Recognition of Cancer using a Bag-of-Words Random Forest for Gastroenterology

S. Francisco[1,2], Miguel T. Coimbra [2,5], R. Gamelas Sousa [3,4]

[1] Faculdade de Engenharia, Universidade do Porto, Portugal.
[2] Instituto de Telecomunicações, Porto, Portugal.
[3] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
[4] Instituto de Engenharia Biomédica (INEB), Porto, Portugal
[5] Faculdade de Ciências, Universidade do Porto, Portugal

In gastroenterology, Computer Aided Diagnosis techniques have allowed physicians to analyze endoscopic images as a first or second opinion, or even as an educational program [4]. Cancer recognition in the gastroenterology track is such a difficult problem that only trained physicians can easily detect. Some pattern recognition solutions have already been published in the past [1-4]. However, these solutions have to be invariant to acquisition constraints (rotation, scale and luminance), goals which are not always achieved.

We extract features from pre-processed endoscopic images, extracted pixel intensities of some regions, and then use Random Forests to create a Bag of Feature Words (BoW). The BoW is built by counting how many times each feature appears on each image, thus giving its information content, and used to build a classification model using Support Vector Machines.

Our dataset 176 chromoendoscopy images of the oesophagus (30 for training and the remainder for testing), exhibiting three different conditions (normal and the pathological tissues with dysplasia and metaplasia). After 50 runs, we achieve $22.03 \pm 3.09\%$ of average error. This results are promissory, especially when compared to the usage of SIFT descriptor and k-means ($32.20 \pm 3.58\%$) to generate the BoW for classification.

Our methodology seems to deal well with few data train samples and it is statistically different and better then the results of the existent methodologies.

References:

[1] Sousa, R et al.., Moura, D.C., Ribeiro, M.D. and Coimbra, M.T. (2014), *Local self similar descriptors: Comparison and application to gastroenterology images*, Engineering in Medicine and Biology Society (EMBC), pp. 4635-4638.

[2] Riaz, F. et al., Silva, F.B., Ribeiro, M.D., and Coimbra, M.T. (2012), *Invariant gabor texture descriptors for classification of gastroenterology images*, Biomedical Engineering, IEEE Transactions, v. 59, n. 10, pp. 2893-2904.

[3] Gross, S., Trautwein, C., Behrens, A., Winograd, R., Palm, S., Lutz, H.H., Sokhan, R.S., Hecker, H., Aach, T. and Tischendorf J.J.W.et al. (2011), *Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification*, Gastrointestinal endoscopy 74.6, pp. 1354-1359.

[4] Balbaque, F.S., Ribeiro, M.D., Rabenstein, T., Goda, K., Kiesslich, R., Harigsma, J., Edebo, A., Toth, E., Soares, J., Areia, M., Lundell, L. and Marschall, H.U. et al.(2013), *Endoscopic assessment and grading of Barrett's esophagus using magnification endoscopy and narrow band imaging: Impact of structured learning and experience on the accuracy of the Amsterdam classification system*, Scandinavian journal of gastroenterology, v. 48, n. 2, pp. 160-167.