A Contour-based Shape Descriptor for Biomedical Image Classification and Retrieval

Daekeun You, Sameer Antani, Dina Demner-Fushman, George R. Thoma

National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

ABSTRACT

Contours, object blobs, and specific feature points are utilized to represent object shapes and extract shape descriptors that can then be used for object detection or image classification. In this research we develop a shape descriptor for biomedical image type (or, modality) classification. We adapt a feature extraction method used in optical character recognition (OCR) for character shape representation, and apply various image preprocessing methods to successfully adapt the method to our application. The proposed shape descriptor is applied to radiology images (e.g., MRI, CT, ultrasound, X-ray, etc.) to assess its usefulness for modality classification. In our experiment we compare our method with other visual descriptors such as CEDD, CLD, Tamura, and PHOG that extract color, texture, or shape information from images. The proposed method achieved the highest classification accuracy of 74.1% among all other individual descriptors in the test, and when combined with CSD (color structure descriptor) showed better performance (78.9%) than using the shape descriptor alone.

Keywords: shape representation, contour feature extraction, biomedical image classification, content-based image retrieval

1. INTRODUCTION

Biomedical image retrieval from large collections can be made more effective and relevant to a query if it can be annotated with information about its imaging type or modality (e.g., X-ray, CT, MRI, photograph, etc.). Successful modality classification of images would enhance the performance of article or image retrieval systems (e.g., OpenI*) by reducing the search space to the set of relevant modalities. For example, to search for *posterior-anterior (PA) chest X-rays with an enlarged heart* in a radiographic collection, the database images could be pre-filtered using automatic categorization by modality (e.g., X-ray), body part (e.g., chest), and orientation (e.g., PA) before any visual similarity between images in the database and the query image is computed. Figure 1 shows two screenshots of OpenI system that clearly demonstrate the effect of modality classification in image retrieval.

OpenI (pronounced "open eye") [1] is an open access biomedical image retrieval system developed by U.S. National Library of Medicine (NLM). It applies multimodal (combined text and image) features to index and retrieve images and full text in the open-access subset of biomedical research articles from the NLM's PubMed Central® (PMC) repository. OpenI currently provides access to over 1.3 million images from over 400,000 medical articles, and its production site has about 20,000 visitors per day with nearly 1 million page views. Search results in such systems with a huge collection of data can be significantly improved by automatically narrowing relevant queries to specific imaging modality of interest. For example, the two screenshots in Figure 1 show retrieved images for a query "abdominal pain". As can be seen, a result with no modality filtering applied contains various imaging modalities; however, when the search is limited to CT (computed tomography) scan by selecting "CT scan" from the left-hand option panel for modality filtering, the result is now highly relevant to CT scan.

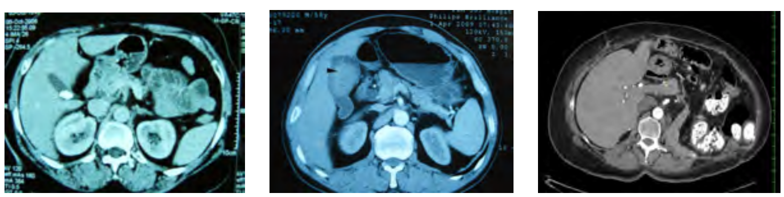
Visual features (descriptors) have been primarily used for biomedical image modality classification with little attention devoted to the use of text metadata (e.g., figure caption) in the articles [2]. Majority of the visual descriptors used in recent approaches reported in [2] extract color and texture information, either separately (e.g., CLD (Color Layout Descriptor)[3] or Tamura (texture)[4]) or in combination (e.g., CEDD (Color and Edge Directivity Descriptor)[5]). Extracted features are then used to train classifiers for various classification tasks, for example, in a hierarchical modality classification discussed in [6].

*http://openi.nlm.nih.gov/

(a) Retrieval result w/o modality filtering (b) Retrieval result w/ modality filtering

Figure 1. Retrieval results w/o and w/ image modality filtering

Visual descriptors that we use in our research [1, 6] mostly encode color and/or texture information from images. We noticed in our research that shape descriptors that capture overall image shape information may be useful in distinguishing certain image types. For example, Figure 2 shows three abdominal CT images with different color. Color information would likely not be the best feature to classify them as the same semantic modality label "CT". Shape information, however, can be a promising solution for this kind of problem. A shape descriptor can capture the outline of the abdominal body shape that other color or texture-based descriptors may not be able to capture and group those that contain certain shapes. Figure 3 shows some sample image types where shape information may be more useful than other visual features to represent and classify. Radiology images generally contain objects with specific shapes that are primarily determined according to the body parts taken by the imaging equipment, and hence could benefit the most from shape descriptors than other modality classes discussed in [6].



(a) (b) (c)

Figure 2. Samples of abdominal CT scans with different colors (in RGB color space, from left to right, (122,195,200), (103,172,208), and (34,34,34), respectively)

(a) Brain MRI (b) Chest x-ray (c) Lung CT (d) Fundus photograph (e) Endoscopy

Figure 3. Sample images that may be represented better by shape than other information

Shape information has been frequently used for object detection and recognition, and an extensive literature exists on various shape representation methods [7]. However, it is difficult to find methods suitable for our purpose since most of the methods in the literature aim to represent a single object by shape that is extracted from contours (outline) or blobs (region) of the object. Shape information is also applied to medical images for specific ROI (region of interest) segmentation (e.g., lung, corpus callosum, blood cell, etc.) [8]. In our research, however, we have no specific objects or ROIs to find from images but need a descriptor that represents an entire image by overall shape information, similar to other visual descriptors that encode visual information such as color or texture.

In this article we propose a shape descriptor that can be used for biomedical image modality classification. We propose a new shape descriptor and demonstrate improvement over comparable shape descriptor from the literature. We use the radiology image modality classification task to evaluate the performance of the proposed shape descriptor. In our approach we adapt the feature extraction method in [9] and consider several image preprocessing techniques for successfully adaptating it in our application. To this end we apply several methods for contour extraction and for extracting features from them. We extract shape descriptors using different methods that are discussed in Section 3 and determine their usefulness by performing image clustering using the LBG algorithm [10]. Our ultimate goal is to apply this shape descriptor to all types of biomedical images.

The remainder of this article is organized as follows. Section 2 provides a brief summary of relevant research in literature and section 3 describes our proposed method. Evaluation results and discussion appear in section 4, and conclusions and future work are given in section 5.

2. RELATED WORK

Shape representation methods found in the literature generally attempt to find an optimal description for specific objects and use it for image indexing and retrieval. Most techniques reported in a recent survey [7] extract some set of numbers from an object of interest that is given as a form of object boundary (contour) or blob (region), and the extracted descriptors can then be used to search for similar shapes in a database or to assign a class label that the unknown object can be classified into. Among many proposed methods, we mention two papers here that may be the most relevant to our proposed method.

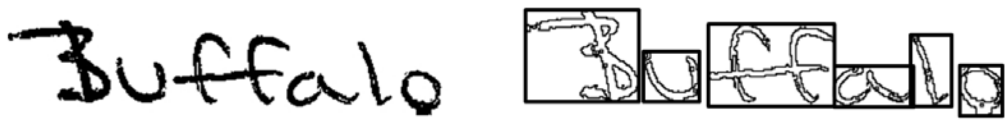
Dalal and Triggs [11] proposed a method, called HOG (Histograms of Oriented Gradients), for human detection (i.e., is it an image of a human(s)?). They divided the image into small subimages ("cells"), computed a local 1-D histogram of gradient directions (or, edge orientations) over the pixels of each cell, and concatenated all histograms to one HOG descriptor. Also, several variants of spacial structure, gradient computation, and normalization methods were applied to improve the performance. They tested the descriptor on datasets containing human or non-human images that were cropped (size of 64×128 or similar) from general photos.

Bosch et al. [12] proposed a shape descriptor (PHOG: Pyramid of Histograms of Orientation Gradients) that is mainly inspired by the image pyramid representation of Lazebnik et al. [13] and the HOG descriptor. They first extracted edge contours using Canny edge detector on monochrome versions of images, and used a 3×3 Sobel mask to compute orientation gradients. Edge gradients of each contour point are then counted to extract a HOG descriptor from each cell at each level of the pyramid. All HOG descriptors are then concatenated to one descriptor, as is done with the HOG descriptor [11]. They tested the descriptor on two datasets, Caltech-101 and TRECVID, containing images falling into 101 and 39 semantic categories (e.g., car, leopard, airplane, etc.), respectively.

3. METHOD

3.1 Overview

Our proposed shape descriptor is motivated by the feature extraction method proposed in [9] where 8-directional chain code features are extracted from 3×3 subimages for handwritten optical character recognition (OCR). We used the feature extraction method in our previous work for OCR [14]; however, we have never attempted to apply the method for shape representation of biomedical images due to obvious differences between the two types of images, viz., text images and biomedical images. The differences include the number of contours under consideration for feature extraction, size (resolution) of input images, and the large variety of shapes that are represented by the contour feature. Figure 4 shows two sample images that illustrate the differences. The input text string may be segmented into rectangle boxes shown in the contour image. Feature extraction is performed on each box, and these features represent shapes of the letters. Note that only a few contours are present in each box. However, the brain MRI (Magnetic Resonance Imaging) shown in Figure 4(b) contains many contours that are more complicated in shape and arrangement, compared to the text string. All contours from the brain image may not be useful to classify the image as an MRI, and hence it may be important to identify the most promising contour(s) for successful modality classification. We address this issue later in section 3.3.4. Figure 5 illustrates the structure of the proposed method.



(a) A text string and contours

(b) A brain MRI and contours

Figure 4. Contours extracted from a text string and a brain MRI

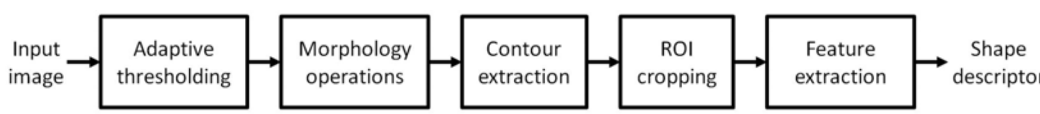
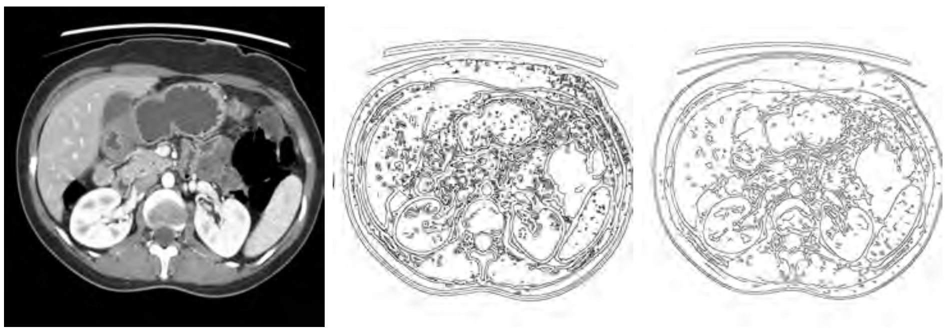


Figure 5. Our proposed shape descriptor extraction algorithm

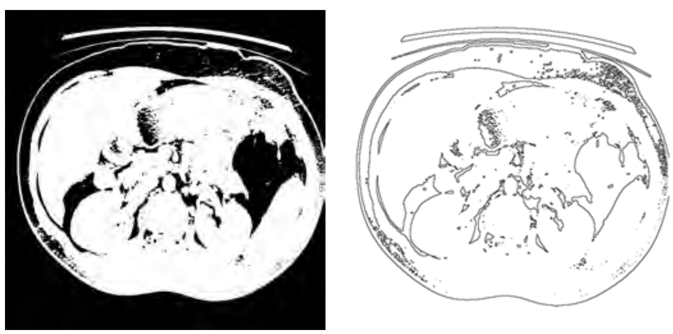
While our proposed method may be similar to HOG or PHOG in terms of feature extraction, our method differs from them in two ways: i) we extract shape information from contours, not from edge gradients, and ii) our method has only four directional slopes (comparable to the orientation bins in the two methods), which results in a much lower dimensional shape descriptor than the two. We compare our method with PHOG and other visual descriptors for radiology modality classification in section 4.

3.2 Adaptive thresholding for contour extraction

Contour extraction is a critical step in our method since shape information is computed from the contours. We examined two approaches: edge detection-based and adaptive thresholding-based methods [15, 16] and choose the latter for the following reasons. Edge detection-based methods tend to generate more contours (details or often noise) than necessary for describing shapes in images. They may need additional pre or post-processing such as blurring or binarization in order to eliminate noisy contours, or obtain a binary image from edge detection, respectively. Figure 6 shows examples of three contour detection results. The input image was first blurred to remove noisy contours in edge detection results. Figure 6(b) is obtained by Sobel edge detection followed by an adaptive thresholding method [16]. For Canny edge detection in Figure 6(c), two thresholds, 60 and 100, are used. As shown in the two results (Figure 6(b) and 6(c)), edge detection-based methods successfully detect details of the tissue structure and outlines of organs; however, they generate too many noisy edges as well even when image blurring was applied to the input image. In contrast, adaptive thresholding method may not be capable of detecting such details that are detected by the edge detection-based methods. As shown in Figure 6(e), the method missed several edge contours of organs and tissue structure; however, the result may be sufficient to describe the overall shape of the CT scan by the outer contour of the body, and hence provides sufficient information to classify the image. Details of inner regions that are successfully detected in Figure 6(b) and 6(c) may add confusion rather than discrimination power in our application. Another reason for applying adaptive thresholding in contour extraction is that radiology images can be divided into two image regions, viz., body or foreground and background, and the method can separate the two regions fairly well (Figure 6(d)). The details, in general, can vary significantly from slice to slice (location where the picture is taken) even for the same body part (i.e., abdominal), and hence leaving them out may lead to better performance.



(a) Input abdominal CT scan (b) Sobel edge detection (c) Canny edge detection



(d) Adaptive thresholding of the input (e) Contours extracted from (d)

Figure 6. Contour extraction results using three different methods. Each edge detection operator was applied to a blurred image of the input.

3.3 Feature extraction

We examine various feature extraction methods and conditions such as size of square grid (subimages), weighting method among multiple contours, feature normalization, and cropping image regions of interest (ROI). Along with the clustering method mentioned in the *Introduction*, we also trained SVM classifiers to assess the new methods by 10-fold cross-validation accuracy on training images.

3.3.1 Chain code representation

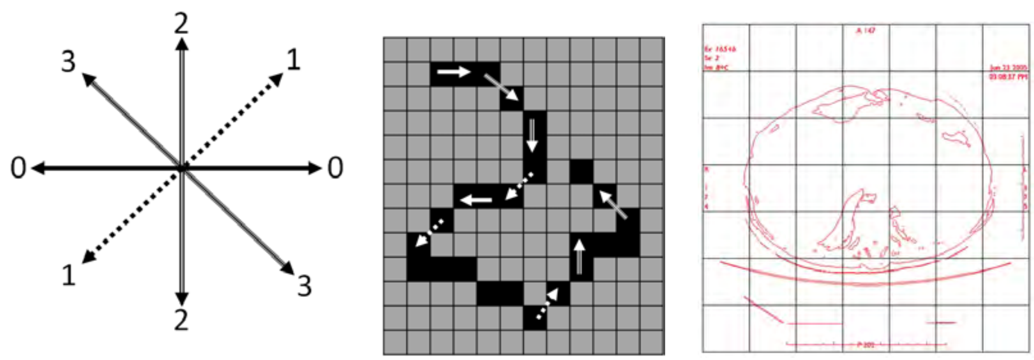
Figure 7 illustrates a 4-directional chain code and how the contour feature is extracted from a contour. Instead of the 8-directional chain code in [9], we use a 4-directional chain code since 8-directional chain code is meaningful only if tracing order in a contour matters, which is not relevant for our goal.

As shown in Figure 7(c) we extract the chain code feature from subimages that are obtained by dividing a contour image by a 7×7 square grid. Then, we obtain a histogram of each directional code (0∼3) from each subimage. The histogram from each subimage is then normalized by the total number of directional codes from the subimage, and all the histograms are concatenated to form one feature vector. For example, assume that the example in Figure 7(b) is a subimage. Then histogram of the 4-directional chain code is [9 5 5 6] and normalizing the histogram by the total count (25) is [0.36 0.2 0.2 0.24]. All 4-dimensional histograms (one from each subimage) are then concatenated into one 196-dimensional vector (7×7×4). Once all feature vectors are extracted from a dataset (either training or testing) we perform a normalization again as follows,

$$f_{Ii} = \frac{f_{Ii}}{f_{i_{max}}} \tag{1}$$

where f_{Ii} is i^{th} element in a feature vector of I^{th} image in a dataset and $f_{i_{max}}$ is obtained by

$$f_{i_{max}} = \max_{I}(f_{Ii}) \tag{2}$$



(a) 4-directional chain code (b) Feature extraction from a contour (c) 7×7 square grid on a contour image

Figure 7. Feature extraction from subimages

3.3.2 Size of square grid

We tried different square grids ranging from 3×3 to 10×10, and found 7×7 to be optimal in terms of classification accuracy and dimensionality of the feature vector. Small grids such as 3×3 generally may miss shape details since a relatively large region is covered by one cell. A grid size larger than 7×7, on the other hand, does not show better performance than the 7×7 size. We choose 7×7 for grid size in this research unless different size is specified.

3.3.3 Noise removal by morphology

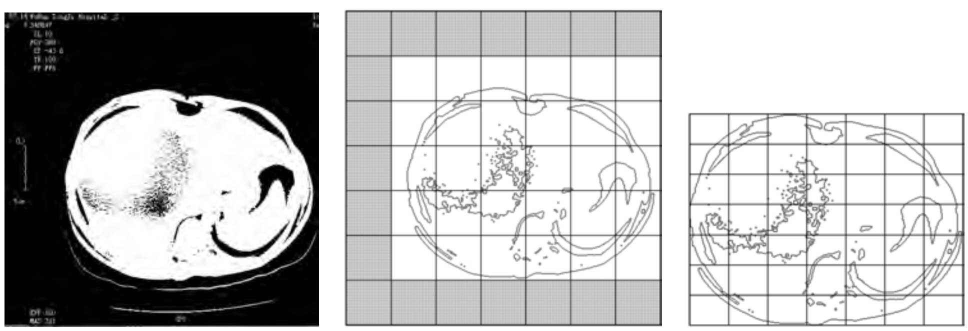
In order to form better shape contours and eliminate noisy edges, morphology operations (dilation and erosion) are applied to a binary image before detecting contours. Figure 8 shows the effect of an opening operation applied to the initial binary image shown in Figure 8(a). Text appearing on the corners in the image are successfully eliminated, and the external contour (indicated by black arrow in Figure 8(c)) becomes cleaner in Figure 8(b), compared to the input image. Final contours are then extracted from Figure 8(b) and fed to a feature extraction module.

(a) Initial binarization result (b) After opening operation (c) Contours extracted from (b)

Figure 8. Morphology operation applied to binarization result to extract better contours

3.3.4 Weighting by contour length

Compared to each rectangular box in Figure 4(a) that mostly contains one contour for feature extraction, images in this research usually generate several tens or hundreds of contours. The number of extracted contours depends on the extraction method used as shown in Figure 6, or on image types. For example, X-ray images typically generate fewer contours than other modality classes. Our hypothesis is that a longer contour has more useful information for shape representation. For example, the longest external contour (pointed to by a black solid arrow) in Figure 8(c) provides an accurate shape of the abdominal CT scan. As discussed in section 3.2, the other contours may vary from slice to slice and hence provide less similarity information than the external contour. We multiply the chain code features extracted from each contour by a length ratio between its contour length and the longest contour's length so that the longest one contributes 100% to the final descriptor and others in proportion to their length. We found this weighting scheme slightly improves classification accuracy.



(a) Initial binarization result (b) 7×7 grid on initial contours (c) 7×7 grid on contours after ROI cropping

Figure 9. ROI cropping

3.3.5 Cropping image region of interest (ROI)

Radiology images often have a large background region around the main body region that is generally located in the center of the images. As shown in Figure 9, subimages and contours in each subimage are quite different according to where the grid is located, and the difference may result in quite different shape descriptions for the same image. For example, the shaded subimages in Figure 9(b) contain no contours; however, corresponding subimages in the cropped image in Figure 9(c) contain the most important external contour, and this difference may result in a significant dissimilarity between the two shape descriptors. Figure 10 shows an effect of ROI cropping that would affect image classification and retrieval. We performed image clustering by descriptors extracted from original images and cropped images. Compared to the sample clusters in Figure 10(a) and 10(b) that are obtained without ROI cropping, the cluster shown in Figure 10(c) contains both types of images (w/ or w/o background region). As can be seen from Figures 10(a) and 10(b), images w/ and w/o background are clustered separately when no ROI cropping is applied. Both results are apparently reasonable and may be meaningful; however, we prefer the result with the ROI cropping applied since we find no meaningful information for classification and retrieval from the background regions.

(a) Without cropping 1: a sample cluster containing images with background

(b) Without cropping 2: a sample cluster containing images without background

(c) With cropping: images w/ or w/o background are collected together

Figure 10. Effect of ROI cropping

In order to find a region of interest, we examine (x,y) coordinates of all contour points, find a bounding box of each contour, and then find an overall bounding box. ROI cropping needs morphology operations (discussed in section 3.3.3) to achieve a better result. Without morphology operations various factors such as overlay text, ruler, and other noise outside of the body region are mostly included in the cropped images, which detracts from the cropping effect.

3.3.6 Negative image

We use a contour extraction function (cvFindContours) provided in OpenCV [17] to find contours of white blobs and holes (contours of black blobs within a white blob). For example, from the initial binary image shown in Figure 11(a), contours are extracted only from the head, shown in Figure 11(b). Figure 11(d) shows a different extraction result that is obtained from the negative (Figure 11(c)) of the initial binary image. Contours from the negative image are almost the same in the head region, but it also provides a frame contour of the image that is not extracted from the initial binary image. This is the same effect as extracting contours from black blobs in the initial binary image instead of white blobs. Several experiments using shape descriptors generated from negative images showed that the descriptor alone achieves a classification accuracy similar to descriptors from initial binary images; however, we found that combining the two descriptors into one vector by concatenating them improved classification accuracy by about 5% compared to the accuracy of individual descriptors. We also found that a 3×3 grid is the most efficient way of utilizing negative images since the only difference from the two results (Figure 11(b) and 11(d)) is the presence of the frame contour, and a 3×3 grid is found to be sufficient to capture the difference.

(a) Initial binary image (b) Contours (c) Negative image (d) Contours from negative image

Figure 11. Contour extraction from the initial binary image and its negative

4. EVALUATION

4.1 Test setup and evaluation

We collected two datasets (for training and testing) from a subset of the open-access collection of PubMed Central®(PMC) for medical image retrieval track of ImageCLEF2012 [2]. One set contained 2,419 and the other 2,576 radiology images. Table 1 shows number of images of different modality in our training and test sets.

Table 1. Number of images by modality in our datasets

Modality	Training set	Test set
Magnetic resonance imaging (MRI)	523	228
Computed tomography (CT)	1,146	712
X-ray	489	1,229
Ultrasound	207	380
Scintigraphy (PET)	54	27
# of total images	2,419	2,576

We performed modality classification tests using the visual descriptors below.

- Shape: PHOG, Proposed descriptor
- Color and texture: CEDD, FCTH (Fuzzy Color and Texture Histogram)[18]
- Color: CLD, CSD (Color Structure Descriptor)[3]
- Texture: Tamura
- Combined descriptors: Proposed+CLD, Proposed+CSD, Proposed+Tamura

We adjusted several parameters of PHOG such as number of bins (K), orientation range ([0 180] or [0 360]), and pyramid level (L). For this proposed shape descriptor, we used 7×7 and 3×3 square grids for initial and negative binary images, respectively. All preprocessing methods, viz., morphology operations, ROI cropping, and weighting by contour length, discussed in section 3 were applied. Other descriptors are tested with no parameter adjustments.

SVM classifiers [19] (RBF kernel, $\gamma = 0.002$, $C = 100$) are trained using the training images and tested on the test images. Classification accuracy is reported as performance metric.

To evaluate the proposed shape descriptor, we first tested it individually and then combined it with other descriptors such as CLD, CSD, and Tamura. We simply concatenated two descriptors and tested the combined descriptor in the same way.

Table 2. Classification accuracy

Descriptors		Feature type	Feature Dim.	Accuracy (%)
PHOG	K=40, L=3, [0 360]	Shape	3,400	71.0
	K=16, L=3, [0 180]		1,360	60.1
	K=16, L=2, [0 180]		336	61.1
	K=8, L=3, [0 360]		680	67.4
	K=4, L=3, [0 180]		340	62.7
	K=4, L=3, [0 360]		340	62.2
CEDD		Color and texture	144	71.7
FCTH		Color and texture	192	70.0
CLD		Color	16	50.8
CSD (128 bins)		Color	128	61.9
Tamura		Texture	18	70.6
Proposed		Shape	232	**74.1**
Proposed + Tamura		Shape and texture	250	**77.8**
Proposed + CLD		Shape and color	248	**74.9**
Proposed + CSD		Shape and color	360	**78.9**

4.2 Evaluation results and discussion

Table 2 shows evaluation results of each visual descriptor on a test set. Our proposed descriptor achieved the highest classification accuracy in the individual descriptor test. PHOG with K=40, CEDD, FCTH, and Tamura showed very similar accuracy, around 71%. Color information alone such as CLD and CSD showed relatively lower accuracy than other descriptors. The lower performance may be explained as follows. All the images in our datasets are radiology images that are mostly gray-scale images. Some images, for example, color doppler or PET (Positron Emission Tomography), may contain color pixels; however, color portions are generally much smaller than gray portions within an image. Hence color information alone may have less discriminative power than other information such as texture or shape. CEDD and FCTH, however, encode both color and texture information, and the texture information may complement the lack of discrimination power of color information, and as a result they could achieve better performance.

PHOG descriptor achieved its highest accuracy with bin size of 40. K=40 with orientation range of [0 360] was reported as an optimal parameters in [12]; however, under consideration of the feature dimension and accuracy together, it may not be an optimal descriptor for our modality classification. Compared to the last parameters with K=4 in the result, the feature dimension is ten times larger, but the accuracy was improved only about 9%. As can be seen, the PHOG with K=8, L=3, and orientation range of [0 360] may be an optimal method for its relatively lower feature dimension (yet higher than our descriptor and others) and an accuracy of 67%. Also, it is difficult to find a strong relationship between classification accuracy and each parameter. As shown in Table 2, PHOG results with different L, orientation range, or K but all others identical, showed no significant difference in accuracy.

Our proposed descriptor achieved the highest accuracy in the individual descriptor test. Combined descriptors achieved better performance than the shape descriptor alone. Early fusion (e.g., concatenating multiple feature vectors) is a common way of combining different types of descriptors for the modality classification problem [2]. CSD combined with the proposed shape descriptor achieved the highest accuracy. Combining them improved accuracy of CSD alone about 17% and that of the proposed descriptor alone about 5%. Shape descriptors generally encode no color information. Hence combining color descriptors with shape descriptors may add useful information that can not be captured by a shape descriptor alone, and that may result in some improvement in the accuracy.

Our proposed method showed better performance in terms of both accuracy and feature dimension compared to the PHOG descriptor which is the only shape descriptor in this test. It is difficult to select one result for performance comparison since PHOG was extracted with different parameters, which resulted in different feature dimension and accuracy. We compare three PHOG results with our shape descriptor here. First comparison is made with PHOG with K=40, L=3, and orientation range of [0 360] that achieved the highest accuracy among all PHOG results. The accuracy is close to (yet lower than) that of our descriptor; however, the feature dimension of 3,400 is much higher than the proposed. In our previous research [6], we combined 15 different visual descriptors (e.g., CEDD, CLD, FCTH, Tamura, EHD[3], etc.) and the dimensionality was 1,361. Hence the accuracy of 71.0% appears much lower than others in consideration of the highest feature dimension, even though the accuracy is actually quite good in the result. Second comparison is made with PHOG with K=8, L=3, and orientation range of [0 360] that achieved the second highest accuracy in PHOG results. As discussed above, this could be an optimal parameter set of PHOG in the test. The accuracy is close to 70% and the feature dimension is 680; however, compared to our method, it achieved a lower accuracy using more feature elements. Our last comparison is made between the proposed and PHOG with K=4, L=3, and orientation range of [0 180] that may be the most similar feature extraction scheme to our method. With such parameters, images are divided into 8×8 subimages at level 3 of the pyramid (c.f., 7×7 subimages in the proposed). Also, the number of bins (K=4) and orientation range are the most similar with the proposed method. However, the accuracy is now decreased to about 62% that is close to CSD's accuracy and much lower than the proposed (about 12%). From all these comparisons, we conclude that our proposed method performs better than PHOG in terms of accuracy and descriptor's dimensionality for radiology modality classification.

In our previous research [6], we also performed modality classification and reported classification accuracy. Those results, however, are not directly comparable with this article since the classification in [6] includes additional modality classes such as *Photo*, *Illustration*, and *Microscopy* and their sub-classes.

5. CONCLUSION

A new shape descriptor for biomedical image classification and retrieval is presented in this article. We adapted a feature extraction method that has been successfully applied for character recognition (OCR). In order to deal with potential shortcomings of the method in applying it to a quite different problem, we applied several typical image processing techniques such as adaptive thresholding and morphology and considered various feature extraction conditions. Contours are extracted from preprocessed binary images, and 4-directional chain code features are then extracted from the contours in subimages. We evaluated our method through radiology modality classification, and compared it with several visual descriptors that extract different types of information from images (e.g., color and/or texture). Experiment results showed that the proposed shape descriptor achieved the highest classification accuracy of 74.1%, and combining the descriptor with color-based descriptor enhanced the accuracy of each individual descriptor. Compared to another shape descriptor (PHOG), our descriptor showed better classification accuracy (by about 7%) using fewer feature dimensions (about 34%), compared to the optimal PHOG result.

Future work includes: i) apply the shape descriptor to our entire dataset (including microscopy, photography, illustration figures, etc.) and identify image types that can be accurately separated by the shape descriptor for modality classification, and ii) apply the descriptor to general databases such as Caltech-101 and assess the usefulness of the method as a general purpose shape descriptor.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the Lister Hill National Center for Biomedical Communications, an R&D division of the National Library of Medicine, at the National Institutes of Health, U.S. Department of Health and Human Services.

References

[1] Demner-Fushman, D., Antani, S., Simpson, M. S., and Thoma, G. R., "Design and development of a multimodal biomedical information retrieval system," *JCSE* **6**(2), 168–177 (2012).

[2] Müller, H., de Herrera, A. G. S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., and Eggel, I., "Overview of the ImageCLEF 2012 medical image retrieval and classification tasks," in [*CLEF (Online Working Notes/Labs/Workshop)*], (2012).

[3] Sikora, T., "The MPEG-7 visual standard for content description-An overview," *Circuits and Systems for Video Technology, IEEE Transactions on* **11**(6), 696–702 (2001).

[4] Tamura, H., Mori, S., and Yamawaki, T., "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions on* **8**(6), 460–473 (1978).

[5] Chatzichristofis, S. A. and Boutalis, Y. S., "CEDD: Color and edge directivity descriptor: a compact descriptor for image indexing and retrieval," in [*Proceedings of the 6th international conference on Computer vision systems*], *ICVS'08*, 312–322, Springer-Verlag, Berlin, Heidelberg (2008).

[6] You, D., Rahman, M. M., Antani, S., Demner-Fushman, D., and Thoma, G. R., "Text- and content-based biomedical image modality classification," *Proc. SPIE 8674, Medical Imaging 2013: Advanced PACS-based Imaging Informatics and Therapeutic Applications* .

[7] Mingqiang, Y., Idiyo, K. K., and Joseph, R., "A survey of shape feature extraction techniques," *Pattern Recognition, Peng-Yeng Yin (Ed.) (2008) 43-90* , 43–90 (Nov. 2008).

[8] He, L., Peng, Z., Everding, B., Wang, X., Han, C. Y., Weiss, K. L., and Wee, W. G., "A comparative study of deformable contour methods on medical image segmentation," *Image and Vision Computing* **26**(2), 141 – 163 (2008).

[9] Kim, G. and Govindaraju, V., "A lexicon driven approach to handwritten word recognition for real-time applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **19**(4), 366–379 (1997).

[10] Linde, Y., Buzo, A., and Gray, R., "An algorithm for vector quantizer design," *Communications, IEEE Transactions on* **28**(1), 84–95 (1980).

[11] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection," in [*Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*], **1**, 886–893 vol. 1 (2005).

[12] Bosch, A., Zisserman, A., and Munoz, X., "Representing shape with a spatial pyramid kernel," in [*Proceedings of the 6th ACM international conference on Image and video retrieval*], *CIVR '07*, 401–408, ACM, New York, NY, USA (2007).

[13] Lazebnik, S., Schmid, C., and Ponce, J., "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in [*Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*], **2**, 2169–2178 (2006).

[14] You, D., Antani, S., Demner-Fushman, D., Govindaraju, V., and Thoma, G. R., "Detecting figure-panel labels in medical journal articles using MRF," in [*ICDAR*], 967–971 (2011).

[15] http://www.xdp.it/cximage/.

[16] You, D., Apostolova, E., Antani, S., Demner-Fushman, D., and Thoma, G. R., "Figure content analysis for improved biomedical article retrieval," *Document Recognition and Retrieval XVI* **7247**, 72470V–10, SPIE, San Jose, CA, USA (2009).

[17] http://opencv.willowgarage.com/wiki/.

[18] Chatzichristofis, S. and Boutalis, Y., "FCTH: Fuzzy color and texture histogram - a low level feature for accurate image retrieval," in [*Image Analysis for Multimedia Interactive Services, 2008. WIAMIS '08. Ninth International Workshop on*], 191–196 (2008).

[19] Chang, C.-C. and Lin, C.-J., "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27 (2011). Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.