Matching of Structural Prototypes for Content-Based Medical Image Retrieval

Benedikt Fischer, Michael Sauren, Mark O. Güld, Ilja Bezrukov, Thomas M. Deserno
Department of Medical Informatics, RWTH Aachen University of Technology, Aachen, Germany

ABSTRACT

The relevant contents of medical images can often be described as a composition of objects with distinct relationships. Each object is represented as a graph node with regional features as node attributes e.g. centroid coordinates. The relations between objects are represented as graph edges with annotated relational features, e.g. relative size. For a given setting, e.g. a hand radiograph, a generalization of the relevant objects, e.g. hand bones, can be obtained by the statistical distributions of the attributes. These yield a structural prototype graph which consists of one node per relevant object. In contrast to the aforementioned graph, the mean and standard deviation of each regional or relational feature are used to annotate the prototype nodes or edges, respectively. The prototype graph can then be used to identify the generalized objects in new images. As new image content is represented by hierarchical attributed region adjacency graphs (HARAGs) which are obtained by region-growing, the task of object identification corresponds to the problem of inexact graph matching between the small prototypes and HARAGs. For this purpose, five approaches are evaluated in an example application of bone-identification in 96 hand-radiographs: Nested Earth Mover's Distance, Graph Edit Distance, a Hopfield neural net, Pott's Mean Field Annealing and Similarity Flooding. The discriminative power of 34 regional and 12 relational features for each object is judged by sequential forward selection. The structural prototypes improve recall by up to 17% in comparison to the approach without relational information.

Categories / Keywords: Classification, Statistical methods, Neural nets

1. PURPOSE

Due to the high inter- and intra-patient variability of depicted regions of interest it is infeasible to define precise normative values for the attributes of an object-class. For example, the regional attributes of the metacarpal bones in radiographs are not discriminative enough to allow a distinction between the single fingers. Even for a human observer it is impossible to identify the corresponding finger without knowledge of a bone's surrounding. However, this becomes easier when relations to other bones (position, size, etc.) are known.

A structural prototype graph therefore aims to describe the relevant objects (nodes) and their relations (edges) as a scene. Since only relevant objects are contained, the graph is very small. In the example application, the 19 hand bones are represented by an identical number of nodes.

During the partitioning of images with unknown content and resolution, each pixel could belong to different segments. Hence, a generic data representation is required, which preserves these ambiguities. For this purpose, an iterative region-growing concept is applied, transforming the image to a hierarchical attributed region adjacency graph (HARAG) [1]. The resulting graphs are rather large. In the example application, the HARAGs for images of 256 pixels maximum width and length lead to graphs between 1,000 and 6,000 nodes.

As prototype and image are both represented as graphs, the object identification is equivalent to the problem of inexact graph matching for which many established solutions exist [2]. Due to the huge search space, we chose the following approximation algorithms: the Nested Earth Mover's Distance [3], the Graph Edit Distance (GED) [4], a Hopfield neural network (NNGM) [5], Pott's Mean Field Annealing [6], as well as Similarity Flooding [7].

2. METHODS

2.1 Structural Prototype Graph

The prototype graph represents objects by nodes and object-relations by edges. Nodes and edges are attributed by the unimodal Gaussian distributions of the regional and relational attributes, respectively.

Means and standard deviations are obtained for manually labeled segments. The similarity of nodes and edges of the prototype to the HARAG is computed by the Mahalanobis distance with an additional class-dependent weight for each dimension. The weights were adjusted on the training data by sequential forward selection so that the classification rate was maximized.

2.2 Graph Matching

The Earth Mover's Distance (EMD) is based on a solution to the transportation problem in linear optimization and has been used for image retrieval as a metric between two distributions, e.g. color histograms [8]. An application of the EMD for graph-matching is presented in [3] as a nested EMD (NEMD). In a first step, the EMD is computed for each pair of nodes, taking into account node and edge attributes. The results of these computations are then used as input for a "outer" EMD, computing the overall similarity between the compared graphs and leading to a final node matching.

The well-known Graph Edit Distance transforms one graph into another with pre-defined transform costs for substitution, deletion and insertion of nodes or edges [4]. The matching is optimal if the sum of all transforms is minimal. In order not to cover the whole search space, a heuristic is used additionally: Beginning with an empty partial matching, in each iteration step the transform with the minimal costs is chosen to extend the partial matching. This is repeated until an end-state (full matching) is reached. To further reduce the complexity, a heuristic function estimating the costs following a selection is used.

For the Hopfield net, an association graph (AG) between prototype and HARAG nodes is used [5]. Each AG node represents one matching hypothesis. A pre-filtering is used to only include pairs of similar nodes in the AG. The regional similarity is used as initial potential for the AG nodes. The final matching is derived by a gradient descent optimization process. Matching hypotheses are strengthened or weakened by similarity to regional or relational distributions. The process iterates until a stable state is reached.

Pott's Mean Field Annealing is a variant of the simulated annealing technique [6]. The aim is to approximate the stochastical simulated annealing by a deterministic approach leading to better run-times and less sensitivity to parameter settings. For the application an AG and a Hopfield net as described above is used, yet the solution space is constrained such that the state vector is binary containing only a single component set to one. This reduces the solution space from 2^k to k for each AG-node.

Similarity Flooding [7] also uses an AG to describe possible node matchings. The main idea is that for two nodes of different graphs to be similar, their neighborhoods in each graph must also be similar. The nodes of the AG therefore spread or "flood" their similarity to neighboring matching hypotheses. This process is iterated until stability or a user-defined number of iterations is reached.

2.3 Experiments

Aim of the experiments was the identification of 19 hand bones in 96 hand radiographs. The HARAG segmentations led to 96 graphs with a total of 221,615 regions and 1,000 to 6,000 nodes per graph, on average 2,300 nodes. The ground truth was established by manually labeling the bones for each image if they had been segmented, resulting in 1,290 labels. Fig. 1 displays the labels (left) and the percentage of how often each bone class occurred in the segmentation.

34 regional attributes were computed for shape, intensity, and texture. 12 relational attributes were used for topology, size, and intensity relations.

The experiments were conducted using a six-fold cross validation.

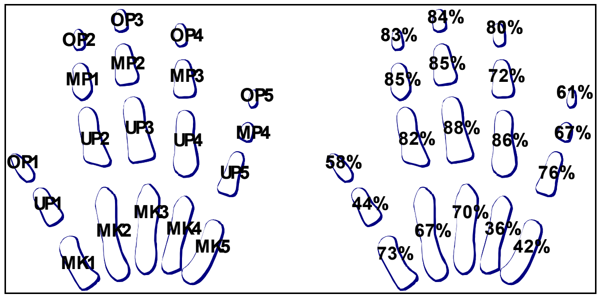

Fig. 1: Bone class labels and percentage of occurrences in the ground truth data.

3. RESULTS

The results as shown in Tables 1 and 2 clearly demonstrate the beneficial use of structural information. Except the Pott's Mean Field Annealing, which is almost identical to the regional classification, all methods always improve the classification results. The only exception is the bone group MP using GED matching, which performs roughly equal to only regional analysis. In particular, the classification using only regional, i.e. no relational information, classifies 60.16% correctly, the NEMD 65.0% (8% increase), the GED 65.2% (9% increase), the NNGM even 70.5% (17% increase), the PFMA 60.5% (0.2% increase), the SF 69.8% (16% increase). The computation time for the prototype-generation was 15 seconds per prototype. The classifications experiments took on average about 20, 21, 18, 30, 5, and 5 seconds per graph, respectively. Runtimes were taken on a 2 GHz PC with 2 GB RAM.

4. NEW OR BREAKTRHOUGH WORK

Although graph-matching has been used for image or object retrieval for a long time, the graphs involved are usually limited to a small number of nodes. Yet in a generic setting, where the image content and resolution is unknown prior to the segmentation, a generic data-structure such as the HARAG is needed, resulting in large graphs with thousands of nodes. Hence the matching process for these problem instances is very difficult. The presented approach therefore features the concept of a small prototype graph as a generalization of the sought objects which is to be matched with the large generic HARAG. For this uncommon approach in image retrieval five well-known optimization techniques (respectively distance metrics) are evaluated and compared to matching without considering the relations between objects. In contrast to active shape models, that consider only a very special constellation of shapes, we presented a generally applicable framework for structural object description and scene analysis in medical imaging with an example evaluation for the difficult task of hand-bone identification.

5. CONCLUSION

Considering that only 19 of on average 2,300 nodes, i.e. 0.8%, in the HARAGs represent hand bones, an overall classification rate of 70.5% is considered as satisfactory. Even more so, as the algorithms may have classified regions similar to the real bones but not labeled in the ground truth as such. Currently the segmentation quality is being improved significantly, which is crucial for the information content of the HARAGs and the training of the prototypes. In addition to further matching algorithms, we will investigate different models for node and edge prototypes in combination with Bayesian classification.

Table 1. Recall in percent of ground truth by bone-group (see Fig. 1); regional: no relations used, NEMD: Nested Earth Mover's Distance, GED: Graph Edit Distance, NNGM: Hopfield neural net, PMFA: Pott's Mean Field Annealing, SF: Similarity Flooding

group	regional	NEMD	GED	NNGM	PMFA	SF
MK	64,5%	72,1%	69,9%	75,4%	64,9%	64,9%
Uph	70,4%	70,6%	74,0%	78,1%	70,4%	75,1%
MP	64,7%	70,7%	64,0%	75,7%	64,7%	76,3%
OP	42,5%	48,7%	55,0%	54,4%	42,8%	62,9%
overall	**60,2%**	**65,0%**	**65,6%**	**70,5%**	**60,3%**	**69,8%**

Table 2. Recall in percent of ground truth for each bone (see Fig. 1); regional: no relations used, NEMD: Nested Earth Mover's Distance, GED: Graph Edit Distance, NNGM: Hopfield neural net, PMFA: Pott's Mean Field Annealing, SF: Similarity Flooding

class	regional	NEMD	GED	NNGM	PMFA	SF	class	regional	NEMD	GED	NNGM	PMFA	SF
MK1	67,1%	70,0%	72,9%	74,3%	68,6%	38,6%	MP2	54,9%	62,2%	51,2%	64,6%	54,9%	64,6%
MK2	70,3%	76,6%	68,8%	78,1%	71,9%	68,8%	MP3	72,0%	82,9%	75,6%	82,9%	72,0%	85,4%
MK3	68,7%	71,6%	80,6%	79,1%	68,7%	76,1%	MP4	72,2%	79,2%	73,6%	87,5%	72,2%	87,5%
MK4	68,6%	74,3%	60,0%	74,3%	65,7%	85,7%	MP5	59,4%	56,3%	54,7%	67,2%	59,4%	67,2%
MK5	40,0%	67,5%	57,5%	67,5%	40,0%	67,5%	OP1	37,5%	51,8%	57,1%	57,1%	37,5%	58,9%
UP1	64,3%	61,9%	52,4%	69,0%	64,3%	69,0%	OP2	35,0%	43,8%	55,0%	53,8%	35,0%	57,5%
UP2	73,4%	65,8%	72,2%	74,7%	73,4%	69,6%	OP3	54,3%	55,6%	59,3%	54,3%	54,3%	66,7%
UP3	70,2%	79,8%	72,6%	75,0%	70,2%	73,8%	OP4	46,8%	45,5%	57,1%	61,0%	48,1%	70,1%
UP4	80,7%	68,7%	85,5%	85,5%	80,7%	80,7%	OP5	35,6%	47,5%	44,1%	44,1%	35,6%	59,3%
UP5	58,9%	72,6%	76,7%	82,2%	58,9%	79,5%	**overall:**	**60,2%**	**65,0%**	**65,6%**	**70,5%**	**60,3%**	**69,8%**

SUBMISSION INFORMATION

This paper has not been submitted elsewhere.

REFERENCES

1 Lehmann TM, Beier D, Thies C, Seidl T: Segmentation of medical images combining local, regional, global, and hierarchical distances into a bottom-up region merging scheme. Procs SPIE 2005; 5747:545-555

2 Conte D, Foggia P, Sansone C, Vento M: Thirty years of graph matching in pattern recognition. International Journal of Pattern Recognition and Artificial Intelligence 2004; 18(3): 265-298

3 Kim DH, Yun ID, Lee SU: A new attributed relational graph matching algorithm using the nested structure of Earth Mover's Distance. International Conference on Pattern Recognition 2004; 48-51

4 Bunke H, Allermann G: Inexact graph matching for structural pattern recognition. Pattern Recognition Letters 1983; 1: 245-253

5 Lappe C, Fischer B, Thies C, Güld MO, Kohnen M, Lehmann TM: Optimierung eines konnektionistischen Graphmatchers zum inhaltsbasierten Retrieval medizinischer Bilder. In: Tolxdorff T, Braun J, Handels H, Horsch A, Meinzer HP (Hrsg) Bildverarbeitung für die Medizin 2004, Springer-Verlag, Berlin, 2004; 338-342

6 Peterson C, Söderberg B. A new method for mapping optimization problems onto neural networks. International Journal of Neural Systems 1989; 1: 3

7 Melnik S, Garcia-Molina H, Rahm E. Similarity flooding - A versatile graph matching algorithm and its application to schema matching. Proc. 18th ICDE Conf. 2002; 117-128

8 Rubner Y, Tomasi C, Guibas LJ: The Earth Mover's Distance as a metric for image retrieval. International Journal of Computer Vision 2000; 40(2): 99-121