hypothesis, i.e. a mapping of one node of the first graph (query graph), to one node of the second one (range graph). For any node q of a query graph Q and any node r of a

range graph R , the node mapping q on r is expressed by a node (q, r) in MG. The MG-generation, the type of edges, and the iterative processes to calculate the quality of matches differ in both approaches and need separate explanation.

2.3.1 Similarity Flooding

For SF, a mapping is inserted into MG only for node-pairs within the graphs, i.e. nodes q_1 and q_2 of the query graph are mapped onto nodes r_1 and r_2 of the range graph, only if q_1 and q_2 share the same relational edge e between them as r_1 and r_2 . If this is the case, then the nodes (q_1, r_1) and (q_2, r_2) are inserted into MG and connected bidirectionally by edge e . The philosophy behind this procedure expresses that regions of the query and the range graph are more likely to be similar, if they are connected to other regions in the same way, e.g. in the inclusion-hierarchy, adjacency or other relations.

The direct comparison of the nodes, i.e. regions, which are mapped onto each other by the MG-nodes is computed by a distance function over the nodes' feature vectors. Therefore, each MG-node is assigned an initial distance value σ^0 . In a fixpoint operation, these initial distance values are "flooded" along the edges to neighboring nodes. Equation (1) shows how the distance value of a node (q, r) is obtained for the next iteration $i+1$:

$$\sigma^{i+1}(q, r) = \sigma^i(q, r) + \sum_{((a,b),e,(q,r)) \in \text{MG}} \sigma^i(a, b) \cdot \omega_e(q, r) \quad (1)$$

$$\text{with } \omega_e(q, r) = \frac{\omega_e}{\#\text{incident edges to } (q, r)}$$

Here, the next distance value $\sigma^{i+1}(q, r)$ of a node (q, r) in MG is assigned the sum of its current distance value $\sigma^i(q, r)$, and the cumulated weighted distance values of all its adjacent nodes. The values of the neighboring nodes are weighted by ω_e i.e. dependant on the type of relation e , and normalized by the number of incident nodes. This way, the user can express which relation should be considered more or less important, i.e. the adjacency may be regarded more important than the inclusion over scales. The iteration is repeated until the Euclidean length of the residual vector $\Delta(\underline{\sigma}^k, \underline{\sigma}^{k+1})$ becomes less than a threshold ε for some $k > 1$, and where $\underline{\sigma}^k$ and $\underline{\sigma}^{k+1}$ are the normalized versions of σ^k and σ^{k+1} , respectively. If no fixpoint is reached, the iteration is stopped after a user-specified number of maximum iterations. This stopping criterion can also be used to limit the influence of the neighboring nodes as after k iterations, all neighbors within a path length smaller than or equal to k have influenced the distance value of a node.

After the iteration process, distance values for the nodes of Q and R have been computed. These provide a mean to compare single image regions, but the global distance of the represented images needs to be evaluated. For this purpose, the global distance σ_{total} of equation (2) compares images independent of the number of iterations and number of mappings:

$$\sigma_{\text{total}} = \frac{\sum_{(q,r) \in \text{MG}} \sigma^k(q, r)}{k \cdot N} \quad (2)$$

In equation (2), k denotes the number of iterations, $\sigma^k(q, r)$ the final distance value of a node (q, r) , and N the number of nodes of MG.

The following modifications of the original approach [15] were necessary in order to adapt the method for its application to image retrieval.

- In order to avoid unlikely mappings and to keep the MG small in the first place, the initial distance σ^0 is computed even before the generation of a MG-node. Only mappings leading to distance values above a user-selected threshold θ are inserted into the MG. By indexing of the features used in the feature vectors, this leads to a rapid narrowing down on likely match partners. Otherwise, the method would not be applicable to large image databases.
- Equation (1) is an adjusted version of the original version, to account for the differing importance for the various relations. Additionally, the equal influence of all neighboring nodes (apart from the relational weight) is guaranteed. In the original approach, nodes with more neighbors gained more distance.
- The original approach calculates similarities or distances on a node-to-node basis. In order to compare the whole images, the global distance of equation (2) has been added.

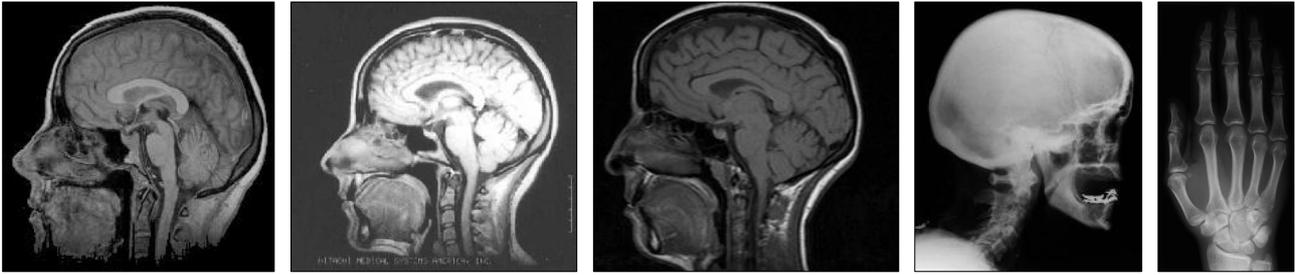


Figure 3: Arbitrary images used for the graph thinning.

2.3.2 Neural Network Graph Matching

The second strategy for graph matching uses a Hopfield-style neural network which incorporates structural matching as well as constraint relaxation by spreading activation in a neural network [16]. In contrast to the Similarity Flooding approach, where only positive impact to the similarity values is spread, the neural network is able to express negative (inhibitive) influences as well. Instead of similarity values, potentials are assigned to the mapping nodes, while the edges between them are used to denote the consistency or inconsistency of the mappings, depending on the sign of their weights. During the construction of the network, edges with negative weights are inserted between nodes which would lead to ambiguous mappings, e.g. between a node (q_1, r_1) and (q_1, r_2) . After the simulation process, the network extracts all stable potentials as the final mappings. In contrast to Similarity Flooding, the network is guaranteed to converge.

The theoretical framework of the original approach [16] did not need adaptations to be applied to image retrieval. For the context of the HARAGs used in the IRMA project, the similarity of relations, which is used for the global distance function, is set to 1, if the relations are the same, otherwise to 0. If, for future evaluations, other spatial relationships, such as “above”, “left of”, etc. are desirable, then this similarity measure can provide quantitative measures for these relations. At the current stage, however, this is not necessary. Therefore the main effort went into a new implementation capable of handling large graphs. While the original implementation rejects simulation runs with more than 300,000 connections based on performance reasons, the re-implemented version is able to handle networks of 15 million connections in twelve minutes on standard PC-hardware running at 1 GHz. The performance benefit is achieved through a partitioning of the network into clusters which exchange their results and missing information in cycles. The cluster-size can be adjusted to the executing PC’s memory and workload and, therefore lead to little swapping even with large graphs. The division into clusters also is the basis for computations in a distributed environment. However, currently, no protocol for the necessary communication between the participating computers has been designed, as the system has not been fully evaluated on single computers, yet.

3. EXPERIMENTS

3.1 Graph thinning

First experiments were conducted to evaluate the effectiveness of the graph thinning and to show the general applicability of the graph matching algorithms. Figure 3 shows five images, which were selected arbitrarily for a graph reduction. All images were segmented using the algorithm from [11] and reduced in size as described in Section 2.2.

3.2 Graph matching

For the evaluation of the graph matching algorithms, synthetic images containing simple objects were generated to allow the controlled validation of the found matches. Figure 4 depicts examples of the images used for testing. The synthetic images lead to small hierarchy graphs of less than 15 nodes. This limits the size of the MG to below 225 nodes and leaves manual checks on the produced graphs feasible. Further, the regions are simple enough to guarantee a sufficient description by features. Therefore, the quality of the computed matches can be easily verified. The experiments are used to evaluate the general applicability of the graph matching algorithms. If they do not lead to acceptable results under these controlled circumstances, a further use under real world conditions, where the sizes of the matching graphs are not human-manageable and where the features are not guaranteed to lead to perfect distinctions does not make sense. The features used for the synthetic images are their mean gray-value and the region-size in pixels. All images consist of 256x256 pixels. For an example of the matching results, the images depicted in Figure 5, top and

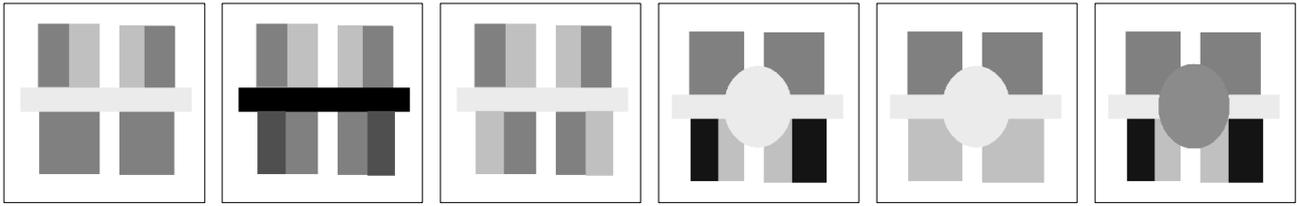


Figure 4: Examples of artificial test images for the graph matching strategies.

middle row are matched. For an easier validation of the given results, the different scales are also shown in the figures. The corresponding HARAGs are depicted in Figure 5, bottom line.

4. RESULTS

4.1 Graph thinning

The first results on graph thinning indicate that the reduction of graphs is possible without significant loss of information. An image of P pixels leads to a graph of $\mathbf{O}(2P)$ nodes, if the nodes are inserted only once into the graph. The thinning reduces the graph size well below $\mathbf{O}(P)$. Examples of the thinning obtained on the graphs obtained from the images of Figure 3 are provided in Table 1. In all examples, reduction of about 98% is achieved regardless of initial image size, imaging modality, and body region examined.

Image	Modality	Dimensions (pixels)	# Nodes	# Nodes of thinned graph	Reduction in %
1	MRI	200x200	43783	751	98.28
2	MRI	276x262	72316	1268	98.24
3	MRI	256x228	58368	798	98.63
4	x-ray	177x212	40753	564	98.62
5	x-ray	500x256	128000	1414	98.89

Table 1: Sample graph sizes corresponding to the images of Figure 4.

4.2 Graph matching by Similarity Flooding:

All results were obtained with the threshold $\theta = 10$ for the maximum distance before insertion into the MG and $\epsilon = 0.1$ for the fixpoint-determination. The flooding process took between 5 and 10 iterations until a fixpoint was reached. The computations took less than 2 seconds for all examples on a standard PC operating at 1 GHz. All regions were matched to visually similar regions in the other graphs. In some cases, multiple possible matchings for regions were computed.

As an example for a typical outcome, the final results of matching the graph of Figure 5, top row to the one of Figure 5, middle row, are provided in Table 2. Here, obviously the regions with the same color and size were matched onto each other correctly (mapping-pairs (2,3), (3,4), (4,5), (5,6), (6,8), (7,7), (10,10), and (11,9)). Of these matches, also the neighboring aspects are accounted for, e.g. neighboring nodes 7 and 11 of *rect1* are mapped to neighboring nodes 7 and 9 of *rect2*. The same holds for nodes 6 and 10 which are mapped to nodes 8 and 10, respectively. The mapping (1,2) seems remarkable, as for a human observer, the identical shapes of both the top-level regions 1 would suggest a mapping (1,1). However, the shape-information was not included for the matching vector. The mapping can be explained by the fact, that if only the graph structure is concerned, the graph of *rect1* is an exact subgraph of the graph *rect2*, if node 1 of the first graph is assigned node 2 of the second one. Further, both regions "1" are not of identical color, but by the merging process which calculates the mean for all merged pixels, region 1 of *rect2* is darker than region 1 of *rect1*. Hence, the assignment (1,2) can also be seen as successful. In summary, all regions were matched to regions which are similar according to the used features and relations.

4.3 Matching by Hopfield-style neural network

All results were obtained using the same parameters derived from a calibration run on the matching of the graph of image *rect1* to itself. The number of mapping nodes ranged between 57 and 132, which were reduced to final 9 and 13,

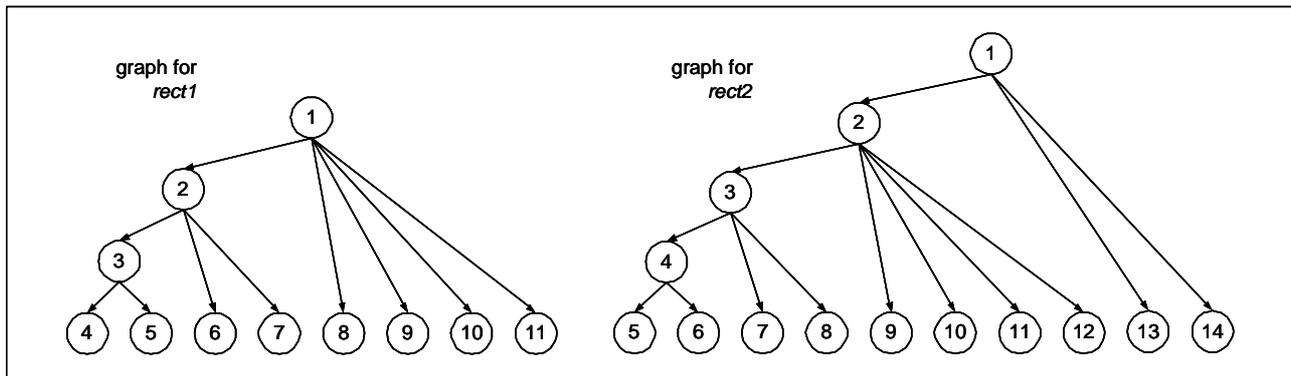
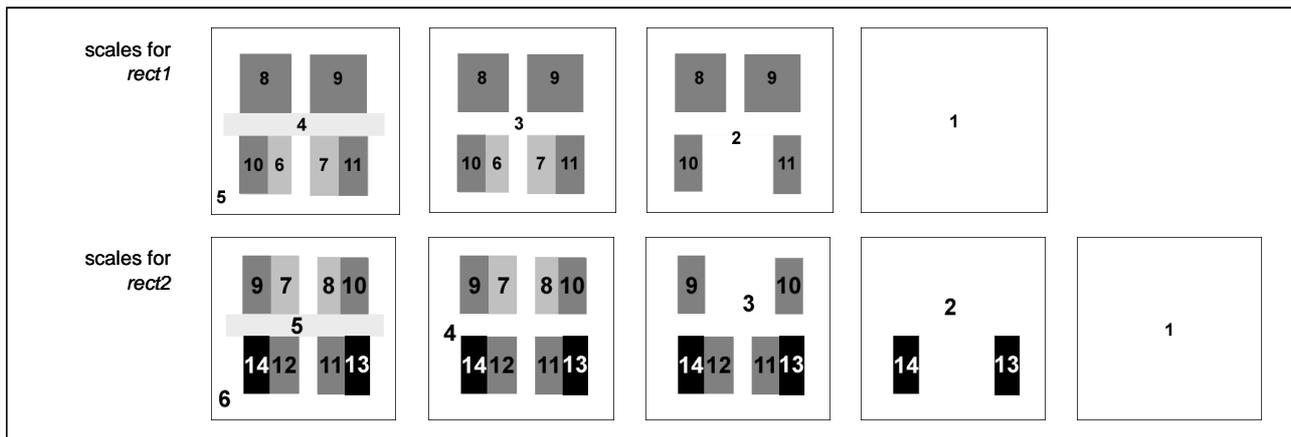


Figure 5: Images *rect1* (top row) and *rect2* (middle row) with their resulting scales from left to right and the corresponding hierarchy graphs, (bottom). The numbers shown are region labels. The adjacencies have been omitted.

respectively, corresponding to the number of query-graph nodes. It took between 129 to 531 iterations until a convergence was achieved. The runtimes were below one second on the same machine that was used for Similarity Flooding. The global distances are relatively high for the matching of identical graphs to each other, yet all distances express the expected relations, i.e.

- distances between identical graphs are lower than between other matchings,
- distances between an image and its mirrored version are smaller than between non-identical, non-mirrored images, and
- distances are identical if query and range graph are swapped, i.e. it is independent of which graph is used as the query graph.

The results for the example matching of *rect1* and *rect2* are summarized in Table 2. The mappings are identical to the ones obtained by Similarity Flooding except for the matchings (8, 11) and (9,12). The two differing mappings are acceptable since the colors of the mapped regions match, even though the size is not the same. The fact that the matches are of less quality is well expressed by the fact, that their finally activity is well below the values of visually more similar regions.

5. DISCUSSION

The presented methodology allows the comparison of medical images by means of hierarchical attributed region adjacency graphs (HARAGs). With the presented thinning, the sizes of the HARAGs are diminished while preserving the possibility to exploit the local features necessary for medical applications. The two graph matching strategies of [15]

Similarity Flooding			Neural Network		
Querynode	Rangenode	Distance	Querynode	Rangenode	Activity
1	2	2.939	1	2	7.437
2	3	2.348	2	3	7.151
3	4	1.742	3	4	6.737
4	5	1.454	4	5	7.235
5	6	1.457	5	6	6.421
6	8	2.032	6	8	4.896
7	7	2.032	7	7	4.921
8	12	2.705	8	11	1.057
9	12	2.712	9	12	1.055
10	10	2.668	10	10	5.736
11	9	2.668	11	9	5.686

Table 2: Final results for matching *rect1* and *rect2* by Similarity Flooding and Hopfield-style neural network.

and [16] have been adapted for image retrieval and first encouraging experiments on synthetic data have been conducted. The experiments show the generally applicability of both techniques for image retrieval. The implementations allow an easy integration of suitable global and regional features and offer a weighting of the influence of the graph structure versus the node contents. As evaluated by synthetic images, the image retrieval task can be fulfilled, provided that the node and edge attributes are suitable descriptions of the regions and their relations. Further experiments can therefore use the implemented algorithms to discover adequate feature sets and distance measures for desired contexts. The testbed for a full-scale evaluation on the complete database of approximately 10,000 medical images has been created and is ready to be used for exhaustive assessment. Overall, a flexible system for content-based image retrieval is formed which is not limited to a specific application or medical context, but which is general enough to cover all imaging modalities and body regions examined.

ACKNOWLEDGEMENT

This work was performed within the project image retrieval in medical applications (IRMA), which is supported by the German Research Community (Deutsche Forschungsgemeinschaft, DFG), grants Le 1108/04 and Le 1108/06. For further information please visit <http://irma-project.org>.

REFERENCES

1. Liu L, Sclaroff S: Medical image segmentation and retrieval via deformable models. Proceedings ICIP, 3: 1071-74, 2001.
2. Long LR, Thoma GR: Computer assisted retrieval of biomedical image features from spine X-rays: progress and prospects. Proceedings 14th CBMS 2001, 46-50, 2001.
3. Shunshan Li, Tiange Zhuang, Hui Chen: Medical image retrieval based on mutual correlation method. Proceedings SPIE 4323: 406-412, 2001.
4. Shyu CR, Matsakis P: Spatial lesion indexing for medical image databases using force histograms. Proceedings CVPR, 2: 603-608, 2001.
5. Shapiro L, Haralick R: A metric for comparing relational descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI 4(7): 90-94, 1985.
6. Eshera M, Fu K: An image understanding system using attributed symbolic representation and inexact graph-matching. Journal of ACM 8(5): 604-618, 1986.
7. Bunke H, Shearer K: A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters 19: 255-259, 1998.

8. Baader F, Küsters R. Computing Least Common Subsumers in Description Logics with Existential Restrictions. Proceedings IJCAI, 96-101, 1999.
9. Petrakis EGM: Fast Retrieval by Spatial Structure in Image Databases. *J. Vis. Lang. Comput.* 13(5): 545-569, 2002.
10. Huet B, Hancock ER: Relational object recognition from large structural libraries. *Pattern Recognition* 35: 1895-1915, 2002.
11. Thies C, Malik A, Keysers D, Kohnen M, Fischer B, Lehmann TM: Hierarchical feature clustering for content-based retrieval in medical image databases. Proceedings SPIE 5032(1): 598-608, 2003.
12. Nagao M, Matsuyama T: Edge preserving smoothing, *Computer Graphics Image Processing* 9(4): 394-407, 1979.
13. Carson C, Belongie S, Greenspan H, Malik J: Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI* 24(8): 1026-1038, 2002.
14. Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohnen M, Schubert H, Wein BB: Content-based image retrieval in medical applications. *Methods of Information in Medicine* 2004, in press.
15. Melnik S, Garcia-Molina H., Rahm E: Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching. Proceedings 18th ICDE, 117-128, 2002.
16. Schädler K, Wysotzki F: Comparing Structures using a Hopfield-style Neural Network. *Applied Intelligence* 11: 15-30, 1999.