Segmentation Of Bone And Soft Tissue Regions In Digital Radiographic Images Of Extremities

S. Kubilay Pakin[a], Roger S. Gaborski[b], Lori L. Barski[c], Dave H. Foos[d], and Kevin J. Parker[e]

[a]University of Rochester, Electrical and Computer Engineering,
204 Hopeman Bldg., Rochester NY 14627 USA
[b]Rochester Institute of Technology, Rochester NY USA
[c,d]Eastman Kodak Co., Rochester NY USA
[e]University of Rochester, Electrical and Computer Engineering, Rochester NY USA

ABSTRACT

This paper presents an algorithm for segmentation of computed radiography (CR) images of extremities into bone and soft tissue regions. The algorithm is a region-based one in which the regions are constructed using a growing procedure with two different statistical tests. Following the growing process, tissue classification procedure is employed. The purpose of the classification is to label each region as either bone or soft tissue. This binary classification goal is achieved by using a voting procedure that consists of clustering of regions in each neighborhood system into two classes. The voting procedure provides a crucial compromise between local and global analysis of the image, which is necessary due to strong exposure variations seen on the imaging plate. Also, the existence of regions whose size is large enough such that exposure variations can be observed through them makes it necessary to use overlapping blocks during the classification. After the classification step, resulting bone and soft tissue regions are refined by fitting a 2^{nd} order surface to each tissue, and reevaluating the label of each region according to the distance between the region and surfaces. The performance of the algorithm is tested on a variety of extremity images using manually segmented images as "gold standard". The experiments showed that our algorithm provided a bone boundary with an average area overlap of 90% compared to the gold standard.

Keywords: Segmentation, region growing, voting, surface fitting

1. INTRODUCTION

Accurate diagnosis of disease states in medical radiographic images often depends on the detection of small, low contrast details in the image. The visibility of these details is dependent of the film's tonescale curve. The tonescale response curve of traditional silver halide films is determined by the film's built-in chemical emulsion and the development process. The manufacturer attempts to optimize the film for a range of exposure techniques and exam types. Digital radiographic imaging systems record raw image data without the built in tonescale curve and with greater dynamic range. The increased dynamic range of the digital image along with digital image processing techniques allow for manipulation of the image before it is viewed by the radiologist. Segmentation of the ROI (region of interest) in the code value histogram is an important first step in performing image enhancement processing. Once identified, the range of code values corresponding to the ROI can be optimally mapped, via a human brightness model, to the film density response of a laser printer or to the luminance response of a CRT for visual interpretation. Robust algorithms exist for segmentation of direct exposure and collimation regions in radiographic imagery.

Previous methods have focused on segmenting the code value histogram based on knowledge of the body part being imaged and the characteristics of the corresponding histogram.[1] For a particular body part the histogram contains certain features (peaks, valleys, etc.) that can be used to map code values to anatomical features. This technique lacks robustness in certain exam types and exposure conditions. Another approach focused on segmenting

Further author information: (Send correspondence to S.K.P.)
S.K.P.: E-mail: pakin@ece.rochester.edu
R.S.G.: E-mail: rsg@cs.rit.edu
L.L.B.: E-mail: lori.barski@kodak.com
D.H.F.: E-mail: dfoos@kodak.com
K.J.P.: E-mail: parker@ece.rochester.edu

Medical Imaging 2001: Image Processing, Milan Sonka, Kenneth M. Hanson, Editors,
Proceedings of SPIE Vol. 4322 (2001) © 2001 SPIE · 1605-7422/01/$15.00

the image into bone and tissue regions based on a neural network classifier using texture measures as features.[2] The time required to obtain the texture features makes the approach impractical. The definition of an robust and effective method to refine the estimate of the ROI by segmenting the bone and soft tissue components of the image has remained elusive.

The algorithm proposed in this paper can be considered as a hybrid technique that combines the features of region-based and histogram-based techniques.[3] Histogram-based techniques follow the Bayesian approach for estimation of the label field by modeling the image as a sample from a Gaussian distributed random field. Also, the intensity values of the pixels are assumed to be independent random variables whose parameters of their density function (mean and variance values since they are Gaussian distributed) depend on the label of the pixel. Assuming equally likely labels for every pixel results in well-known K-means algorithm.[4] There are also numerous papers in the literature which impose spatial constraints by modeling the label field as a Gibbs or Markov random field.[5–7] In general, these algorithms provide better results than K-means algorithm but they are also known for their significant computational burden.

The main problem associated with the usage of histogram-based techniques in the context of the segmentation of CR images is that CR images do not follow the image model that histogram-based techniques rely on due to the exposure variations. Our algorithm attempts to address this problem in several steps. First, the image is divided into regions in which the pixels share certain statistical similarities. The feature extraction and region growing stages that perform this task will be covered in Section 2. After region growing stage, each region is labeled as bone or soft tissue. The voting procedure, which performs the labeling task, consists of clustering of regions using K-means algorithm with 2 classes. Note that performing region-based instead of pixel-based clustering effectively introduces spatial constraints to the label field. The effect of exposure variations is compensated by processing the image locally using the neighborhood systems and sub-images. The voting procedure will be explained in detail in Section 3. The last stage of the algorithm refines the labels of the regions by fitting a 2^{nd} order bivariate polynomial to each tissue and investigating the average distance between each region and the surface of the label that they belong. The details of this stage is provided in Section 4. Finally, the experiments and the conclusions are stated in Sections 5 and 6, respectively.

2. CONSTRUCTION OF REGIONS

2.1. Feature Extraction

In the first stage of the algorithm, two features are computed for each pixel. The features are denoted as μ_i and σ_i, where i denotes the index of the pixel site, and they reflect the local characteristics around the pixel that they belong.

- μ_i : The median value of the intensities of the pixels that are inside the neighborhood of pixel $\mathbf{x}_i$ assuming 8-connectivity.[4] The region growing algorithm depends on local differences which are suppressed to some extent in sample mean computation. Because of this fact, median value is used instead of sample mean to avoid the smoothing on region boundaries. More sophisticated schemes for feature extraction such as iterative methods that preserve line structures have been avoided due to their high computational cost.[8]
- σ_i : The standard deviation of the intensities of the pixels that are inside the neighborhood of pixel $\mathbf{x}_i$ assuming 8-connectivity.

2.2. Region Growing

The two feature values computed in the previous step are utilized for investigation of the statistical similarity of each adjacent pixel pair where adjacency is defined in an 8-connectivity sense. For the pixel pair $(\mathbf{x}_i, \mathbf{x}_j)$, the following statistical tests are utilized to determine whether they are sufficiently similar to be in the same region.

- *F-test :* The pixels $(\mathbf{x}_i, \mathbf{x}_j)$ and their corresponding neighbors are assumed to be the realizations of the random variables whose distributions are $N(\eta_i, \nu_i^2)$, and $N(\eta_j, \nu_j^2)$, respectively. The well-known F-test, which derives from the fact that if $\nu_i = \nu_j$, then the RV σ_i^2/σ_j^2 has F-distribution with degrees of freedom $(\mathrm{K}-1, \mathrm{K}-1)$, where K is the size of both ensemble,[9] is used to investigate the similarity of the variances of the distributions in order to test the hypothesis that both ensemble are indeed governed by the same distribution. Note that,

σ_i^2 and σ_j^2 are the sample variance values and have been computed in the feature extraction step. F-test is formulation as follows :

$$\frac{\sigma_i}{\sigma_j} < \gamma_{\mathrm{F}}, \tag{1}$$

where γ_{F} is a predetermined threshold. Note that γ_{F} is always greater than 1, and the reciprocal of the Eqn 1 has to be considered if $\sigma_i < \sigma_j$.

- *Mahalanobis distance :* We exploit the interpretation of Mahalanobis distance that is used for investigating the closeness (statistical) of a point to an ensemble whose distribution is known. Thus, we end up with the following two tests in which $\mathbf{x}_i$ and $\mathbf{x}_j$ are compared to the ensemble that consists of $\mathbf{x}_j$ and $\mathbf{x}_i$ and their neighbors, respectively.

$$\frac{|\mu_i - \mu_j|}{\sigma_i} < \gamma_{\mathrm{M}} \text{ and } \frac{|\mu_i - \mu_j|}{\sigma_j} < \gamma_{\mathrm{M}} \tag{2}$$

where γ_{M} is a predetermined threshold.

For a particular adjacent pixel pair $(\mathbf{x}_i, \mathbf{x}_j)$, if all the above statistical tests are satisfied, that pixel pair is said to be connected. Following the investigation of the similarity of each adjacent pixel pair, regions that are defined as maximal set of pixels all belonging to the same connected components are formed.

During the stages that follows region growing, the algorithm basically attempts to label each region as either bone or soft tissue. Hence, it is vital to adjust the thresholds γ_{M} and γ_{F} in order to construct regions whose pixels belong to only one type of tissue, otherwise some of the pixels will inevitably be assigned incorrect tissue labels. While high threshold values will cause leaking between regions that correspond to different tissue types, too low values will result in over segmentation which deteriorates the performance of the stages that follows growing.

3. VOTING PROCEDURE

Following the region growing stage, each region is assigned a tissue label using the voting procedure. "Voting" is a broad term and it has been used under several contexts in the literature such as Hough transform related grouping techniques,[10] and majority voting schemes used for combining the decisions of several classifiers.[11] The procedure we propose assigns bone or soft tissue votes to the regions that have been constructed in the previous step using local 2-class clustering.

As mentioned in Section 1, the exposure variations seen on the image plate prevents the digital radiographic images from having a bimodal histogram. The significant overlap between the histograms of bone and soft tissue intensities makes the use of global clustering algorithms inefficient. The voting procedure is essentially governed by the following observation: Because of the relatively high x-ray absorption of bone, locally, bone pixels are darker than soft tissue pixels. Exposure variations makes this observation invalid globally.

During the voting procedure, the image to be segmented is analyzed using overlapping blocks. The following notation is introduced in order to explain the voting procedure:

- $\mathcal{R}_m$: m^{th} region.
- $\mathcal{V}_m(\mathrm{B})$: The votes for a bone label in the m^{th} region.
- $\mathcal{V}_m(\mathrm{S})$: The votes for a soft tissue label in the m^{th} region.
- $\mathcal{B}_k$: k^{th} block.
- $\mathcal{R}_{ik}$: i^{th} region in k^{th} block.
- $\mathcal{N}_{ik}$: Neighborhood system of $\mathcal{R}_{ik}$.
- $\mathcal{R}_{jik}$: j^{th} region in $\mathcal{N}_{ik}$.
- $\mathcal{L}_{jik}$: The label of $\mathcal{R}_{jik}$ as the result of clustering of the regions in $\mathcal{N}_{ik}$.

The voting procedure can be summarized as follows:

- For each $\mathcal{B}_k$
  - For each $\mathcal{R}_{ik}$
    - Compute the average intensities of $\{\mathcal{R}_{jik}\}$.
    - Cluster $\{\mathcal{R}_{jik}\}$ using k-means algorithm with 2 clusters.
    - For each $\mathcal{R}_{jik}$
      - If $\mathcal{L}_{jik}$ is bone, increment $\mathcal{V}(\mathrm{B})$ of corresponding $\mathcal{R}_m$.
      - If $\mathcal{L}_{jik}$ is soft tissue, increment $\mathcal{V}(\mathrm{S})$ of corresponding $\mathcal{R}_m$.

Since overlapping blocks are used, a particular region, $\mathcal{R}_m$, which has been constructed in growing stage, could quite probably appear in different blocks. Furthermore, the borders of blocks can pass through $\mathcal{R}_m$, thus different patches of $\mathcal{R}_m$ could appear in different blocks. Hence, according to the above formulation, two different symbols, $\mathcal{R}_{ik}$ and $\mathcal{R}_{i'k'}$, may correspond to different patches of a particular region, $\mathcal{R}_m$. Even though, another symbol for this mapping between the regions in the image and the regions in the blocks is not reserved here for the sake of convenience, it is very crucial in the implementation. The reason for this importance is that only the votes of $\mathcal{R}_m$'s are kept, but $\mathcal{R}_{ik}$'s are processed. Since a particular region, $\mathcal{R}_m$, will appear in different blocks and neighborhood systems, it will acquire several votes. After all the blocks and regions are processed the tissue labels of the regions are determined as follows:

- For each $\mathcal{R}_m$
  - If $\mathcal{V}_m(\mathrm{B}) > \mathcal{V}_m(\mathrm{S})$, $\mathcal{R}_m$ is a bone region.
  - If $\mathcal{V}_m(\mathrm{B}) < \mathcal{V}_m(\mathrm{S})$, $\mathcal{R}_m$ is a soft tissue region.
  - If $\mathcal{V}_m(\mathrm{B}) = \mathcal{V}_m(\mathrm{S})$, the label of $\mathcal{R}_m$ is undecided.

The regions whose labels are undecided are reconsidered and their labels are assigned in the next stage which will be discussed in the next section.

4. REFINEMENT OF LABELS VIA SURFACE FITTING

Although the voting procedure results in correct tissue labels for most of the regions, the probability of misclassification increases in the areas of image where regions that belong to same type of tissue densely clustered. Such dense clusters tend to occur especially in the soft tissue close to tissue boundary. Since the thickness of the tissue varies greatly in such areas, the magnitude of the image gradient is usually significant compared to other areas of the images.

Because of the reasons above, the tissue labels are reevaluated by the last stage of the algorithm in order to ameliorate the segmentation result provided by the voting procedure, and to assign tissue labels to the regions left undecided. The last stage essentially attempts to represent bone and soft tissue regions of the image with 2^{nd} order bivariate polynomials whose coefficients are computed in a LS sense. After the computation of the surfaces, the labels of the regions whose RMS fitting error to the surface that corresponds to its tissue type is significant are reversed. The regions whose labels have been left undecided at the end of the voting procedure are assigned labels during the first iteration.

The steps of the label refinement with surface fitting step can be stated as follows:

- Divide the image into 4 quadrants along the principal axes of tissue regions.
- Iterate the following steps until no label is changed.
  - Compute the surfaces in each quadrant for bone, $\mathcal{P}_{\mathrm{B},k}$, and soft tissue, $\mathcal{P}_{\mathrm{S},k}$, using LS fitting, for $k = 1 \ldots 4$.
  - For each $\mathcal{R}_m$

* Compute $\epsilon_{\mathrm{B},m} = \|\mathcal{P}_{\mathrm{B},k} - \mathcal{R}_m\|_{\mathrm{RMS}}$ and $\epsilon_{\mathrm{S},m} = \|\mathcal{P}_{\mathrm{S},k} - \mathcal{R}_m\|_{\mathrm{RMS}}$
* If the label of $\mathcal{R}_m$ is bone and $\epsilon_{\mathrm{B},m} > \epsilon_{\mathrm{S},m}$, then change the label.
* If the label of $\mathcal{R}_m$ is soft tissue and $\epsilon_{\mathrm{B},m} < \epsilon_{\mathrm{S},m}$, then change the label.

One drawback of the surface fitting algorithm is that, since it depends on the LS data fitting concept, it inherently assumes that a label might be incorrect only if the size of its corresponding region is relatively small. A big region has the ability to bend the surface to be fitted towards itself strongly due to the well-known fact that LS data-fitting is very sensitive to noise. Hence, even if a relatively big region is misclassified during the voting procedure, its RMS fitting error will turn out to be small and this will make the label correction impossible.

The surface fitting stage can also be conceptualized as a clustering scheme in which the regions are clustered into two classes using a different kind of metric than the usual ones (e.g. Mahalanobis distance) that are used in conventional clustering algorithms. From this point of view, the voting procedure provides the initial classification.

5. EXPERIMENTS

The performance of the algorithm was tested on 14 extremity images which include wrist, knee, hand, ankle and foot exams. Images that are manually segmented under the supervision of a radiologist were used as the gold standard. The original images were in three different sizes, 2500×2048, 2048×2500, and 2392×1792, with 12-bits per pixel.

For the experiments, 2^{nd} level approximation coefficients of the wavelet decomposition which exploits the wavelet filter Daubechies4 were used.[12] After the computation of approximation coefficients, the resulting images, which are essentially the versions of the originals that are acquired by filtering and downsampling by 4, are cropped in order to feed the algorithm with an image that is free of artifacts that occur towards the image boundaries. For each image used in the experiment, blocks of size 64×64 which overlap 32 pixels in all directions were used. A common misclassification case in the images in which the bone tissue is clustered towards the center (such as knee) is that, some of relatively small regions that are close to the tissue boundary can be labeled erroneously. This is due to the fact that comparing the regions close to the tissue boundary with the bone tissue surface is effectively an extrapolation. In order to correct such misclassifications, binary opening operation is applied to both bone and soft tissue regions after surface fitting stage.

In Figure 1, two different images that were used during the experiments and their corresponding segmentation results are displayed in 1^{st} and 2^{nd} row, respectively. The red regions in the images of the 2^{nd} row show the bone tissue that is detected by the algorithm. Also, a visual comparison between the result of algorithm and manual segmentation is provided in the last row. In those images, red pixels correspond to ones that were classified as bone in both cases. In addition to that, blue region indicates the pixels that were classified as bone by the algorithm but as soft tissue during manual segmentation, and vice versa for green region.

In Table 1, the summary of the experiments, which contains the type of extremities and the sizes (after cropping) of the images, and the computational time, is provided. In order to evaluate the performance of the algorithm statistically, the labeling process can be viewed as a binary hypothesis testing problem. If the events that a pixel belongs to bone, and to soft tissue are assumed as the null and alternate hypothesis, respectively, the performance of the algorithm can be evaluated by computing the ratio of correct detections to total number of decisions where being correct is defined with respect to the gold standard. In fact, this ratio is the well-known parameter "accuracy" which is equal to

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{3}$$

where (TN+TP), FN, and FP are the number of correct decisions, false alarms (green regions in last row of Figure 1), and misses (blue regions in last row of Figure 1), respectively.

The experiments were conducted on an SGI Indigo2 workstation. It can be seen from Table 1 that the algorithm provides over 90% accuracy (except an outlier image which is the case 6) with reasonable computational time. The observer variability has to be taken into consideration when the accuracy results are interpreted. For this purpose, a particular wrist image (case 5) is manually segmented 10 times and the accuracy values of those experiments are computed by assuming the original manual segmentation "gold standard". The manual segmentations resulted in an average accuracy value of 96.53% with a standard deviation of 0.174. No attempt has been made to streamline or optimize the code for more rapid completion of the analysis.

CASE	SIZE	TIME (sec)	ACCURACY.
13 (knee)	501×312	59	92.7%
18 (wrist)	381×556	68	89.9%
20 (knee)	606×416	91	92.5%
22 (knee)	526×366	78	95.5%
23 (knee)	566×341	77	96.0%
25 (foot)	441×486	79	92.8%
38 (foot)	596×441	90	92.3%
47 (knee)	596×286	66	95.1%
5 (wrist)	426×576	87	93.1%
52 (ankle)	551×486	101	95.1%
53 (hand)	451×326	49	90.7%
55 (ankle)	606×379	80	93.8%
6 (foot)	381×581	71	82.4%
9 (ankle)	321×581	74	92.5%

Table 1. Summary of the experiments

6. CONCLUSION

In this paper, we presented a novel algorithm for segmentation of CR images of extremities. The algorithm requires no supervision and performs the segmentation task with reasonable computational complexity. The experiments showed that the algorithm is robust in parameter space. In other words, the optimal parameter values, which are found experimentally, for the images that share common characteristics are close to each other. This algorithm can be useful for a variety of applications, including image enhancement and automatic recognition of x-ray exam type.

ACKNOWLEDGMENTS

We would like to thank Thomas Gaborski and Dr. Saara Totterman for their help in acquisition of manual segmentations which have been used in segmentation evaluation and observer variability studies.

REFERENCES

1. J. Capozzi and R. Schaetzing, "Method and apparatus for automatic tonescale generation in digital radiographic images," *U.S. Patent 5,164,993* , 1992.
2. L. Barski, R. Gaborski, and P. Anderson, "A neural network approach to the segmentation of digital radiographic images," in *ANNIE 93 Artificial Neural Networks in Engineering*, St. Louis, MO, 1993.
3. K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. on Image Proc.* **7**, pp. 1684–1699, 1998.
4. M. Nadler and E. P. Smith, *Pattern Recognition Engineering*, John Wiley & Sons, 1993.
5. S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution, and the bayesian restoration of images," *IEEE Trans. Patt. Anal. Machine Intell.* **PAMI-6**, pp. 721–741, 1984.
6. J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B* **48**, pp. 259–302, 1986.
7. T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans on Signal Proc.* **40**, pp. 901–914, 1992.
8. J. G. Tamez-Pena, *Four-Dimensional Reconstruction and Visualization of Complex Musculoskeletal Structures*, Ph.D. Thesis, Univ. of Rochester, 1999.
9. V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley & Sons, 1976.
10. G. L. Foresti, V. Murino, C. S. Regazzoni, and A. Trucco, "A voing-based approach for fast object recognition in underwater acoustic images," *IEEE Journal of Oceanic Eng.* **22**, pp. 57–65, 1997.
11. L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: An analysis of ots behavior and performance," *IEEE Trans. on Sys., Man, and Cyber. - Part A: Sys and Humans* **27**, pp. 553–568, 1997.
12. I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. in Pure and Applied Math.* **41**, pp. 909–996, 1988.

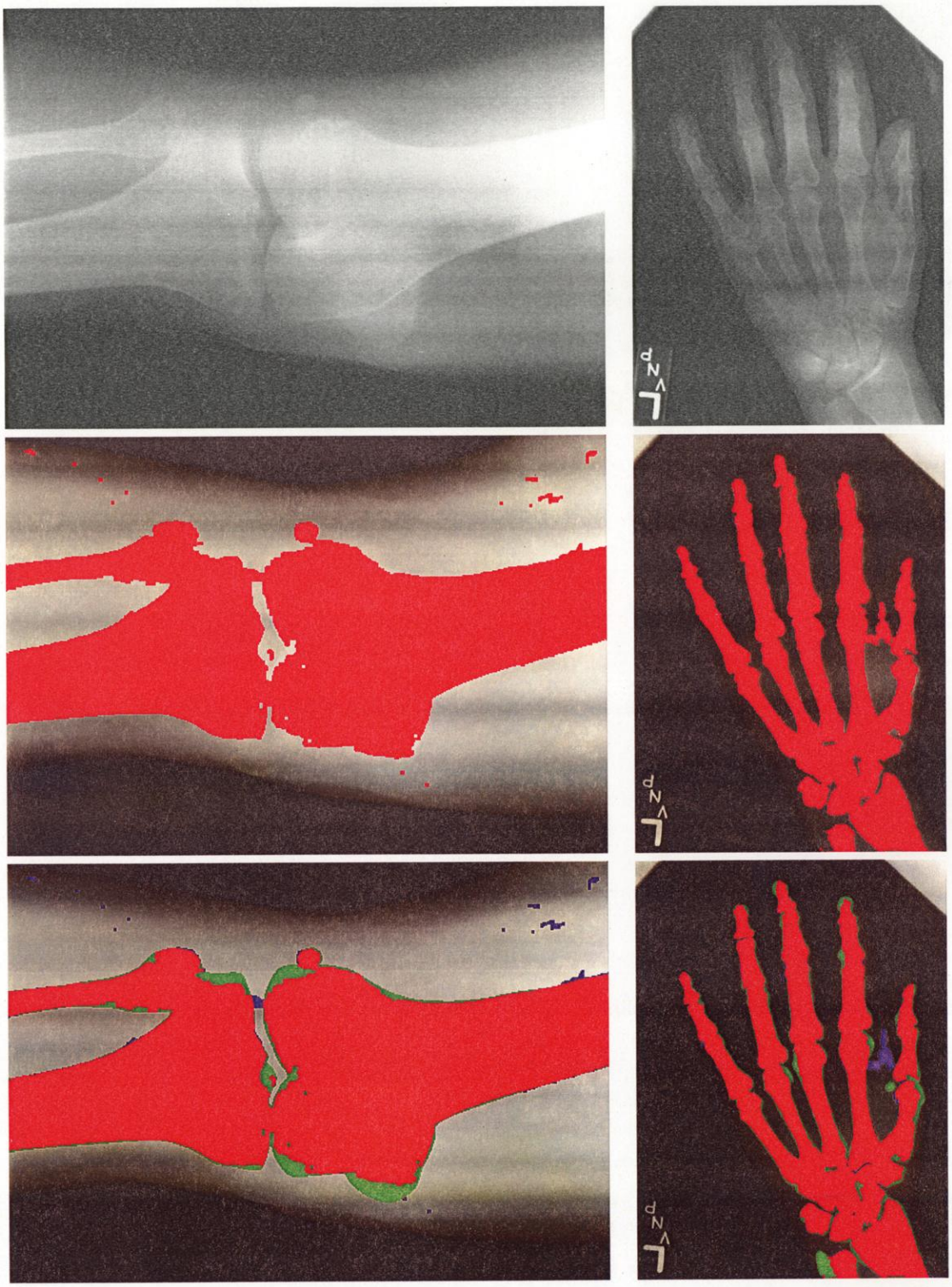

Figure 1. From top to bottom, left to right: Knee image; bone region of knee at the end of segmentation; comparison of the result of the algorithm with manual segmentation; wrist image; bone region of wrist at the end of segmentation; comparison of the result of the algorithm with manual segmentation.