Segmentation of medical images combining local, regional, global, and hierarchical distances into a bottom-up region merging scheme

Thomas M. Lehmann[1], Daniel Beier, Christian Thies, and Thomas Seidl[a]

Department of Medical Informatics
[a] Chair of Computer Science IX
Aachen University of Technology (RWTH), Aachen, Germany

ABSTRACT

Segmentation of medical images is fundamental for many high-level applications. Unsupervised techniques such as region growing or merging allow automated processing of large data amounts. The regions are usually described by a mean feature vector, and the merging decisions are based on the Euclidean distance. This kind of similarity model is strictly local, since the feature vector of each region is calculated without evaluating the region's surrounding. Therefore, region merging often fails to extract visually comprehensible and anatomically relevant regions.

In our approach, the local model is extended. Regional similarity is calculated for a pair of adjacent regions, e.g. considering the contrast along their common border. Global similarity components are obtained by analyzing the entire image partitioning before and after a hypothetical merge. Hierarchical similarities are derived from the iteration history. Local, regional, global, and hierarchical components are combined task-specifically guiding the iterative region merging process. Starting with an initial watershed segmentation, the process terminates when the entire image is represented as a single region. A complete segmentation takes only a few seconds.

Our approach is evaluated contextually on plain radiographs that display human hands acquired for bone age determination. Region merging based on a local model fails to detect most bones, while a correct localization and delineation is obtained with the combined model. A gold standard is computed from ten manual segmentations of each radiograph to evaluate the quality of delineation. The relative error of labeled pixels is 15.7 %, which is slightly more than the mean error of the ten manual references to the gold standard (12 %). The flexible and powerful similarity model can be adopted to many other segmentation tasks in medical imaging.

Keywords: Segmentation, Region Growing, Region Merging, Distance Measure, Similarity Measure

1. INTRODUCTION

Segmentation of medical images is fundamental for many applications of high-level image analysis such as object identification or object-based image compression [1]. For image interpretation and automatic description of image content, more than one object is considered. Typically, this results in a partitioning of the given image. The partitions are called regions and should represent visual important objects displayed in the image. In the medical domain, bones, vessels, organs and other hard or soft tissues are considered as objects and different levels of details are required with respect to the medical task. Usually, large data volumes are acquired in health care and, therefore, must be processed automatically. This is possible using unsupervised techniques such as region growing or region merging [2].

ADAMS & BISCHOF have developed a supervised segmentation algorithm that requires the definition of k seed points or seed regions at the initial algorithm stage [3]. These seed regions define the starting points for a region growing process. In the subsequent steps, the image pixels that lie at the border to at least one of the k regions are examined. The pixel that is most similar to its adjacent region is added to this region. This process is repeated until each pixel is assigned to

---

[1] Corresponding author: Priv.-Doz. Dr. Thomas M. Lehmann, Assoc. Prof., Department of Medical Informatics, Aachen University of Technology, Pauwelsstr. 30, D - 52057 Aachen, Germany, email: lehmann@computer.org; web: http://irma-project.org/lehmann, phone: +49 241 80-88793.
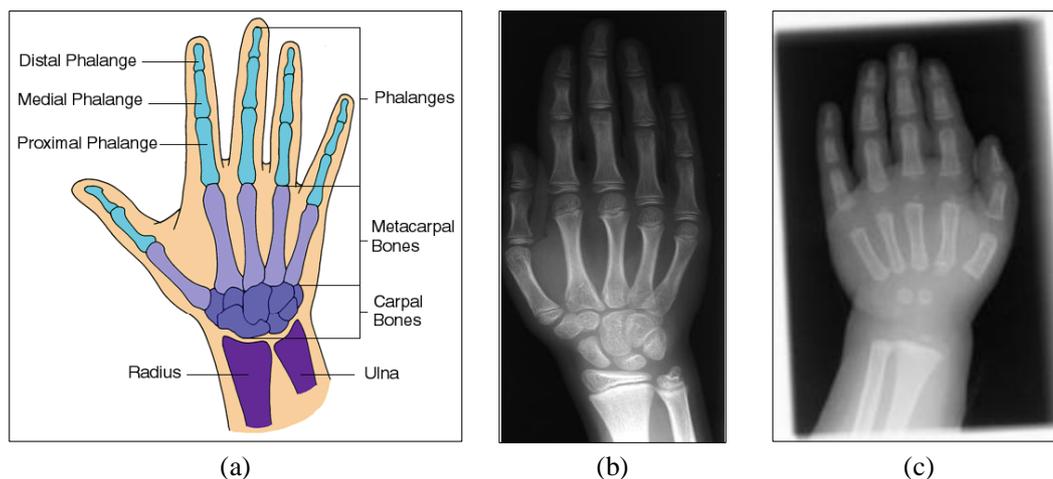
Figure 1: Skeletal radiograph of a human hand: (a) taxonomy of bony elements; (b) and (c) examples acquired in clinical routine. Note the variety in position, exposure, and form of relevant bones characterizing the age of the patient.

one of the k regions. The output of this algorithm is one possible partitioning of the image, i.e. a single-scale description of the image content. A variant of this algorithm was proposed by MEHNERT & JACKWAY [4] that overcomes the dependency of the processing order of the pixels. Further expansions of this approach mostly aim at the automatic seed pixel generation [5, 6].

However, the growing process is controlled by the similarity or distance measure, which is usually based on a mean feature vector describing each region and the Euclidean distance. This opens the so called semantic gap between the low-level nature of features that are used in computerized image processing and the high-level nature of image understanding by human observers [7]. Hence, unsupervised segmentation often fails to extract regions, which are visually comprehensible and anatomically relevant.

Clustering in feature space was proposed to overcome this problem [8]. To preserve local connectivity of image partitions, the pixel coordinates are added to the feature vectors. Each region is described by a statistical distribution, e.g. a Gaussian. The image is then modeled as the mixed distribution of the single distributions, the so-called Gaussian Mixture Model (GMM) [9]. The expectation maximization (EM) algorithm optimizes the parameters of the single distributions in an iterative process until a certain stopping-criterion is met. The result is a partitioning of the image, but again, the number of regions must be specified a-priori defining a certain level of details on a single scale. Also, the decomposition is incomplete, since not all pixels are labeled to a certain region. Furthermore, the labeled regions may be unconnected.

To achieve a hierarchical multi-scale description of image content, a region-based segmentation technique was proposed that is based on the growth of regions [10]. In the initial stage, each image pixel is considered as one region. Then, the most similar regions are merged using a clustering by region growing without seed pixels or other constraints [11]. This process is repeated until the entire image forms a single region. This approach does not require any a-priori knowledge about the given image, operates fully automatic, and further provides a hierarchical description of the image content. However, if applied to medical images, the obtained regions often do not correspond to relevant objects. In other words, the semantic gap opens. In particular, the similarity measure that is usually applied is strictly local. The feature vector of each region is calculated without evaluating the region's surrounding.

In this paper, we present a fast bottom-up region merging algorithm that is based on a flexible and powerful model for the description of regions and their similarity relations but still operates unsupervised. This algorithm is evaluated on skeletal radiographs of the hand (Fig. 1), where bony objects form reliable references for both, localization and delineation determining the identification of an object at a certain image position and its precise contouring, respectively.
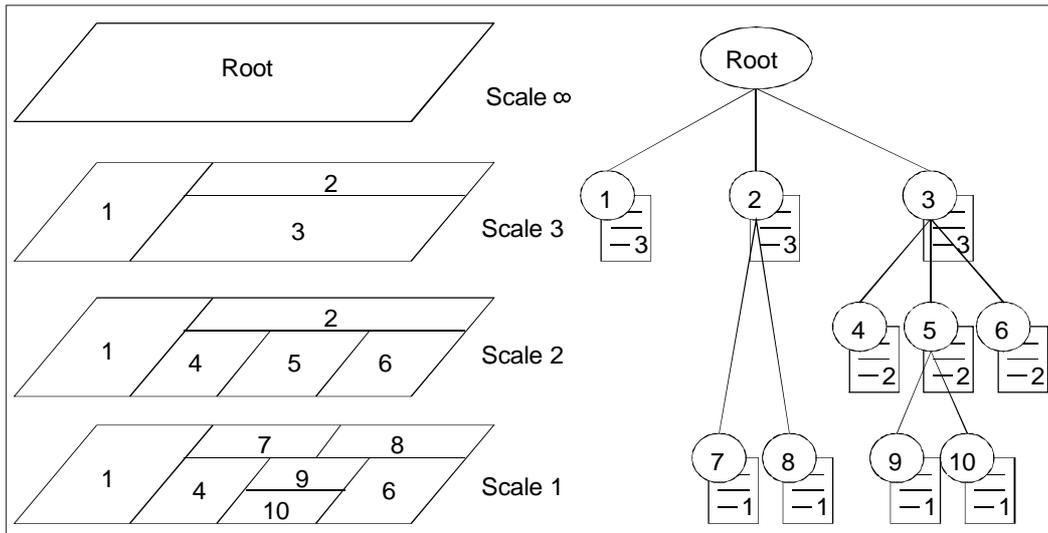
Figure 2: The HARAG is a tree-like structure recording all merges performed within the bottom-up image segmentation algorithm. The left side of this figure displays the subsequent image partitioning (referred to as scale) due to the segmentation algorithm. The right side depicts the corresponding HARAG. Each vertex (circle) is attributed with a feature vector (box), and the inter-scale inclusion relation is denoted by the edges. For a better overview, the neighborhood adjacencies have been omitted.

2. METHODS

2.1. Unsupervised Segmentation

The proposed segmentation scheme is based on bottom-up region merging [10]. The principal outline of the algorithm is fairly simple. First, an initial partitioning of the input image is produced using the morphological watershed transform [12]. The merge process is performed efficiently using the nearest neighbor graph (NNG) [13] and recorded by a hierarchical attributed region adjacency graph (HARAG) [14]. The HARAG is a tree-like data structure that reflects the adjacencies of regions as well as the inclusion relations induced by the subsequent merging process (Fig. 2).

2.2. Local Region Similarity

During the merge, a region R is described independently of its surrounding by calculating a feature vector $\mu(R)$ from the region's elements, i.e. the pixels of the particular region. Common vector entries are gray or texture values such as those proposed by HARALICK et al. [15]. The similarity of two regions is then determined by a distance measure that compares the regions' mean feature vectors. An instance of a local similarity measure is the Euclidean distance. For two regions R_a and R_b and their D-dimensional feature vectors $\mu(R_a)$ and $\mu(R_b)$, it is defined as:

$$d(R_a, R_b) = \sqrt{\sum_{i=1}^{D} \left(\mu_i(R_a) - \mu_i(R_b)\right)^2} \tag{1}$$

The local view constitutes the commonly used way of describing regions and modeling their similarity. Features are computed for each region individually by taking into account only pixels of the particular region. Figure 3 (upper row) emphasizes the problems related to this strictly local approach:

1. Segmented regions smear over object boundaries. This results from the highly inhomogeneous gray values within the meaningful objects that are represented in equation (1) by a mean value, only.
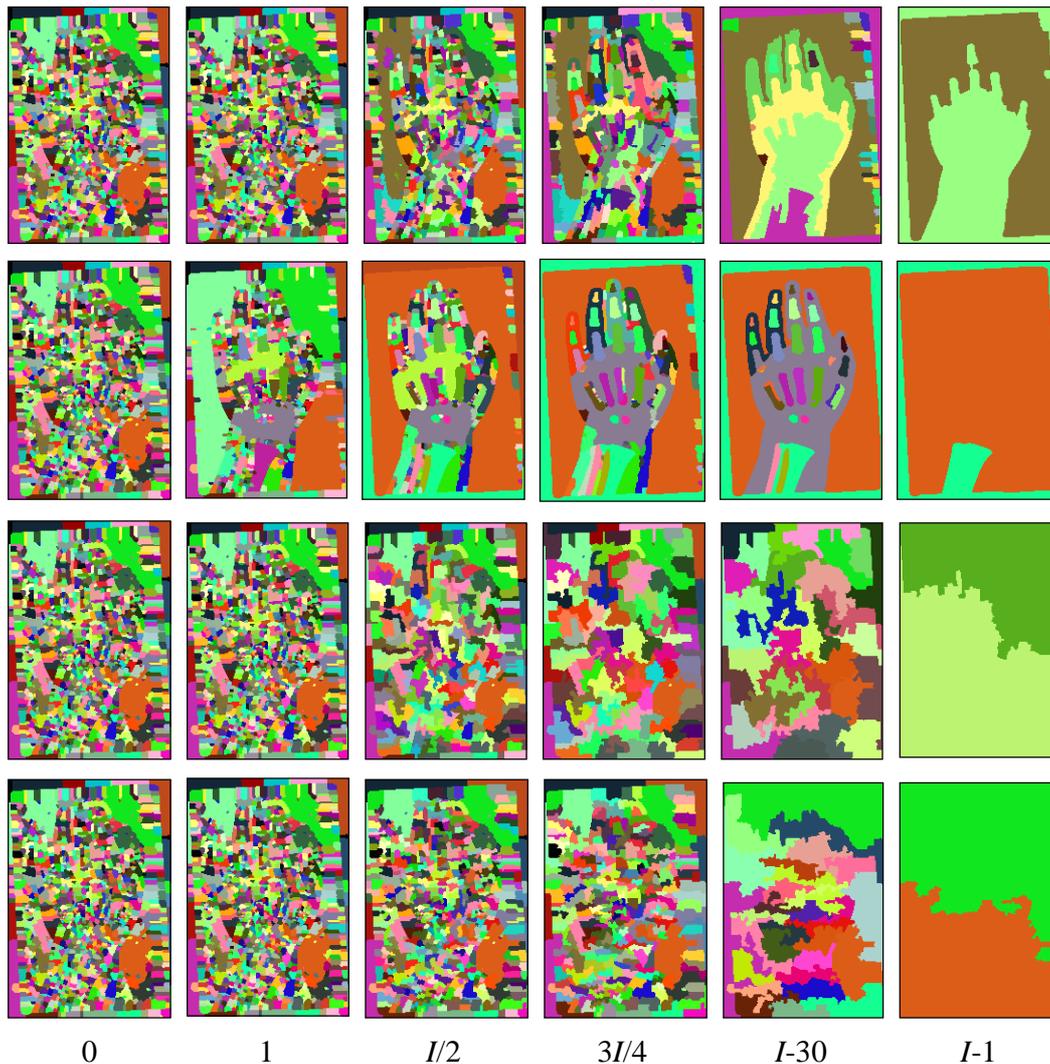
Figure 3: Multi-scale partitioning of Figure 1c using only local (top row), regional, global, or hierarchical (bottom row) distance measures, where I determines the number of iterations.

2. In all scales, very small but also very large regions co-exist. Since the multi-scale decomposition is used to reflect different levels of details, a rather homogeneous distribution of region sizes is required in each of the levels.

3. Irregular shapes are constructed during the merge. Since the HARAG acknowledges only the fact that two merge candidates are neighbored, other important parameters such as the length of connecting border or its regularity remain unconcerned.

4. Some regions are maintained over a large number of iterations (e.g., in Fig. 3 top row, the green rectangular-shaped artifact in the very top left part of the image is visible from level 0 up to I-30, with I denoting the total number of iterations, which is about several thousands. This results from handling each scale individually.

2.3. Extended Region Similarity

To extend the strictly local nature of image processing, four views on the similarity of regions are defined, i.e., the local, regional, global, and hierarchical view, which are denoted by l, r, g, and h, respectively. To satisfy a certain task of image segmentation, several components of these categories are combined:

$$d_C(R_a, R_b) = \prod_{n=1}^{N} d_n(R_a, R_b)^{w_n} \qquad (2)$$

Here, N denotes the number of single similarity measures, which is usually larger than four, and w_n the weight for similarity measure $d_n(R_a, R_b)$. This measure provides a more flexible and powerful description of region similarities than traditional approaches that are based on the calculation of mean feature vectors.

Local view: The local view of regions based on the Euclidian distance is defined in equation (1). Since the features in use may be highly correlated, the Mahalanobis distance is used

$$d_l(R_a, R_b) = [\mu_i(R_a) - \mu_i(R_b)]^T K^{-1} [\mu_i(R_a) - \mu_i(R_b)] \qquad (3)$$

where K denotes the covariance matrix of the distribution of the features.

Figure 3 (top row) shows the result of partitioning Figure 1c using only the local view on image similarity. The related problems have been addressed already in Section 2.2. Neither hard or soft tissue structures form individual objects at any of the levels. Obviously, these results are not suitable for further image analysis.

Regional view: Pairs of adjacent regions and their mutual dependencies such as feature value distributions at their common border are considered when regional similarity measures are calculated. An instance of a regional measure accumulates the values returned by a Canny edge filter along the common border of adjacent regions:

$$d_r(R_a, R_b) = \frac{1}{|B_{a,b}|} \sum_{(p_a, p_b) \in B_{a,b}} \max\{e(p_a), e(p_b)\} \qquad (4)$$

Here, $e(p)$, $B_{a,b} = \{ (p_a, p_b) \in R_a \times R_b \mid p_a \in n_4(p_b) \}$ and $n_4(p)$ denotes the binary edge value returned by the edge filter for image pixel p, the common border of regions R_a and R_b, and the four-neighborhood of p, respectively. The contrast along a common border forms another regional measure that is calculated directly from the border pixels' intensities, and does not require any pre-processing of the image, e.g. computing the Canny edge filter or its Laplacian. Let $g(p)$ denote the gray value at position p, the contrast along the common border is then defined as:

$$d_r'(R_a, R_b) = \frac{1}{|B_{a,b}|} \sum_{(p_a, p_b) \in B_{a,b}} [g(p_a) - g(p_b)]^2 \qquad (5)$$

Another regional measure has been proposed by PAVLIDIS & LIOW, where regions after a split and merge are rejected if their contours do not correspond to edges [16].

The benefit from regional distance measures is exemplified in Figure 3 (second row). Based on a binary edge image that is morphologically dilated to an edge thickness of three pixels, merging neighbored regions that are separated by an image border is omitted. For instance at scale I-30, hard as well as soft tissue form separated objects, since the contours are clearly present in the original radiograph (Fig. 1c).

Global view: The global view considers the entire image to define a similarity of two adjacent regions. For instance, the segmentation process can be guided to result in regions of similar size by the global view, e.g. evaluating the variance of all region sizes before and after a hypothetical merge. To avoid exhaustive computations, a partitioning with similar region sizes is achieved implicitly through the minimization process performed in each algorithm iteration. An instance of a global measure is then simply defined by the sum of two regions' sizes:

$$d_g(R_a, R_b) = |R_a| + |R_b| \qquad (6)$$

Nonetheless, the global nature of this view is maintained. Another global view is provided by the Ward criterion [17], which is actually not a distance measure but a decision rule that minimizes the global squared error for a partitioning. The central idea of the Ward criterion is based on error minimization. That pair of regions is merged whose unification leads to the smallest change in the global error of the subsequent partitioning.

The third row of Figure 3 shows the partitioning obtained when selecting $d_C = d_g$ from equation (6). Note that neither gray nor textures values of image pixels guided this segmentation. Consequently, the regions formed in all scales $I > 0$

do not correspond to any relevant image objects. However, the size of regions grows homogenously. The higher the scale, the larger the regions.

Hierarchical view: The inter-scale relations between regions is considered for the hierarchical view on the segmentation process. In particular, the algorithm is forced to a homogeneous distribution of the iteration age of all regions. For instance, this is accomplished by defining a hierarchical distance measure as:

$$d_h(R_a, R_b) = l(R_a) + l(R_b) + 1 \tag{7}$$

where $l(R)$ denotes the creation level of region R, i.e. the iteration I in which the region was created. For instance, regions created in the initialization phase have creation level 0. Due to the minimization process performed in each iteration, region pairs with small creation levels are merged in early stages of the segmentation process resulting in a scale-consistent region growth. Again, this implicit definition avoids exhaustive computations.

Figure 3 (bottom row) visualizes the effect of hierarchical components of the distance measure. Here, $d_C = d_h$. Again, merging is not guided by the image pixel information and, consequently, the regions do not correspond to meaningful objects. However, merges take place all over the image. For instance, the small artificial regions at the right collimation border, which are present in all image-guided examples (Fig. 3, first and second row) disappear earlier. Note that this example has been computed without calculating the transitive closure for regions adjacencies. Hence, the line-wise processing order from top to bottom becomes visible in the rather horizontal orientation of the regions.

2.4. Evaluation

The bottom-up region merging approach using a combined distance is used for content-based image retrieval in medical applications (IRMA). Here, the images are automatically classified according to their imaging modality, direction to the patient, examined body region, and imaged biological system before local analysis is performed [18]. This initial categorization is based on global features, i.e. a single vector describing the entire image, and allows to design task-specific combinations for the distance d_C.

The image category selected for experimental evaluation is that of a plain radiograph (modality) acquired in posteroanterior projection (direction) of the musculoskeletal system (biosystem) of the left hand (body region). These images are routinely acquired to determine the skeletal age of infants. Therefore, the objective of segmentation is to extract all bones, i.e. 14 phalanges, 5 metacarpal bones, up to 8 carpal bones (depending on the age of the infant), ulna, and radius (Fig. 1). Although these bones are obviously visible for human observers, automatic segmentation is extremely difficult due to the gradients of gray scales resulting from the various thickness of hard and soft tissues absorbing the x-rays [19].

A gold standard or ground truth is required to analyze the quality of segmentation quantitatively. In particular, the delineation of objects is analyzed. Since single manual references are insufficiently reproducible, a set of ten manual segmentations has been provided by human observers. Currently, 407 hand radiographs are contained in the IRMA reference database, where ten references have been selected arbitrarily. Altogether, 100 manual segmentations were obtained. A reliable gold standard was computed summing each of the ten binary references and thresholding them again at gray level seven. In other words, a pixel is regarded to display bone if at least seven observers have labeled it accordingly.

The HARAGs were computed automatically with a weighting of 0.5, 30, 1, and 1 for local, regional, global and hierarchical components of the distance measure d_C respectively. The regional view is combined from equations (4) and (5) with $w(d_r) = 20$ and $w(d_r') = 10$. For each bony object, that node of the highest level was preferably selected with all region pixels laying within the reference object. The error E was determined by the relative number of falsely labeled pixels and the overlap O of segmented region B and the reference objects A:

$$O = \frac{\#\{A \cap B\}}{\#\{A \cup B\}} \tag{8}$$

where $\#\{R\}$ denotes the number of pixels in region R. The intra-observer variability is compared to the error obtained using the automatic method.
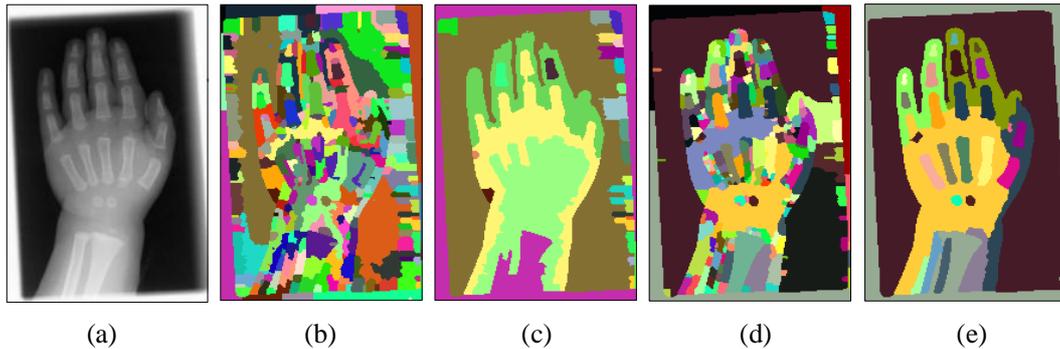
| (a) | (b) | (c) | (d) | (e) |

Figure 4: Examples of partitioning: (a) plain radiograph of a human's left hand; (b) segmentation using local similarity only (iteration $3I/4$); (c) segmentation using local similarity only (I-30); (d) segmentation combining local, regional, global, and hierarchical similarity ($3I/4$); (e) segmentation combining local, regional, global, and hierarchical similarity (I-30).

3. RESULTS

Figure 4a shows a radiograph of size 250x169 pixels. It was segmented using the Euclidean distance of mean gray values to determine the similarity of regions. Figure 4b and 4c display the partitioning obtained in iteration $3I/4$ and I-30, respectively, where I denotes the total number of iterations (iteration 0 contains the initial watershed segmentation and iteration I is the entire image represented with only one region). The segmentation fails to separate bone from soft tissue. Replacing the Euclidean distance with the Ward criterion [17] leads to similar poor results. Figure 1d and 1e show the corresponding iteration levels using the combined distance measure, which is computed on a 2.6GHz Pentium PC in 2.5 seconds. Most bones are extracted accurately.

The results of quantitative analysis are visualized in Figure 5. The left column exemplifies the variety of manual references and the ground truth calculation. The other rows of Figure 5 show the results of the automatic segmentation using a combined distance measure and selecting the appropriate scale for each of the bony components. Images 7 and 4 were selected for small errors while images 6 and 10 represent a large amount of misclassified pixels.

Table 1 comprehensively shows the quantitative results obtained for all radiographs. Minimum, mean and maximum error is 4.4 %, 15.7 %, and 26.0 %, respectively. Since the minimum, mean, and maximum error of manual reference segmentation is 6.3 %, 12.3 %, and 22.9 %, respectively, the automatic segmentation is of similar precision as the manual.

Frequently, the similarity of two regions is measured in terms of the overlap O. The mean correspondence of ground truth and automatic segmentation is 85.1 %. In comparison to the relative error E, results in terms of overlap O are slightly better. However. this is due to the different normalization in equation (8).

4. DISCUSSION

The description of regions and their similarity is crucial for the segmentation results of region merging approaches. Traditionally, regions are described by a mean feature vector computed from all region pixels, and the similarity of two regions is determined straight-forwardly by a simple distance measure such as the Euclidean or the Mahalanobis distance. However, this model is insufficient to cope with the large variety of structures in the medical imagery. Significant improvement is obtained using a more flexible and powerful similarity model which is integrated into a bottom-up region merging approach that operates fully automated and processes average-sized images within a few seconds.

The selected application only exemplifies the potential of this approach for various tasks in medical image segmentation. Taking a closer look at the original image reveals that the bones are distinguished by sharp edges, but hold an internal gradient in their gray values. In addition, the surrounding tissue is displayed with a gray value intensity
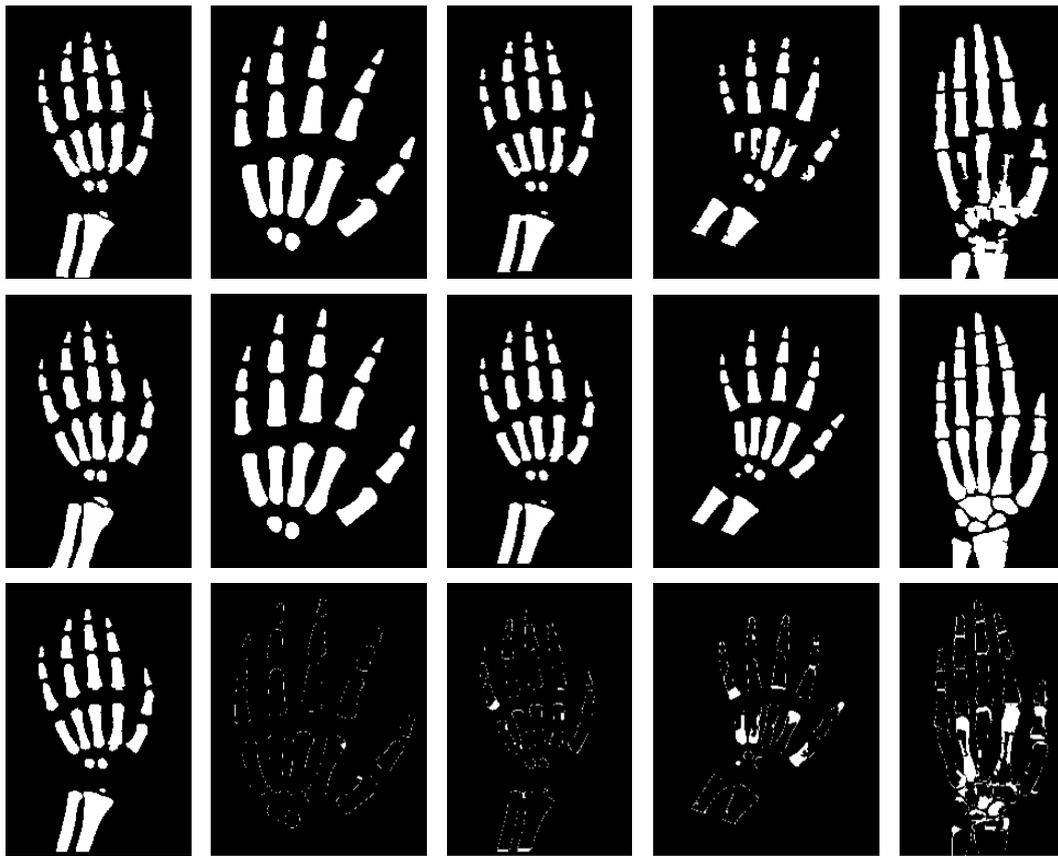
Figure 5: Results: (left column) two manual reference segmentations (top, middle row) and the gold standard for image 04 (bottom row); (columns 2 to 5) automated segmentation (top), gold standard (middle), and error pixels (bottom) for images 7, 4, 6, and 10 visualizing the smallest and largest errors.

similar to the one of the bones. Thus, the description of the image content is modeled more accurately using the combined measure that emphasizes an edge-based component.

Colums 2 and 3 in Figure 5 emphasize that errors of 10 % or less are rather invisible when visually inspecting the automatic segmentations. For images resulting in errors of 20 % or more, the discrepancies in delineations are obviously notable. Here, the segmentation algorithm fails to delineate a small number of objects such as the carpal or metacarpal bones, while the phalanges are delineated correctly. The problems occur either by very young patients, where skeletal mass is very low (Fig. 5, Column 4), or for adults, where the carpal bones do not appear separately in the x-ray images (Fig. 5, Column 5). Regarding an application for skeletal maturity assessment based on this general segmentation approach, further tests are required to analyze the actual error in age determination.

For applications in content-based image retrieval, additional features are computed from the extracted regions [20]. Here, application-specific properties can be defined for size, shape, position, and relative position of relevant objects. Content-based access to images is then provided formulating an interval query on the regions attributes. In contrast to computer-assisted diagnoses, content-based image retrieval is looped interactively by query refinement. Hence, erroneous delineations of individual bones or other relevant object can be tolerated.

Nr	#{A}	#{ A ∪ B } − #{ A ∩ B }	E	#{ A ∪ B }	#{ A ∩ B }	O
01	5657	826	14.6 %	5925	5099	86.0 %
02	7671	1477	19.3 %	8195	6718	82.0 %
03	6080	619	10.2 %	6368	5749	90.3 %
04	5340	483	9.0 %	5551	5068	91.3 %
05	7735	1037	13.4 %	7961	6923	87.0 %
06	4790	1060	22.1 %	5098	4038	79.2 %
07	8600	378	4.4 %	8800	8422	95.7 %
08	7054	1420	20.1 %	7437	6017	80.9 %
09	7169	1273	17.7 %	7588	6315	83.2 %
10	9373	2435	26.0 %	9879	7444	75.4 %
mean			15.7 %			85.1 %

Table 1: Results.

5. CONCLUSION

So far, the bottom-up region merging method has been limited to a simple model of region similarity based on the mean feature vector. Applied to medical images, this yields poor results since anatomic structures are segmented incorrectly. In this paper, a flexible and powerful approach to model the similarities of image regions is presented and integrated into a region merging algorithm. Extending the local view to regional, global, and hierarchical components of a combined similarity measure significantly improves segmentation results, i.e. extracting more objects of interest (localization) in a better quality (delineation). Therefore, unsupervised segmentation becomes applicable for medical imagery showing a large variety of relevant structures. In particular for content-based image retrieval, where the desired structure is defined by the user at the time of the query, hierarchical partitioning can now be computed already at time of data entry.

ACKNOWLEDGEMENT

REFERENCES

1. Lehmann TM, Meinzer HP, Tolxdorff T: Advances in biomedical image analysis. Past, present and future challenges. Methods of Information in Medicine 2004; 43(4): 308-14.
2. Haralick RM, Shapiro LG: Image segmentation techniques. Computerized Visualization and Graphical Image Processing 1985; 29: 100-32.
3. Adams R, Bischof L: Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 1994; 16(6): 641-7.
4. Mehnert AJH, Jackway PT: An improved seeded region growing algorithm; Pattern Recognition Letters 1997; 18(10): 1065-71.
5. Hojjatoleslami SA, Kittler J: Region growing: A new approach. IEEE Transactions on Image Processing 1998; 7(7): 1079-84.
6. Stewart RD, Fermin I, Opper M: Region growing with pulse-coupled neural networks: an alternative to seeded region growing. IEEE Transactions on Neural Networks 2002; 13(6): 1557-62.
7. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000; 22(12): 1349-80.
8. Belongie S, Carson C, Greenspan H, Malik J: Recognition of images in large databases using color and texture. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002; 24(8): 1026-38.
9. Greenspan H, Goldberger J, Ridel L: A continuous probabilistic framework for image matching. Journal of Computer Vision and Image Understanding 2001; 84(3): 384-406.

10. 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 2003; 5032(1): 598-608.
11. Tilton JC; Method for recursive hierarchical segmentation by region growing and spectral clustering with a natural convergence criterion. Disclosure of Invention and New Technology: NASA Case Number GSC 14,328-1, 2000.
12. Vincent L, Soille P: Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 1991; 13(6): 583-98.
13. Haris K, Estradiadis SN, Maglaveras N, Katsaggelos AK: Hybrid image segmentation using watersheds and fast region merging. IEEE Transactions on Image Processing 1998; 7(12): 1684-99.
14. Sonka M, Hlavac V, Boyle R: Image Processing, Analysis and Machine Vision. Chapman & Hall, London, 1993.
15. Haralick RM, Shanmugam K, Dinstein I: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 1973; 3(6): 610-21.
16. Pavlidis T, Liow YT: Integrating Region Growing and Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(3): 225-233, 1990.
17. Ward JH: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(1): 236-245, 1963.
18. 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 43(4): 354-361, 2004.
19. Aja-Fernandez S, de Luis-Garcia R, Martin-Fernandez MA, Alberola-Lopez C: A computational TW3 classifier for skeletal maturity assessment. A computing with words approach. Journal of Biomedical Informatics 2004; 37(2): 99-107.
20. Thies C, Ostwalt T, Fischer B, Lehmann TM: Object-based modeling, identification and labeling of medical images for content-based retrieval by querying on intervals of attribute values. Procs SPIE 2005; 5748, in press.