INTELLIGENT RETRIEVAL
Open Journal

Review

Intelligent Shape Feature Extraction and Indexing for Efficient Content-Based Medical Image Retrieval

Phillip A. Mlsna*; **Nikolay M. Sirakov**

Department of Electrical Engineering, Department of Mathematics and Statistics Northern Arizona University, Flagstaff, AZ 86011, USA

****Corresponding author***
Phillip A. Mlsna
Department of Electrical Engineering, Department of Mathematics and Statistics Northern Arizona University, Flagstaff, AZ 86011, USA

Article information
Received: November 19th, 2023; **Revised:** January 16th, 2024; **Accepted:** January 30th, 2024; **Published:** February 26th, 2024

Cite this article
Mlsna PA, Sirakov NM. Intelligent shape feature extraction and indexing for efficient content-based medical image retrieval. 2024; 2(1).
doi: https://doi.org/10.70705/ppp.ir.2024.v02.i01.pp15-17

ABSTRACT

We detail the process of creating an efficient and innovative system for retrieving medical images based on their content, one that can extract and index crucial information like the morphology of regions. We begin with a review of the system's architecture and its primary parts. We have investigated a quick active contour method based on the geometric heat differential equation for grayscale segmentation to detect areas. A collection of shape-based characteristics is used for region representation. We use a method for organizing features using -dimensional feature vectors. In order to get images, the procedure compares query vectors to indexed feature vectors for similarity. We arrange the feature index using a convex hull model that uses the heat differential equation to decrease the search space. Certain parts of our strategy have been tested and validated via the execution of experiments. Lastly, the computational complexity of the system, along with its benefits and drawbacks, are examined.

INTRODUCTION

Content-Based picture Retrieval (CBIR) approaches, which use automatically derived characteristics like color, texture, and form as search criteria [9], have garnered a lot of attention in response to issues with conventional picture indexing methods [5]. Interest in the CBIR issue is on the rise in medical applications like medical records systems, and it is a difficult area of current study [6]. The capacity to search for photos with certain attributes using non-text information would provide a strong new dimension to medical photography. The most prevalent types of images included in medical image databases are X-ray, MRI, CT, and ultrasound, however many more modalities are also included. To be useful, CBIR approaches for big picture databases need lightning-fast query and retrieval times. Our study has shown that doing focuses on efficiently searching for relevant traits by extracting and indexing them.

Among the various illnesses recognized by medical professionals, shape information is among the most crucial and effective criteria [8, 13, 14]. We are focusing on developing new and efficient methods for extracting and intelligently indexing important information regarding radius shape in order to construct a broad picture database capabilities. It is assumed in this article that the photos are first treated to maximize contrast and eliminate noise. Next, we need to extract the regions that are of interest. We provide a quick active contour technique that relies on the geometric heat differential equation, but there are other ways to segment images [10, 14, 15]. Subsequently, a collection of extracted features including boundary support, convex hull, number of concavities and their geometric properties, boundary extrema, and number of boundary segments are used to characterize the region's borders. In order to retrieve images, the algorithm uses a similarity measure that is specific to the form characteristics that were extracted. Combining this shape data with others, such grayscale or texture, allows for more complex searches, which in turn makes the retrieval process more broadly useful. Lastly, we detail the computational complexity of our strategy, along with its benefits and drawbacks.

2 General System

A content-based image retrieval system consists of two main functions: 1) extracting and indexing feature informa- tion from the images and 2) searching the index to retrieve images whose features are most similar to the query. To be meaningful, an indexing system should address the fol- lowing steps [1]: segmentation; feature extraction, repre- sentation and classification; and feature indexing. Figure 1 illustrates the major steps involved in such a system, both in preparing the index of feature vectors and in querying this index to determine which images most closely match a given query. Features may be specified either directly or extracted from a provided query

image.
The retrieval system should be capable of efficiently comparing the query features with the set of indexed fea- tures contained in the image database [11]. For this purpose, an appropriate similarity metric is needed. The choice of data structure is key to searching the index very efficiently. In very large databases, it is important to avoid searching the entire index. Hence, techniques for reducing the search space are quite valuable.

3 Segmentation and Convex Hull Models

To address the segmentation problem, we employ a novel active contour model based on the geometric heat differen- tial equation (GHDE):

In Eq. 1, is a convex closed curve, is the curvature, and is a normal vector. This model leads to an algorithm used
to guide the evolution of the curve to capture the image
region:
normal vectors, , in the domain of , with the origin assumed at the center of the image.
To distinguish multiple regions, the segmentation pro- cess in Eq. 2 is applied repeatedly. The first stage results in one or more "shells" as illustrated in Figs. 2-3. Each shell may contain one or more distinct regions. If a shell repre- sents multiple regions, these will generally subdivide until each shell contains a single region. Finally, the algorithm is run on each shell to determine the complete boundary of its region. Figures 4-7 show an example of this process.
In order to extract the convex hull of a region, we have developed a model that employs Eq. 2 in the form

In Eq. 2, is the step for time in the -th iteration and
. Curve evolution is stopped at the region boundary by the penalty function , which is designed for grayscale images. The algorithm uses unit where is a vector designed to stop curve evolution in concert with when the convex hull is reached. Figures 8- 10 show an example of this.
We have tested these algorithms using a tool based on tail to differentiate between little variations in form. As a result, a huge collection of medical pictures could be generated from anatomical forms that are quite similar. It is possible to use local form characteristics to reduce the collection of retrieved photos.
Among the many global picture properties that our system collects and uses for shape representation, region matching [7], and image indexing [1, 2], are the image's re-gions and the distances between related regions in other images. The number of boundary segments, boundary extrema, concave segment count, segment geometric properties, convex hull, and overall boundary support are all local shape attributes that may be used to filter the pictures that are returned.
The consistency [12], a kind of directional chain code, is another helpful characteristic for representing area boundaries. You can compare the four finite numerical sequences that each consistency creates throughout the query procedure quickly.
Such a representation is invariant to scaling and translation but not to rotation and boundary origin point. To overcome these disadvantages, we first determine the boundary sup- port representation. The consistency of this support curve is invariant to rotation and choice of origin since rotation of the region corresponds to translation in support space. Polygon approximation or curve evolution is used to elimi- nate the insignificant boundary points.
Yet another promising feature is the control point set produced by closed B-spline approximation of the region boundary [3, 4]. This tends to produce relatively compact representation of a region's major shape features. To a great extent, region shapes can be effectively compared using the B-spline control point vectors. Our work is progressing to- ward a more generalized shape similarity metric based on control points combined with other attributes.

5 Feature Index Organization for Retrieval

The feature organization strategy is strongly dependent upon the feature vector used, the query types supported, and the image semantics. Thus we create a feature vector to the Medical Image Database (MID), where is the th feature of the th area. An interval-defined vector feature space is represented by all the vectors for. By applying this idea to the description of regions using features, we may express each area as a vector with dimensions in the -dimensional space. There is a point in -dimensional vector feature space that each area represents. Additionally, a -D ellipsoid has its center at this location. Therefore, we utilize the following concept as a similarity measure: two regions are considered comparable if and only if their respective ellipsoids are contained inside each other [11].
Our method for calculating the degree of form similarity between picture areas taken from different sources[11,12] using the -D feature vector is based on the aforementioned theoretical ideas. As seen in Figure 1, MID queries may be created either by directly specifying the feature vector or by using an example medical query picture. In order to retrieve results for the query region(s), the retrieval system computes and extracts the features of the -D feature vector. In order to get the response to a query, the weighted query feature vector is compared to the feature vectors of the database regions using the Euclidean distance in -D feature space [11]. The system will return MID photos that are closest to the query feature vector in terms of weighted feature vector distance.

Reliability in query response time is critical for large biological picture databases. To do this, it is recommended to eliminate unnecessary comparisons of feature vectors in order to decrease the size of the index search space that a certain query must traverse. Our approach achieves this by dividing the search space according to the local shape properties of each region's convex hull and concave border segments. In order to quickly get the convex hull of any area, we use Equation 3. After that, all of the collected areas are sorted into 344 distinct convex hull types using consistencies [11, 12] and three axes of observation. Depending on the perspective used, the convex hull of a certain area might belong to one of many classes. Additionally, the quantity of concave border segments allows for further subdivision within each class of convex hulls. The boundary support function is used to calculate the later number [11]. Hence, the query procedure restricts itself to convex hull classes and concavity subclasses that match. with the question. The total search speed is increased by a factor of several hundred using this strategy.

6 Discussion and Conclusion

The segmentation and convex hull definition tools have been implemented and tested using Mathematica, while the support definition and part of shape representation and matching tools have been developed in C++.

The image segmentation approach of Eqs. 2-3 requires approximately arithmetic operations, wherede- notes the maximum number of regions in an image and is the number of normal vectors used. Assuming a very large image with a flow with 1000 vectors and 100 distinct re- gions, the algorithm will require approximately operations. Consequently, only a very short time would be required to segment this image using a C++ tool running on a modern 2.4 GHz personal computer. While this segmen- tation approach is quite fast, it can be somewhat sensitive to both noise and poor contrast. Ongoing work aims toward improvement in these areas.

Our search space partitioning and indexing approach provides a clear advantage over those described by An- tani [1]. Partitioning by convex hull type and concavity count produces a major reduction in the number of indexed regions needed to compare to the query region. We are con- tinuing to characterize and explore this potential.

Shape features used for content-based image retrieval present some interesting challenges and opportunities. A wide variety of possible features can be applied in novel ways to provide improved query accuracy and speed. This paper has described the recent progress made by our team in the segmentation, feature extraction, and indexing areas intended for medical image database applications. Further progress targets several areas: improvements to our GHDE segmentation method, the development of more general- ized and efficient shape similarity metrics, optimization of our search space partitioning and indexing approach, refine- ment of the set of shape features useful in medical applica- tions, and advances in query speed and accuracy.

REFERENCES

[1] [1] “Eval-uation of Shape Indexing Methods for Content-Based Retrieval of X-Ray Images” (vol. 5032) published in 2003 by S. Antani, L. R. Long, G. R. Thoma, and D.J. Lee in the Proceedings of the IS&T/SPIE 15th Annual Symposium on Electronic Imaging, Storage and Retrieval for Media Databases.

In their 1999 article “Shape Indexing by Multi-Scale Representation” published in Image and Vision Computing, A. Del Bimbo and P. Pala discuss this topic at length.

Proc. of IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), Santa Fe, NM, Apr. 7-9, 2002, pp. 139-143, “Efficient Region Boundary Approximation Using Adaptive Smoothing and Second- Order B-Splines” [3] T. El Doker and P. A. Mlsna.

Paper presented at the 2002 IEEE Midwest Symposium on Circuits and Systems (MWSCAS) in Tulsa, OK, by T. El Doker and P. A. Mlsna, titled “Efficient Unsupervised Estimation of Second-Order B-Spline Contour Descriptors,” is cited as [4].

P. Enser’s “Pictorial Information Retrieval” from the Journal of Documentation (vol. 51, no. 2, 1995, pp. 126-170) is cited as an authoritative source.

[6] “World-Wide Maze” by R. Jain appeared in IEEE Multimedia, volume 2, issue 3, in 1995.

In the 2001 edition of Computational Imaging and Vision, edited by R. C. Veltkamp, H. Burkhardt, and H. P. Kriegel, volume 22, state-of-the-art in content-based image and video retrieval, L. Latecki and R. Laka¨mper discuss “Shape Description and Search for Similar Objects in Image Databases” (pp. 69-96).

“Similarity Measurement using Polygon Curve Representation and Fourier Descriptors for Shape Based Vertebral Image Retrieval” (Proc. of SPIE International Symposium on Medical Imaging: Image Processing, 2003, v. 5032), written by D.J. Lee, S. Antani, and L. R. Long.

[9] In “Use of Shape Models to Search Digitized Spine X-Rays,” published in the Proceedings of the IEEE Computer Based Medical Systems conference in Houston, TX, June 23-24, 2000, on pages 255-260, L. R. Long and G. R. Thoma discuss this topic.

[10] “Segmentation and Feature Extraction of Cervical Spine X-Ray Images” by L. R. Long and G. R. Thoma was published in the Proceedings of SPIE Medical Imaging: Image Processing in 1999. The conference was place in San Diego, CA, from February 20-26. The conference was held in volume 3661 and the papers were pp. 1037-1046.

Using shape-based boundary descriptors, N. M. Sirakov, J. Swift, and P. Mlsna conducted an image database query at the 2003 International Conference on Intelligent Systems Design and Applications in Tulsa, OK, from August 10th to the 13th. The paper is available in the proceedings volume 373–382.

“A System for Reconstructing and Visualizing Three-Dimensional Objects” (N. M. Sirakov and F. Muge, 2001, pp. 59-69), published in the International Journal on Computers and Geosciences, volume 27, issue 1, in 2001.

[13] “Vertebral Shape: Automatic Measurements with Active Shape Models” (Publication: Radiology, Volume 211, Issue 2, May 1999, Pages 571–578) by P. P. Smyth, C. J. Taylor, and J. E. Adams.

In their 2001 paper “Estimation of Ori-entation of Cervical Vertebrae for Segmentation with Active Shape Models,” G. Zamora, H. Sari-Sarraf, and S. Mitra presented their findings at the SPIE Medical Imaging: Image Processing conference in San Diego, California, from February 17th to the 23rd.