**ISBN** 978-93-84468-85-9
**Proceeding of 2016 4th International Conference on Image Processing, Production and Computer Science (ICIPCS'16)**
*Phuket (Thailand)*, *Dec. 12-13, 2016*

# Automatic Detection of Biomedical Compound Figure using Bag of Words

Mohammad Reza Zare[1], Henning Müller[2]

[1]School of Information Technology, Monash University, Malaysia,
[2]University of Applied Sciences Western Switzerland (HES-SO) Valais, Sierre, Switzerland

***Abstract***: *Biomedical Figures in medical articles contain a large variety of information such as text, illustrations, and images produced from various modalities. These figures represent an important part of the knowledge stored in medical articles and can often be reused for learning, research, and clinical decision support. Exploiting these figures as part of a query in automatic retrieval of medical articles has become an increasingly important and active research area. The major challenge is that almost half of the biomedical figures in medical articles are compound figures, meaning that they contain more than a single figure type. These compound figures then need to be separated into sub-figures in order to be used as query image to retrieve similar cases or articles. Prior to this, a major step is to detect a compound figure. In this paper, a classification algorithm is proposed for compound figure detection. A Bag of Words (BoW) is employed as image representation technique followed by a Support Vector Machine (SVM) classifier to generate a classification model. This model is able to automatically detect if an image is compound figure. The best accuracy rate obtained by this model is 93.5%.*

***Keywords:*** *Compound Figure, Medical Articles, Bag of Words, Image Classification*

## 1. Introduction

Medical online resources such as PubMed Central (PMC) are regularly used by clinicians and medical researchers. It is a free digital repository and with PubMed a search engine exists that is used for biomedical test search. It contained in late 2015 over 3.5 million biomedical journal publications and also over 3.4 million associated images. Despite the features of PubMed (free and accessible to anyone from anywhere), it is a text-based search engine. There are indications that the search experience of such databases can be enhanced by incorporating images as part of the search [1, 2]. Inspired by the success of Content-Based Image Retrieval (CBIR) [3], the integration of images and text in a search engine has received a remarkable attention in biomedical article retrieval [4-6]. Images in biomedical publications constitute a considerable amount of information that can assist clinicians or medical researchers to evaluate potential usefulness of an article in a clinical situation at hand. Open-I is one example of s system combining text and visual search[1].

However, the image retrieval procedure in this context is different from existing CBIR systems; it is complicated by the fact that an estimated 40% of figures in medical scientific articles are compound figures [7,23]. Compound or multi-panel figures consist of several sub-figures, each showing sub images or results.

Thus, in order to employ the CBIR concept to retrieve medical articles also based on images, it is essential to distinguish the parts of compound figures that are relevant to the query image. To explain the usability of this idea, a look can be taken at the sample image given in Figure 1. It is a compound figure consisting of three sub-

[1] https://openi.nlm.nih.gov/

figures, and the parts are related in the context of an article. In an offline mode, the search engine should be able to detect this image as a compound figure and then separate it into subfigures. Subsequently, each subfigure is classified into modalities, annotated with caption parts and then linked to its respective text sections. This gives an opportunity to the user to retrieve associated articles precisely by finding specific subfigures that are being searched for.

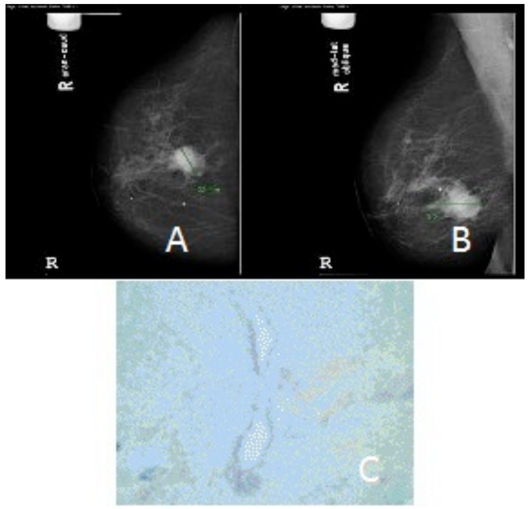


Fig. 1: Example of a compound figure consisting of two subfigures of a same time and one different part.

However, the very first challenging task is the detection of compound figures. The aim of this paper is to develop an intelligent algorithm to decide if the biomedical figures extracted from medical articles are compound figures or not. The rest of the paper is organized as follows: section 2 discusses the challenges existing in dealing with compound figure detection. The proposed detection algorithm is then presented in section 3. The experimental results are presented and analysed in section 4 followed by conclusion in section 5.

## 2. Challenges with Automatic Detection of Compound Figure

Automatic detection of compound figures in a set of biomedical images is a challenging problem in computer vision. In natural imagery, the analysis is mainly based on color. In many cases of natural imagery, having a color background is a dominant weight in characterizing image categories. In medical images such as X-ray images, even though the background is often black, most of the images from a particular category are visually very similar. On the other hand, general bio-medical figures contain a large variety of information such as text (case explications), illustrations (tables, charts, forms, etc.), and images produced by medical devices (X-ray, Magnetic Resonance Imaging, Computed Tomography, etc.).

As such, compound figures can be a combination of any of these modalities and non-compound can be either one of these modalities. There are some particular difficulties when dealing with automatic detection of compound figures as listed below.

### 2.1. Visual Similarity

As shown in Figure 2, there are cases where sub-image of compound figures that are visually similar. This could lead to misclassification as these sub-figures are difficult to distinguish for a single modality image when no clear boundary lines are visible.

Fig. 2: Sample Compound Figures with visual similarity among sub-figures and no separating lines.

### 2.2. Visual Dissimilarity

Unlike Figure 2, there are cases where images belonging to each category (compound & non-compound) are not visually similar as shown in Figure 3. The top row shows sample images from compound figure and the second row is selected sample images from non-compound figures.

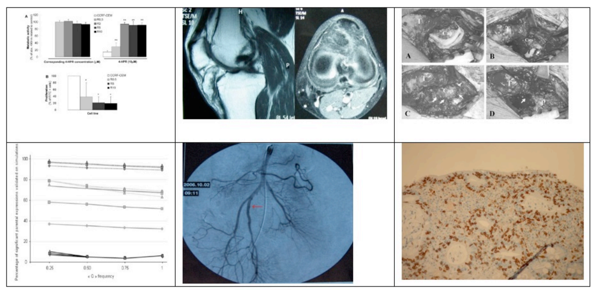

Fig. 3: Sample compound & non-compound figures with visual dissimilarity among them

### 2.3. Separation Line

Figure 4 is a compound figure consisting of three sub-figures without a separation line. Identifying whether this is a single figure or combination of three sub-figures is a challenging task for a machine.

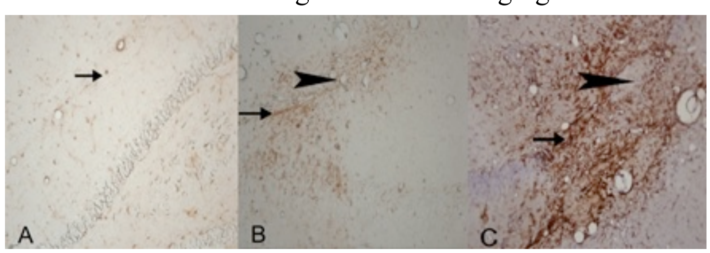


Fig. 4: Example of compound figure without separation line between sub-figures.

## 3. Methodology

This paper presents a method for compound figure detection using classification technique to identify whether a figure is compound or non-compound.

The classification task starts with feature extraction. Defining an appropriate feature extraction technique is one of the open challenges in this field. This representation must be robust enough to handle large variability of the biomedical figures. In addition, choosing the classifier techniques is also important in order to achieve the maximum classification accuracy.

In this experiment, various image representation techniques and classifiers were tested. Based on the results shown in Table 1, the Bag of Words (BoW) and Support Vector Machines (SVM) [8] are used as main image representation technique and classifier, respectively. In the past years, the BoW was successfully employed in various medical image retrieval and classification tasks [9-17]. Among other classifiers, SVMs have shown a better generalization performance in medical domain compared with other classification techniques. [9-11, 19-20]

### 3.1. Bag of Words Implementation

#### A. Detect and Extract Local Features

The Local keypoint (interest point) detection is the first step of the BoW operation. In this step, the relevant components of an image need to be identified. As such, low-level feature space must be partitioned into regions of the feature space that potentially correspond to visual topics. This can be done either by dense grid sampling or interest point. In this experiment, a Difference of Gaussians (DoG) [18] is used for local interest point detection. Upon detection of local interest points, SIFT features are used to represent the description of the local interest area detected by the DoG.

#### B. Vector Quantization

The feature vectors generated in the previous step are then divided into similar groups using a clustering or vector quantization technique to form a codebook. This step usually uses k-means clustering. It is one of the simplest but well known clustering algorithms. Given an image d with a set of features F(d)={f_j,j=1,…,N_(f(d)) }, the k-means algorithm performs clustering on these features vectors through a set of clusters k fixed a priori. The algorithm then randomly chooses k points in that feature vector that are used as the initial centres of the clusters. Then, the algorithm starts partitioning the feature space into N regions.

#### C. Codebook Construction

Once the cluster centres are identified, each feature vector in an image is assigned to a cluster centre using nearest neighbour with a Euclidean metric and finally each image is represented as histogram of these cluster centres by simply counting the frequency of the words appearing in an image.

## 4. Experimental Results and Discussion

In this section, a set of experiments were conducted to assess the performance of the classification algorithm on the ImageCLEF2015 Medical-Classification dataset [7].

The database used in this experiment consists of 20,000 biomedical figures from two categories: compound and non-compound figures. As discussed earlier, figures from each category vary from one another, each refers to a different image type. The labelled training dataset is used in this experiment. To increase the reliability of the results, the experiments are conducted five times, each on a different data set. The data (20,000 images) are divided into training and testing data. 70% are chosen randomly for training purposes and the remaining 30% are used for testing. As such 5 different training datasets and testing datasets were chosen for every experiment. The average of the classification acurracy obtained from every experiment is taken as final score.

To show the performance of the proposed approach, other experiments with various image representation and classification techniques were conducted on the same database with the same number of training and test data. In Table I, the total accuracy obtained using various feature extraction techniques is presented.

The classifier was also trained on various extracted BoW, differing based on the vocabulary sizes. This is to investigate how the classification performance is affected with regards to different vocabulary size as it is the main parameter in construction of BoW. Figure 5 shows the classification rate obtained with respect to various numbers of visual words. The best classification accuracy obtained is 93.5% where V in the visual vocabulary is 300.

The aim of this study was to classify biomedical images into two classes (compound and non-compound). To the best of our knowledge, only little research has been published for this purpose as this task is one of the latest initiatives in ImageCLEF 2015 [7]. Even though the classification accuracy obtained is promising for a database with such complexity, it is believed that there is still a lot of room for improvement.

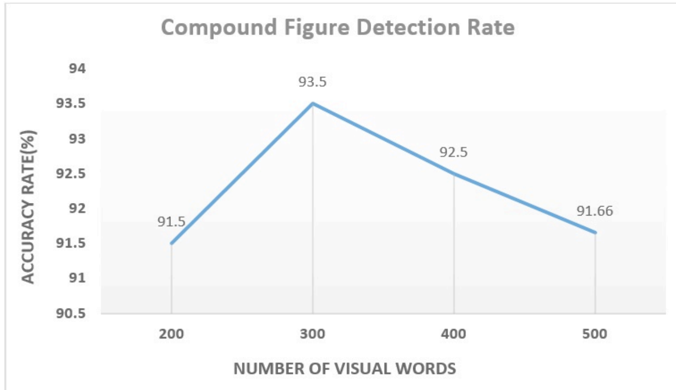

Fig. 5: Classification results with respect to various visual vocabulary sizes.

TABLE I: Compound figure detection results with various image representation techniques

| Feature Extraction | Classification Techniques | Accuracy Obtained |
|---|---|---|
| Gray Level Co-occurrence Matrix (GLCM) [21] | SVM with RBF | 84.5 % |
| GLCM | K-Nearest-Neighbor(KNN), k=9 | 82.0 % |
| Local Binary Pattern (LBP) [22] | SVM with RBF | 89.7 % |
| Local Binary Pattern | KNN, k=9 | 87.5 % |
| Bag of Visual Words (V=200) | SVM with RBF | 91.5 % |
| Bag of Visual Words (V=200) | KNN, k=9 | 89.2 % |
| Bag of Visual Words (V=300) | SVM with RBF | 93.5 % |
| Bag of Visual Words (V=300) | KNN, k=9 | 91.7 % |
| Bag of Visual Words (V=400) | SVM with RBF | 92.5 % |
| Bag of Visual Words (V=400) | KNN, k=9 | 92.1 % |
| Bag of Visual Words (V=500) | SVM with RBF | 91.6 % |
| Bag of Visual Words (V=500) | KNN, k=9 | 91.5 % |

Detailed analysis on the classification results shows that misclassified images mostly belong to compound figures. Further observation on these images revealed the following reasons of misclassification. Figure 6 shows two sample images from com-pound figure that are misclassified as non-compound. Figure 6(A) and 6(B) consists of 2 and 4 sub-figures, respectively.

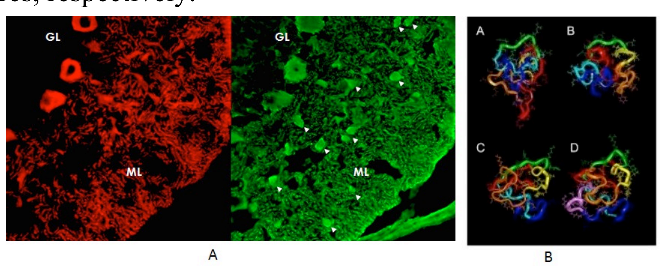

Fig. 6: Sample misclassified images due to similarity among subfigure

As it can be seen, the sub-figures are visually similar in both cases and without a separating line, which makes it difficult for a classification model to distinguish these images as compound figure. This could be due to the similarity exist in their extracted features. In addition, there is no clear separation line between sub-figures which can also lead to misclassification.

A compound figure containing 2 graphs is shown in Figure 7. As it can be observed, even though the sub-figures are not visually similar, yet the classification model could not detect it as a compound figure. The reason is that the two sub-figures do not contain any identifiable space gap.

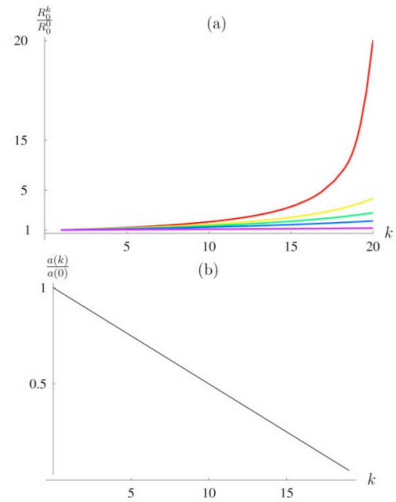


Fig. 7: Sample misclassified images due to lack of uniform space gap.

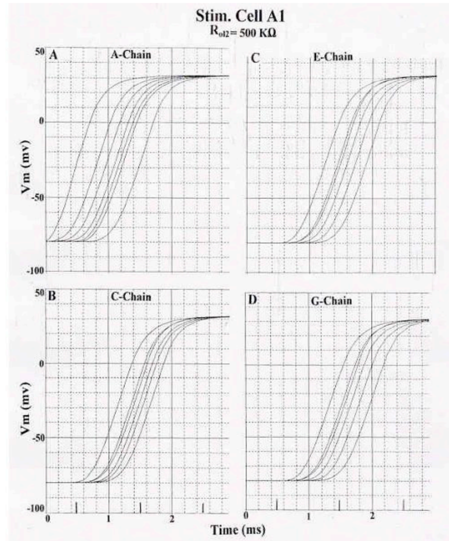


Fig. 8: Sample classification result affected by different vocabulary sizes.

It was also observed that changing the vocabulary size in construction of BoW has an impact on classification result. As a good example, Figure 8 shows a compound figure that is misclassified as non-compound with vocabulary size of 300 (V=300) and correctly classified with vocabulary size of 200 (V=200).

V=300 leads to have several visual words representing the same local content. On the other hand, once it is reduced to 200, a visual word representing larger regions of the feature space, making the visual words less specific but making the image representation more stable across similar images. However, this could vary from image to image due to ambiguous data representation generated by BoW.

In the BoW, an image is represented as an order less collection of local features. As such, one set of these local features may denotes different parts on different object categories and similarly, two different sets of these local features may represent a similar part of an object. Such ambiguous representations lead a classification to be a challenging task. Another challenge with the construction of the BoW is ignoring all information about the spatial layout of the features that may contain useful cues for image classification. These limitations are bold especially when the recognition system deals with heavy occlusion, clutter as well as view point changes that exist in medical compound figure databases. Approaches based on a generative model that is proven to be effective on classification of medical databases [10], will be employed in future work to disambiguate the BoW representation. Further observation of the compound figures in the given dataset revealed the fact that almost half of the sub-figures include associated panel labels such as A, B, C or D. Extraction of these labels could possibly enhance the classification performance especially for the cases with the above mentioned complexities. Various algorithms of deep learning will be also explored in the future in extended work.

## 5. Conclusion

Medical scientific journals are enriched with a vast amount of figures that contain comprehensive information. In this article, the use of these figures to enhance the search results of medical article retrieval system is discussed. The usage of figures in process of such retrieval system is a complex task. This is due to the fact that these figures are compound figures. The research work presented in this article was to develop a classification framework to automatically distinguish compound figure from non-compound figure. In this experiment, the classification task started with extracting a BoW as an image representation technique. The extracted features are then used as an input to SVM classifier to construct a classification model. Despite of the significant accuracy rate obtained by the proposed classification approach, some of the possible future work is identified to address the complexities exist in classification of such medical database.

# 6. References


[1] M.A. Hearst, A. Divoli, J. Ye, M.A. Wooldridge, "Exploring the Efficacy of Cap-tion Search for Bioscience Journal Search Interfaces." In: Proceedings of the Work-shop on BioNLP 2007: Biological, Translational, and Clinical Language Processing, BioNLP '07, pp. 73-80. Association for Computational Linguistics, Stroudsburg, PA, USA (2007)

[2] R. Sandusky, C. Tenopir, "Finding and Using Journal Article Components: Impacts of Disaggregation on Teaching and Research Practice". Journal of the American Society for Information Science and Technology, vol. 59(6), 2008, pp. 970-982

[3] S. Antani, L.R. Long, G.R. Thoma, "Content-Based Image Retrieval for Large Biomedical Image Archives. Studies in Health Technology and Informatics, vol. 107(Pt2), 2004, pp. 829-833

[4] E. Apostolova, D. You, Z. Xue, S. Antani, D. Demner-Fushman, G.R. Thoma, "Image retrieval from scientific publications: Text and image content processing to separate multipanel figures. JASIST, vol. 64(5), 2013, pp. 893-908

[5] L.D. Lopez, J. Yu, C.N. Arighi, C.O. Tudor, M. Torii, H. Huang, K. Vijay-Shanker, C.H. Wu, "A framework for biomedical figure segmentation towards image-based document retrieval. BMC Systems Biology, vol. 7(S-4), 2013, pp. S8-S23

[6] S. Antani, D. Demner-Fushman, J. Li, B.V. Srinivasan, G.R. Thoma, " Exploring use of images in clinical articles for decision support in evidence-based medicine. In: Document Recognition and Retrieval XV, DRR 2008, 15th Document Recognition and Retrieval Conference, part of the IS&T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 29-31, 2008. Proceedings, pp. 68150 (2008)

[7] A. Garcia Seco de Herrera, H. Müller, S. Bromuri, Overview of the ImageCLEF 2015 medical classification task, Working Notes of CLEF 2015 – Cross Language Evaluation Forum, CEUR, 2015.

[8] C. Cortes, V. Vapnik " Support-vector networks. Mach. Learn., vol. 20(3), 1995, pp.273-297

[9] M.R. Zare, A. Mueen, C.S. Woo,"Automatic medical x-ray image classification using annotation". Journal of Digital Imaging, vol. 27(1), 2014, pp. 77-89

[10] M.R. Zare, A. Mueen, M. Awedh, W.C. Seng, "Automatic classification of med-ical x-ray images: hybrid generative-discriminative approach." Image Processing, IET, vol. 7(5), 2013, pp.523-532

[11] U. Avni, H. Greenspan, E. Konen, M. Sharon, J. Goldberger ," X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words." Medical Imaging, IEEE Transactions on, vol. 30(3), 2011, pp. 733-746

[12] T. Deselaers, D. Keysers, H. Ney, " Features for image retrieval: An experimental comparison." Inf. Retr., vol. 11(2), 2008, pp. 77-107

[13] T. Tommasi, F. Orabona, B. Caputo, " Discriminative cue integration for medical image annotation." Pattern Recogn. Lett., vol. 29(15), 2008, pp. 1996-2002

[14] I. Dimitrovski, D. Kocev, S. Loskovska, S. Deroski,"Hierarchical annotation of medical images. Pattern Recognition, vol. 44(1011), 2011, pp. 2436-2449

[15] I. Sadek, D. Sidibe, F. Meriaudeau, "Automatic discrimination of color retinal images using the bag of words approach." In: Proc. SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis, 94141J, pp. 8 (2015)

[16] L. Shen, J. Lin, S. Wu, S. Yu, " Hep-2 image classification using intensity order pooling based features and bag of words. Pattern Recognition, vol. 47(7), 2014, pp. 2419 -2427

[17] I. Dimitrovski, D. Kocev, I. Kitanovski, S. Loskovska, S. Deroski, "Improved medical image modality classification using a combination of visual and textual features." Computerized Medical Imaging and Graphics, vol. 39(0), 2015, pp. 14 – 26

[18] D.G. Lowe, "Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2, ICCV '99, pp. 1150-, IEEE Computer Society, Washington, DC, USA (1999)

[19] H. Pourghasem, H. Ghasemian, "Content-based medical image classification using a new hierarchical merging scheme", Computerized Medical Imaging and Graphics, vol. 32 , 2008, pp.651–661

[20] M. Rahman, S. Antani, G. Thoma, "A Learning-Based Similarity Fusion and Filtering Approach for Biomedical Image Retrieval Using SVM Classification and Relevance Feedback. Information Technology in Biomedicine, IEEE Transactions on, vol. 15(4), 2011, pp. 640-646

[21] R. M. Haralick, K. Shanmugam, I.H. Dinstein," Textural Features for Image Classification. Systems, Man and Cybernetics", IEEE Transactions on, vol. 3(6), 1973, pp. 610-621.

[22] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rota-tion invariant texture classification with local binary patterns." Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24(7), 2002, pp. 971-987

[23] A. Chhatkuli, D. Markonis, A. Foncubierta, F. Meriaudeau, H. Müller, Separating compound figures in journal articles to allow for subfigure classification, SPIE Medical imaging 2013, Orlando, FL, USA.