Temporal Segmentation and Quality Assessment of Digital Colposcopies

Kelwin Fernandes[1,2], Jaime S. Cardoso[1,2], Jessica Fernandes[3]
[1]University of Porto, Porto, Portugal; [2]INESC-TEC Porto, Portugal (kafc@inesctec.pt); [3]Universidad Central de Venezuela, Caracas, Venezuela

Abstract
Cervical cancer remains a significant cause of mortality in low-income countries. Digital colposcopy is a promising and inexpensive technology for the detection of cervical intraepithelial neoplasia. However, diagnostic sensitivity varies widely depending on the doctor expertise. Colposcopies cover four stages: macroscopic view, observation under green light, Hinselmann and Schiller. We focus on the temporal segmentation of the video in these steps and in the ordering of colposcopic images according to their quality. Using a KNN classifier we achieved a precision of 97% in the temporal segmentation. We obtained an accuracy of 76% in the order prediction of frames using SVM.

Author Keywords. Cervical cancer, colposcopic images, histogram distances, temporal segmentation, ranking.

1. Introduction

Digital colposcopy is a promising and inexpensive technology for the detection of cervical intraepithelial neoplasia. However, diagnostic sensitivity varies widely depending on the doctor expertise (IFGO 2009). According to the World Health Organization protocol, detection of preinvasive cervical lesions during a colposcopic screening covers the following steps (see Figure 1): macroscopic view with magnifier white light, observation under green light and exposure to acetic acid solution (Hinselmann) and potassium iodine (Schiller) (IFGO 209). Throughout the procedure, the expert disturbs the cervix area at different times.

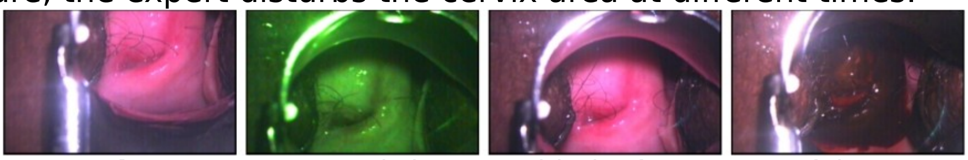

Figure 1: Colposcopic images with the four stages of the colposcopic screening

In order to decide the right diagnostic, the decision maker (either a physician or an automatic tool) needs to select a subset of good quality frames from each stage. The feedback obtained in these Quality Assessment (QA) tasks is usually defined in a numeric scale by subjective ratings (Winkler 2012), which suffers from noise because of inter-human inconsistencies and limitations of the underlying scale. A more natural way to define quality would be induced by a preference function whereby it is defined which object (if any) is the best within an object pair.

2. Materials and Methods

We assume that *transitions* correspond to frames with high motion. Therefore, frames with large Euclidean pixel-wise distance with their adjacent frames were filtered (Fernandes et al. 2015). Then, frames are individually labeled following the protocol stages using a K-Nearest Neighbors model, describing each colposcopic image by their one-dimensional hue and saturation histograms (Fernandes et al. 2015). Similarity between two images is defined by the mean distance -L1 or Earth Mover's Distance- between their histograms. Then, *temporal boundaries* between the diagnosis steps were stablished by modeling the problem as the minimum cost recognition of a word by a Weighted Finite Automaton (Fernandes et al. 2015). For the QA we labeled pairs of frames according to their diagnosis-based quality. We extracted features related to image edges, specular reflection, color statistics, distance to transition intervals and cervix area. Finally, we trained a Support Vector Machine fed with the binary order relation between the features of both frames.

3. Discussion

We gathered a dataset of 56 colposcopies from different patients (143640 images) annotated by a specialist. Every patient was equally weighted in the compilation of the results. For the assessment of the temporal segmentation we used a *leave-one-patient-out cross-validation* approach. Transition frames were correctly identified with an accuracy of 90.86%. Table 1 shows the classification metrics for the temporal segmentation. For the evaluation of the order prediction we annotated 798 pairs of images and performed experiments using a 10-fold cross-validation. We obtained an accuracy of 76.2% using a RBF kernel.

Distance	**Transition**				**Non-Transition**			
	Acc.	**Prec.**	**Rec.**	**F1**	**Acc.**	**Prec.**	**Rec.**	**F1**
L1	0.94	0.94	0.71	0.76	0.95	0.97	0.93	0.90
EMD	0.93	0.91	0.67	0.70	0.92	0.94	0.87	0.83

Table 1: Avg. classification metrics for the temporal segmentation

4. Conclusions

In this work we provided a framework to temporarily segment a colposcopic assessment. The proposed framework achieved a precision of 91.46% in the transition detection and a precision of 96.65% in the temporal segmentation. As we observed in the experiments, L1 distance behaved better than the EMD. Also, we obtained preliminary results in the QA task by means of pairwise ordering of frames.

References

International Federation of Gynecology and Obstetrics (IFGO). 2009. *“Guía global para la prevención y control del cáncer cervicouterino“*. Technical report.

Winkler, Stefan. 2012. *“Analysis of public image and video databases for quality assessment*”. In *Selected Topics in Signal Processing, IEEE Journal of* 6, no. 6 , 616-625.

Fernandes, K., Cardoso, J. and Fernandes, J. 2015. "*Temporal Segmentation of Digital Colposcopies*". In Iberian Conference on Pattern Recognition and Image Analysis. Springer (accepted).