Body Segment Classification for Visible Human Cross Section Slices

Zhiyun Xue, Sameer Antani, L. Rodney Long, Dina Demner-Fushman, George R. Thoma
Lister Hill National Center for Biomedical Communications
National Library of Medicine
Bethesda, USA
{xuez,rlong,santani,gthoma}@mail.nih.gov

***Abstract*—Visible human data has been widely used in various medical research and computer science applications. We present a new application for this data: a method to classify which body segment a transverse cross section image belongs to. The labeling of the data is created with the guidance of an online body cross section tutorial. The visual properties of the images are represented using a variety of feature descriptors. To avoid problems that arise from the large dimensionality of features, feature selection is applied. The multi-class SVM is employed as the classifier. Both the CT scans and the color photographs of cryosections of the whole body (male and female) are used to test the proposed method. The high performance with overall accuracy above 98% on both the 2160 CT dataset and the 1870 cryosectional photos show the method is very promising. Because of its observed effectiveness on visible human data, we will extend our approach to classify figures in biomedical articles.**

Keywords—Body segment classification; visible human data; body cross section

I. INTRODUCTION

The Visible Human Project (VHP) initiated by the National Library of Medicine (NLM) two decades ago is one of the most significant efforts made toward providing comprehensive data for understanding human anatomy [1]. The rich data collected by VHP (which was made available world-wide under a no-cost license agreement) provides a potentially highly valuable resource for the computer science and medical research/education community. The dataset, which consists of complete multi-modal images of both male and female cadavers, has been widely used in various applications for image processing, annotating, visualization, and analysis by hundreds of academic and industrial groups in the world. Representative applications include 3D anatomical model construction and visualization [2], body organ segmentation [3], and multi-modal image registration [4]. Anatomy education using visible human data has also been reported [5]. In this paper, we present a new method to identify which body segment (such as thorax, abdomen) in which a cross sectional image is located, using only the visual features of the given image. To the best of our knowledge, there is no such application in the existing literature. This work was motivated by our need to solve a related image processing problem: figure classification, which is one step in our research and development in integrating text and image for biomedical article retrieval. We have been developing OPENi, an online system which provides ability to search figures in biomedical articles based on the visual contents of the figures, in order to augment traditional text-query searching. To try to optimize retrieval performance and enhance user search experience, we proposed a hierarchical classification framework for figure image analysis. The figures are first classified into *regular figures* or *illustration figures*. Then the regular figures are classified based on the image modality, such as MRI, CT, X-ray, Ultrasound, photograph, etc. The illustration figures can be classified into categories of graphs, diagrams, flow charts, etc. The next level for modality images, such as CT, MRI and X-ray, is body segment classification, for example, head, abdomen, or pelvis. We have developed effective methods for regular/illustration classification [6] and modality classification [7]. Body segment classification is our next goal. To that end, we have adopted the visible human data as an initial experimental data set to test our body segment classification methods; our goal is to move from this data set to processing figures in biomedical articles. We expect some modification of our methods will be necessary, since the figures in articles have different characteristics from those of visible human data (for example, for the same modality of figures, the images may be obtained from different bodies, different machines, different resolution, and different views). In this paper, we report the first step of our work toward body segment classification of images, namely the results of our work on the visible human data. This paper is organized as follows. First, in Section 2, we describe the cross sectional image data we employed and how the images are labeled. Our method, which contains steps of feature extraction, feature selection and classification, is presented in Section 3. Section 4 reports and discusses experimental tests. The concluding remarks and future work are given in Section 5.

II. VISIBLE HUMAN DATA AND LABELING

A. Visible Human Data

The visible human data consists of two entire human bodies (normal male and female) in three image modalities: computerized tomography (CT), magnetic resonance images (MRI), and cryosectional color photographs (CRYO). For the male cadaver, transverse CT scans of the entire body were obtained both before and after the cadaver was frozen, and are

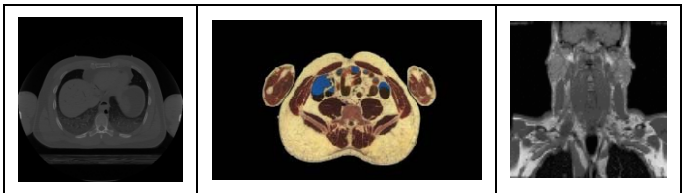
referred to as the *fresh view* and the *frozen view* respectively. For the fresh view, the CT images were collected at one millimeter intervals in the head and neck; at three millimeter intervals in the thorax, abdomen, and pelvis; and at five millimeter intervals in the lower limbs. For the frozen view, the CT images were obtained at one millimeter intervals from head to toe. The size of CT images in both fresh view and frozen view is 512 by 512 pixels. Anatomical axial cryosections that correspond to frozen view CT images were also obtained at one millimeter intervals and then color photographed using a digital camera. The resolution of cryosectional photos is high (2048 by 1216 pixels). Besides CT images and CRYO color photographs, low-resolution (256 by 256 pixels) MRI images are also provided. For MRI images, the head was imaged in the axial plane, and the remainder of the body was imaged in the coronal plane. Fig. 1 shows an example of CT, MRI, and CRYO images of the visible male, respectively. The visible female images were acquired in the same way as those of the male except that anatomical cryosections were taken at 1/3 millimeter intervals. The complete male dataset is approximately 15 GB in size while the whole female dataset is about 40 GB in size. In the dataset, the CT scans and the CRYO anatomic slices of visible human data were aligned and were numbered correspondingly.

B. *Body Slice Labeling*

The labeling of each transverse cross sectional image, i.e., which body segment the image belongs to, is guided by the online LUMEN Cross-Section Tutorial[1]. The tutorial divides the body into six regions: head and neck, upper limb, thorax, abdomen, pelvis, and lower limb, but we consider only five of them: head and neck, thorax, abdomen, pelvis, and lower limb, as illustrated by Fig. 2, because the cross sections of upper limb overlap with those of thorax and other body segments. In addition, the lower limb region defined in our approach doesn't contain the section overlapped with the pelvis region. The guidelines that we used for partitioning the body segments are as follows:

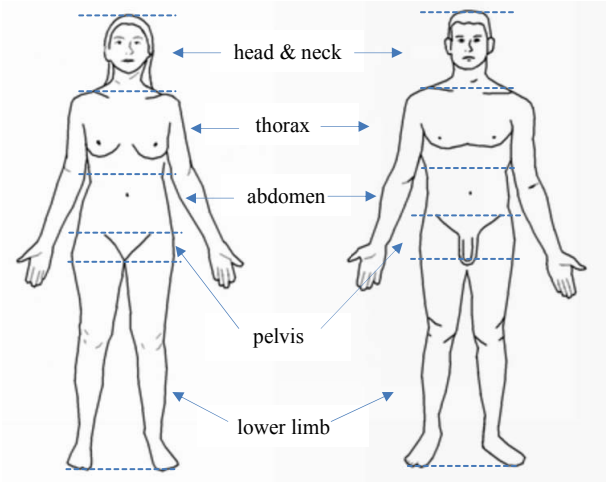
- The separation between the head and neck region and the thorax region is determined by the first presence of the clavicle in the cross section image;
- The separation between the thorax region and the abdomen region is determined by the first disappearance of the inferior lobe in the cross section image;
- The separation between the abdomen region and the pelvis region is determined by the first presence of the ilium in the cross section image;
- The separation between the pelvis region and the lower limb region is determined by the first disappearance of the ischial tuberosity in the cross section image.

Fig. 3 shows the corresponding first and last slice in each segment (the slices are numbered in order from head to toe). Only CT and CRYO cross sections are considered in our tests, because MRI images do not have whole body transverse cross sections.



CT | CRYO | MRI

Fig. 1. Visible human data



http://commons.wikimedia.org/wiki/File:Human_body_features.svg

Fig. 2. Body segments

III. METHOD

The proposed method consists of three steps: feature extraction, feature selection, and classification. The step of feature extraction aims at representing/describing the visual content of the images using features that can preserve class separability as much as possible. Therefore, the features need to be not only sufficient but also discriminative. There are many different features proposed in the literature whose effectiveness is application-dependent. For most cases, using one or a few feature types is inadequate. Therefore, to achieve satisfactory classification results, we first apply various types of descriptors that have been shown to be effective for other applications, i.e., we initially represent the image characteristics using a large number of features. Then, a feature selection procedure is used to remove irrelevant and redundant features, since a large number of features may lead to classifier over-fitting, which degrades classifiers' generalization capabilities. Various powerful techniques have been proposed to solve the problem of multi-class classification, such as neural networks, decision trees, k-Nearest Neighbor, and Naive Bayes. In this paper, we utilize multi-class Support Vector Machines (SVMs) due to their robustness and empirical success.

[1] http://www.meddean.luc.edu/lumen/meded/grossanatomy/x_sec/

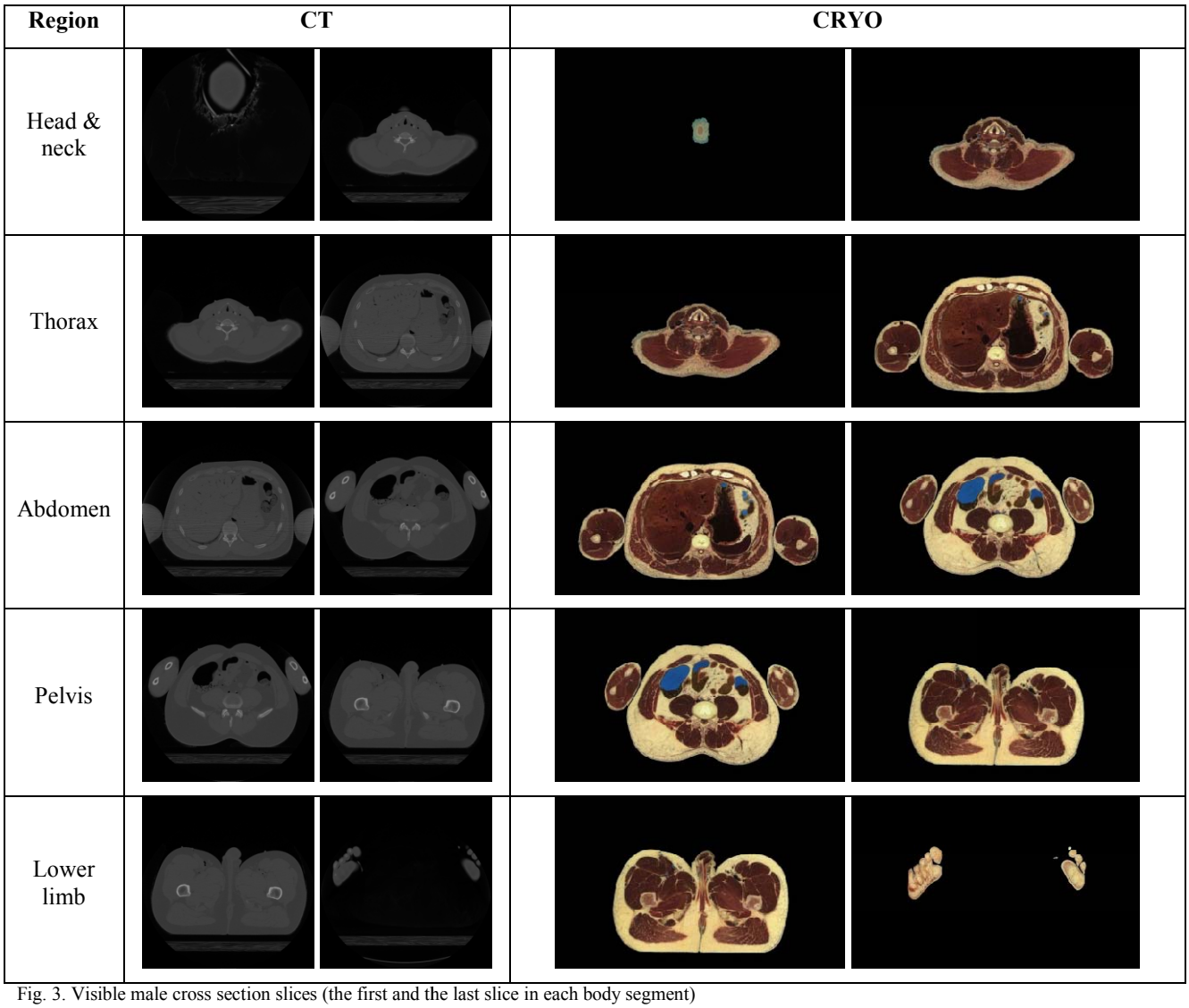


Fig. 3. Visible male cross section slices (the first and the last slice in each body segment)

A. Feature Extraction

With the goal of comprehensively representing the visual characteristics of the original data and capturing the perceptional characteristics that discriminate among images in different regions, we extracted the following features that describe aspects of color, texture, shape, size and spatial structure.

CEDD and **FCTH**: CEDD (color and edge directivity descriptor) [8] and FCTH (fuzzy color and texture histogram) [9] are two descriptors used by the Lucene image retrieval (LIRE) library for image indexing and retrieval. Both features incorporate color and texture information in one histogram which results from the combination of three fuzzy units. The first and second fuzzy units, the part for color information representation, are the same for CEDD and FCTH. They differ in the third fuzzy unit which is for the capture of texture information. Both features are compact and their sizes are limited to less than 72 bytes per image.

CLD and **EHD**: CLD (color layout descriptor) and EHD (edge histogram descriptor) are MPEG-7 features [10]. CLD captures the spatial layout of the dominant colors on an image grid consisting of 8 by 8 blocks and is represented using the DCT (discrete cosine transform) coefficients. EHD represents the local edge distribution in the image, i.e., the relative frequency of occurrence of five types of edges (vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional) in the sub-images.

Bag of Keypoints: SIFT (scale invariant feature transform) [11] features are local features that are relatively invariant to translation, scaling, orientation, and image noise. The algorithm consists of four major steps. First, it detects the maxima and minima in scale space. Then, it identifies key points by removing those extrema with low contrast. Next, it assigns an orientation to each key point. Last, it computes the local image gradient feature measured relative to the orientation of the key point to provide invariance to rotation.

For each key point, a vector of length 128 features distinctively represents the neighborhood around it. After obtaining this feature for each key point, the bag-of-keypoints approach described in [12] was then applied.

LBP: LBP (local binary patterns) [13] operator is a texture descriptor that is robust against illumination changes. The texture information in the image is represented by a histogram of binary patterns. The binary patterns are generated by thresholding the relative intensity between the central pixel and its neighboring pixels. Because of its computational simplicity and efficiency, LBP has been successfully used in various computer vision applications.

Beside the features above, other features we applied are Tamura descriptor [14], primitive length [15], edge frequency, Hu moments [16], color moments [17], autocorrelation coefficients, and image width-to-height ratio.

B. Feature Selection

Feature selection is frequently used in applications when a method of selecting a small number of effective features is not obvious a priori; in these applications, such as ours, we begin by extracting a large number of features from those known in the technical literature. Then we apply a feature selection strategy with the goal of reducing the number of features by selecting an effective subset of the original features. The goodness of features is described by relevance, which can be defined and evaluated by various measures [18]. Alternative feature selection algorithms, which perform as a preprocessing step before learning, have been proposed. Generally speaking, these algorithms can be categorized into two broad groups: wrapper approaches and filter approaches. Wrapper methods use the intended learning algorithm to search through the space of features and evaluate the goodness of each feature subset. Filter approaches do not rely on the specific learning algorithm; rather, features are evaluated based on the general characteristics of the training data. Wrapper methods are generally considered to yield better results, but filter approaches tend to run faster. In our application, we employ the open source machine learning software WEKA [2]. It provides a wide range of attribute selection methods by combining different evaluation measures and different search methods. The specific method we used will be explained in Section IV.

C. Classification

To solve a given multi-class classification problem, one practical approach is to use a combination of binary SVM classifiers. There are several popular strategies on how to carry out binary classification and how to combine the multiple binary results. For example, the one-versus-all strategy [19] converts an N-class classification problem into N binary classification problems, where each binary classifier is trained to separate a given class from the rest of N-1 classes, while the one-against-one strategy [20] requires $N(N\text{-}1)/2$ binary classifiers and each binary classifier is trained to differentiate each pair of classes. For testing, the one-versus-all strategy uses the winner-take-all method in which the unknown example is assigned the class label of the binary classifier that outputs the highest value, while the one-against-one strategy employs max-wins voting in which the class with the maximal number of votes among the binary classifiers determines the label of the given example. There is open source software that implements multi-class SVMs, such as SVM-Light [20], and LibSVM [22]. We use the SVM method given in the WEKA program package. Specifically, we employed the Sequential Minimal Optimization (SMO) algorithm implemented in WEKA [23] which uses one-against-one strategy for multi-class problems,

IV. EXPERIMENTAL TESTS

Two datasets of visible human data were used in the experimental assessment. One is CT cross sections (fresh view). The other is color photos of cryosections. MRI data was not tested because it does not contain a set of whole body axial cross sections. For the CT dataset, both female and male data were included. For cryosectional photos, only male data was tested. Both testing datasets were labeled into five classes of body segments (head and neck, thorax, abdomen, pelvis, and lower limb) as described in Section 2. Table I lists the number of slices in each body segment for the CT and CRYO datasets, respectively. The lower limb segment contains a much larger number of slices because the length of this segment is longer (around 900 slices vs 200+ slices for each of the other segments). In total, there are 2160 images for CT data and 1870 images for CRYO data.

For each image, we extracted thirteen types of features, resulting in a total feature vector length of 2321. To reduce the feature vector length, we encapsulated the attribute selection process with the classifier itself. This can be done by using the meta-classifier "AttributeSelectedClassifier" in the WEKA. With this method, both the attribute selection method and the classifier only get access to the data in the training set (or folds if cross-validation is performed). For the attribute selection method used in the "AttributeSelectedClassifier", we selected the feature evaluator as "CfsSubsetEval" (which evaluates the value of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them) and the search method as "BestFirst" (which searches the space of feature subsets by greedy hill-climbing augmented with a backtracking facility). For the classifier used in the "AttributeSelectedClassifier", we selected SMO. Ten-fold cross-validation was performed. As a result, the length of the feature vector was reduced from 2321 to 52 for the CT dataset. The reduced feature vector length was 93 for the CRYO dataset. The cross-validation classification performance was measured using: TP (true positive) rate, FP (false positive) rate, precision, recall, F-score, and ROC (receiver operating characteristic) area. Table II and Table IV list the values of the above measures for the CT and CRYO datasets, respectively, as output by WEKA. The confusion matrix for each dataset is also provided in Table III and Table V, respectively. The numbers in these tables demonstrate the effectiveness of the proposed method. For example, as shown in Table VI, the overall classification accuracy is 98.8% for the CT dataset and 99.0% for the CRYO dataset.

[2] http://www.cs.waikato.ac.nz/ml/weka/

TABLE I. NUMBER OF SLICES IN EACH BODY SEGMENT

	Head & neck	Thorax	Abdomen	Pelvis	Lower limb
CT	182 (M) 233 (F) 415	88 (M) 194 (F) 282	208 (M) 171 (F) 279	73 (M) 213(F) 286	101 (M) 797 (F) 898
CRYO	255 (M)	262 (M)	221 (M)	219 (M)	913 (M)

TABLE II. CLASSIFICATION RESULTS (CT DATASET)

	TP rate	FP rate	Precision	Recall	F-score	ROC area
Head & neck	0.981	0.001	0.995	0.981	0.988	0.996
Thorax	0.979	0	1	0.979	0.989	0.998
Abdomen	1	0.006	0.959	1	0.979	0.997
Pelvis	0.965	0.001	0.993	0.965	0.979	0.994
Lower limb	0.997	0.009	0.988	0.997	0.992	0.995

TABLE III. CONFUSION MATRIX (CT DATASET)

		Predicted class				
		Head & neck	*Thorax*	*Abdomen*	*Pelvis*	*Lower limb*
Actual class	Head & neck	407	0	0	0	8
	Thorax	1	276	3	0	2
	Abdomen	0	0	279	0	0
	Pelvis	0	0	9	276	1
	Lower limb	1	0	0	2	895

TABLE IV. CLASSIFICATION RESULTS (CRYO DATASET)

	TP rate	FP rate	Precision	Recall	F-score	ROC area
Head & neck	0.996	0.001	0.996	0.996	0.996	0.999
Thorax	0.969	0.004	0.977	0.969	0.973	0.994
Abdomen	0.977	0.005	0.964	0.977	0.971	0.995
Pelvis	0.995	0.002	0.986	0.995	0.991	0.999
Lower limb	0.997	0	1	0.997	0.998	1

TABLE V. CONFUSION MATRIX (CRYO DATASET)

		Predicted class				
		Head & neck	*Thorax*	*Abdomen*	*Pelvis*	*Lower limb*
Actual class	Head & neck	254	1	0	0	0
	Thorax	1	254	7	0	0
	Abdomen	0	5	216	0	0
	Pelvis	0	0	1	218	0
	Lower limb	0	0	0	3	910

TABLE VI. CLASSIFICATION ACCURACY

	Overall accuracy
CT	98.8%
CRYO	99.0%

V. CONCLUSION

We propose a new method for classification of body axial cross section images by body region. Our approach exploits multiple descriptors to represent the color, texture, shape and size attributes of images. Our strategy is to begin with a wide range of features, then to use a standard feature selection method for reducing feature dimensionality, before invoking a multi-class SVM classifier which assigns a body region label to an image, based only on the features for that image. We used two whole body datasets created by the National Library of Medicine's visible human project to test our algorithm: CT cross sections and cryosectional color photographs. Our experiments show that the proposed method yields a classification accuracy of higher than 98% in 10-fold cross validation for both datasets. Our future work will focus on identifying the location of the body segment shown in the biomedical figures.

ACKNOWLEDGMENT

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We are grateful to David Chen in the Office of High Performance Computing and Communications (OHPCC) of NLM for the visible human data used in this study.

REFERENCES

[1] V. Spitzer, M. J. Ackerman, A. L. Scherzinger, and D. Whitlock, "The visible human male: A technical report," Journal of the American Medical Informatics Association, vol. 3, pp. 118-130, 1996.

[2] P. Beylot, P. Gingins, P. Kalra, N. Magnenat-Thalmann, W. Maurel, D. Thalmann, and J. Fassel, "3-D interactive topological modeling using visible human dataset," Proceedings of Eurographics, pp. C33–C44, 1996.

[3] C. Imielinska, D. Metaxas, J. Udupa, Y. Jin, T. Chen, "Hybrid segmentation of anatomical data," Lecture Notes in Computer Science, vol. 2208, pp. 1048-1057, 2001.

[4] Z. Krol, M. Chlebiej, P. Mikolajczak, and K. Hoffmann, "Visible human projects – a step towards the virtual patient: models and simulations," Journal of Medical Informatics and Technologies, vol. 7, pp. 51-58, 2004.

[5] J.A. Juanes, A. Prats, M.L. Lagandara, J.M. Riesco, "Application of the 'Visible Human Project' in the field of anatomy: a review," European Journal of Anatomy, vol. 7, no. 3, pp. 147-159, 2003.

[6] B. Cheng, S. Antani, R.J. Stanley, G.R. Thoma, "Graphical image classification using a hybrid of evolutionary algorithm and binary particle swarm optimization," Proceedings of SPIE Electronic Imaging, San Francisco, California, Jan 2012.

[7] M. Simpson, M. M. Rahman, S. Phadnis, E. Apostolova, D. Demner-Fushman, S. K. Antani, G. R. Thoma, "Text- and content-based approaches to image modality classification and retrieval for the ImageCLEF 2011 medical retrieval track," CLEF 2011 Working Notes, Amsterdam, Netherlands, September 2011.

[8] S.A. Chatzichristofis, Y.S. Boutalis, "CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," In:

Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) Proceedings of the 6th International Conference on Computer Vision Systems. Lecture Notes in Computer Science, vol. 5008, pp. 312-322, Springer- Verlag Berlin Heidelberg, 2008.

[9] S.A. Chatzichristofis, Y.S. Boutalis, "FCTH: Fuzzy color and texture histogram: A low level feature for accurate image retrieval," In: Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services, pp. 191-196, 2008.

[10] M. Lux, "Caliph & Emir: MPEG-7 photo annotation and retrieval," Proceedings of the seventeen ACM international conference on Multimedia, pp. 925-926, 2009, Beijing, China.

[11] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[12] M.M. Rahman, S.K. Antani, G.R.Thoma, "Bag of keypoints-based biomedical image search with affine covariant region detection and correlation-enhanced similarity matching," Proceedings of IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS), pp.261-266, Oct. 2010

[13] G. Zhao, M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915-928, 2007.

[14] P. Howarth, and S. Ruger, "Robust texture features for still image retrieval," IEEE Proceedings of Vision, Image and Signal Processing, vol. 152, no. 6, pp. 868-874, December 2005.

[15] G.N. Srinivasan, G. Shobha, "Statistical texture analysis," Proceedings of World Acad. Sci. Eng. Technol., vol. 36, pp. 1264–1269, 2008.

[16] H. Ming-Kuei, "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory, vol. 8, pp. 179-187, 1962.

[17] M. Stricker, and M. Orengo, "Similarity of color images," In SPIE Conference on Storage and Retrieval for Image and Video Databases III, vol. 2420, pp. 381-392, Feb. 1995.

[18] L.C. Molina, L. Belanche, A. Nebot, "Feature selection algorithms: a survey and experimental evaluation," Proceedings of IEEE International Conference on Data Mining, pp. 306- 313, 2002

[19] M. Aly, "Survey on multi-class classification methods," Caltech, USA, Technical Report, November 2005.

[20] C. Hsu, C. Lin, "A comparison of methods for multi-class support vector machines," IEEE Transactions on Neural Networks, vol. 13, pp. 415-425, 2002.

[21] http://svmlight.joachims.org/svm_multiclass.html

[22] http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multiclass.html

[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA data mining software: an update," SIGKDD Explorations, vol. 11, no. 1, pp. 10-18, November 2009.