Ontology-based Automatic Reclassification of Tissues and Organs in Histological Images

Claudia Mazo[1,2,3], Maria Trujillo[1], Enrique Alegre[4], and Liliana Salazar[5]

[1] Universidad del Valle, Computer and Systems Engineering School, Cali, Colombia
claudia.mazo@correounivalle.edu.co
[2] University College Dublin, CeADAR: Centre for Applied Data Analytics Research, School of Computer Science, Dublin, Ireland
[3] OncoMark Limited, Dublin, Ireland
[4] Universidad de León, Industrial and Informatics Engineering School, León, Spain
[5] Universidad del Valle, Morphology Department, Cali, Colombia

Abstract. Heterogeneous data source produces different types of data that cannot be treated in the same way. In this paper, two sources of data are considered: image and human knowledge. The former is represented using visual descriptors and the latter is represented using an ontology. The integration of these data sources is used in the automatic classification of tissues and organs of the human cardiovascular system together. Firstly, visual descriptors – texture descriptors – are used in the automatic classification using a cascade Support Vector Machine. Secondly, obtained classification results are refined using a histological ontology of the human cardiovascular system to confirm or reclassified. The final classification results are more precise than the obtained using only image data, in all cases.

Keywords: Automatic Classification · Histological Ontology · Histology Images· Image Processing.

1 Introduction

On the one hand, a computer vision problem consists of identifying the fundamental tissues and recognising distinctive patterns, formed by spatial structures among them, in order to infer an organ. The solution of this problem may be used to reinforce learning processes and translate into better-formed professionals without requiring any mayor social or economic investment [3]. On the other hand, histological knowledge representation is used to solve complex tasks, such as: support teaching and medical practices or have natural language interactions, which are challenges. In fact, multiple and heterogeneous data sources produce different types of representations of data that cannot be treated in the same way, being an open problem. In this paper, we present a method that combines the automatic classification based on texture descriptors with the histological ontology in order to improve the classification results.

Ontologies and taxonomies contain relevant knowledge represented with rich structural and semantic information. Approaches that use those knowledge in an automatic classification process are twofold: (i) model the relation between visual and semantic information [9] [10] and (ii) use those ontologies and taxonomies in the classification algorithm [6] [8] [1] [7] [2]. We will focus in the second group to perform the classification process using images and an ontology, after creating a histological ontology [5].

In this paper, we describe the use of an ontology to refine the results of a classification based on image data. We propose two different methods: (i) refining an organs classification and (ii) a classification of epithelial tissue. The proposed approaches allow to reduce the classification error, in all cases.

2 Method

Our proposal consists of three steps: (1) Input: histological image, along with histological and expert knowledge are used as input source. (2) Image processing: given a histological image, the block-based classification method, proposed in [4], is used in the classification of individual blocks. Classified blocks are concatenated in a way to represent an image. (3) The histological knowledge representation: the ontology, presented in [5], is used to refine classification results. We propose two methods, as follows: (3.1) refining organ classification and (3.2) recognising epithelial tissue. The refining processes are described as follows.

2.1 Refining organ classification

A given image is divided into blocks and each block is classified in one of six classes. Four classes are considered discriminative classes, whilst two classes are non-discriminative. Discriminative classes are associated to organs.

Afterwards, classification results are refining by a SPARQL's query, based on RDF triples derived from the frequencies per discriminative class. The RDF's subject is the organ with higher frequency among the classes. The RDF's objects are the remaining classes. RDF triples are built with the predicate *hasPresenceOf*. If the obtained result is empty, then blocks classified in the organ class, used as object in the RDF triple, should be reclassified. A new label will be decided according to the behaviour of false positives in the classification process. In other case, the classification is confirmed and the organ class is not modified.

2.2 Refining epithelial tissue classification

The two non-discriminative classes are used for identifying epithelial tissue, using as input the concatenated blocks forming an image. Firstly, large image regions, classified as light regions, are selected using a threshold value, selected heuristically, in order to decide if there is epithelial tissue. Secondly, if a light region is large enough, there is not loose connective tissue in the neighbourhood. Then, blocks on the neighbourhood are verified not to be labelled as loose connective. Thirdly, RDF triples are built using frequencies of discriminative and non-discriminative classes. The RDF's subject is the organ with higher frequency among the discriminative classes; the object is *TejidoEpitelialRevestimiento* or *EpithelialLining*; and the predicates are *someValuesFrom* and *subClassOf*. The possible results are: (i) the coating epithelial or the lining epithelial is in the subject. In this case, blocks surrounding light region and between muscle region are labelled as the epithelial tissue. (ii) an empty result means that it is highly probably non-presence of epithelial tissue. This result confirm the classification and the blocks are not modified.

3 Evaluation and Results

Tissue samples from organs were stained with Hematoxylin and Eosin. Initially, 1500 blocks belonging to five histological images of different organs and persons acquired at 10× objective were manually labelled. *Img-He* and *Img-He1* represent the heart images; *Img-MA* represents the muscular artery image; *Img-EA* represents the elastic artery images; and *Img-LV* represents the large vein image. We left this dataset publicly available at http://biscar.univalle.edu.co/datasets. Algorithms were implemented in C++, using the CImg library, SPARQL and Protégé in a computer of 8-cores and 8Gb of RAM.

Results of the refining classification process were evaluated based on the True Positive (TP) and the False Positive (FP). Figure 1(a) contains the total of TP and FP obtained using the initial and the refined classifications, by organ. It can be observed that the TP is increased after the refining. The highest increment in the TP is obtained with the elastic artery. Figure 1(b) contains a graphical representation of the TP and the FP by blocks obtained after identifying epithelial tissue in the set of test images. The reclassified blocks correctly classified between 0 to 7 per image and the recognition of epithelial tissue process increases TP rate classification between 0% and 2.333% according to the area which contain epithelial tissue. It is important to highlight that epithelial tissue regions occupy a smaller proportion in histological images, for this reason rates of improvement are highly variable between images and less than 3%. Additionally, the behaviour after the epithelial recognition process corresponds to increasing TP in images with epithelial tissue.

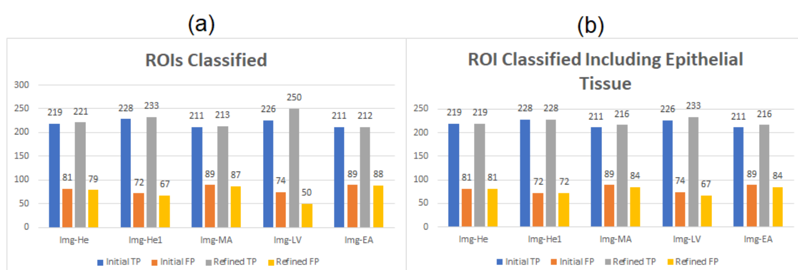

Fig. 1. Results (a) Block diagram of the TP and the FP of the initial and the refined classifications, by organ. (b) TP and FP blocks with automatic classification and improvement process.

4 Conclusions

Our refinement proposal enables us to obtain more consistent information and to reduce margins of error and uncertainty through corroboration and verification, by comparing analysis of different data sources separately. Besides, more information of histological knowledge as a system, a composition, but also as structures, relations, regions, layers, sectors, tissues and cells is obtained or inferred from an image. Epithelial tissue is identifies in 10× magnification images that cannot be done manually. As future work, automatic identification of micro-circulation organs using macro-circulation identified in this work could be proposed.

5 Acknowledgment

This project has received funding from the EI and from the European Union's Horizon 2020 research and innovation programme under the Marie Slodowska-Curie grant agreement No 713654

References

1. Abdollahpour, Z., Samani, Z.R., Moghaddam, M.E.: Image classification using ontology based improved visual words. In: 2015 23rd Iranian Conference on Electrical Engineering. pp. 694–698 (May 2015)
2. Breen, C., Khan, L., Kumar, A., Wang, L.: Ontology-based image classification using neural networks (2002)
3. Izet, M.: E-learning as new method of medical education. Acta Informatica Medica **16**(2), 102–117 (2008)
4. Mazo, C., Alegre, E., Trujillo, M.: Classification of cardiovascular tissues using lbp based descriptors and a cascade svm. Computer Methods and Programs in Biomedicine **147**, 1 – 10 (2017)
5. Mazo, C., Salazar, L., Corcho, Ó., Trujillo, M.P., Alegre, E.: A histological ontology of the human cardiovascular system. J. Biomedical Semantics **8**(1), 47:1–47:15 (2017)
6. Othmani, A., Meziat, C., Loménie, N.: Ontology-driven image analysis for histopathological images. In: Advances in Visual Computing - 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29-December 1, 2010. Proceedings, Part I. pp. 1–12 (2010)
7. Paulson, P., Hohimer, R., Doucette, P., Harvey, W., Seedahmed, G., Petrie, G., Martucci, L.: A methodology for integrating images and text for object identification. In: Prospecting for Geospatial Information Integration. pp. 1057–1066 (2006)
8. Smith, B., Arabandi, S., Brochhausen, M., Calhoun, M., Ciccarese, P., Doyle, S., Gurcan, M.: Biomedical imaging ontologies: A survey and proposal for future work. Journal of Pathology Informatics **6**(37) (2015)
9. Wu, Z., Jiang, S., Li, L., Cui, P., Huang, Q., Gao, W.: Vicept: link visual features to concepts for large-scale image understanding. In: Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010. pp. 711–714 (2010)
10. Yang, J., Jiang, Y.G., Hauptmann, A.G., Ngo, C.W.: Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval. pp. 197–206. MIR '07, ACM (2007)