Temporal Segmentation of Digital Colposcopies

Kelwin Fernandes[1,2(✉)], Jaime S. Cardoso[1,2], and Jessica Fernandes[3]

[1] INESC TEC, Porto, Portugal
[2] Universidade do Porto, 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. Therefore, automation of this process is needed in both, detection and visualization. Colposcopies cover four steps: macroscopic view with magnifier white light, observation under green light, Hinselmann and Schiller. Also, there are transition intervals where the specialist manipulates the observed area. In this paper, we focus on the temporal segmentation of the video in these steps. Using our solution, physicians may focus on the step of interest and lesion detection tools can determine the interval to diagnose. We solved the temporal segmentation problem using Weighted Automata. Images were described by their chromacity histograms and labeled using a KNN classifier with a precision of 97 %. Transition frames were recognized with a precision of 91 %.

Keywords: Cervical cancer · Colposcopic images · Histogram distances · Temporal segmentation · Weighted finite automata

1 Introduction

Despite the possibility of prevention with regular cytological screening, cervical cancer remains a significant cause of mortality in low-income countries. This being the cause of more than half a million cases for year, and killing more than a quarter of a million in the same period [1].

Digital colposcopy is a promising and inexpensive technology for the detection of cervical intraepithelial neoplasia. The diagnostic sensitivity with these resources ranges between 67 to 98 %, depending on the expertise of the doctor [1]. The resection of lesions in the first visit could reduce the costs involved in a scheme of successive visits. Also, it would ensure the appropriate care of patients with poor adherence to treatment.

According to the protocol proposed by the World Health Organization (WHO) [1], detection of preinvasive cervical lesions during a colposcopic screening covers the following steps (see Fig. 1): macroscopic view with magnifier white light, followed by observation under green light for diagnosis of aberrant vascularization

R. Paredes et al. (Eds.): IbPRIA 2015, LNCS 9117, pp. 262–271, 2015.
DOI: 10.1007/978-3-319-19390-8_30

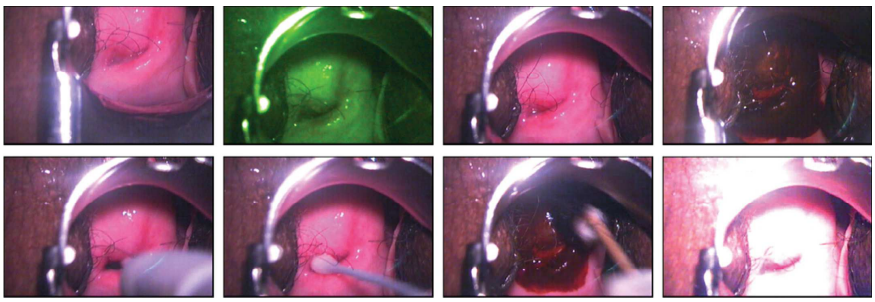

Fig. 1. Top: Diagnosis steps. From left to right: macroscopic observation, green filter, Hinselmann and Schiller. **Bottom:** Transition frames. The first three frames have occlusions of the cervix area and the last one presents a strong illumination difference after removing the green filter (Color figure online).

and then evaluate the cervical characteristics after exposure to acetic acid solution (Hinselmann) and potassium iodine (Schiller) [1]. Although Hinselmann and macroscopic observation cannot be differentiated on healthy patients, these two steps can be distinguished using contextual information. Throughout the procedure, the expert disturbs the cervix area to achieve better focus, to move from one step to the next, to clean the cervix area, etc. Figure 1 shows four transition frames. These scenes do not bring useful information for the diagnosis and should not be considered in the detection of lesions.

The goal of the project is to provide a more effective tool for the diagnosis of pre-invasive lesions, for environments with different resource availability and with different training staff. Our aim is to develop a diagnostic tool that can automatically identify neoplastic tissue from digital images. During the first phase of the study, we aim to achieve automatic recognition of each of the phases mentioned in the colposcopic study of a patient, in order to fine tune the diagnosis of cervical lesions. The temporal segmentation generated by our tool can be used by further techniques to detect lesions.

2 State-of-the-art

Colposcopic image processing has been a topic of interest in the last decade among computer vision researchers [2–7]. These works cover different topics from preprocessing and region segmentation [2,4], specularity removal [2,4] to computer-aided diagnosis systems that partially handle the WHO protocol [5–7].

Das *et al.* proposed a specular reflection (SR) detection algorithm based on the intensity level of the three RGB channels and reconstruct these areas using a smooth interpolant that fills the damaged area [2]. Then, they segment the cervix area using the K-means clustering algorithm. Gordon *et al.* segment the cervix region using unsupervised clustering via Gaussian Mixture Modeling [4]. They use as features the a channel of the *CIE - Lab* color space and the distance of a pixel to the center of the image. Then, they apply ad hoc rules to detect the clusters that represent the ROI. Roubakhsh *et al.* [3] extracted a set

of features obtained by correspondence analysis (CA), color gradient, statistics measurements of the color histograms (e.g. skewness, energy) and the red level of the image. Then, they fed these features into a Neuro-Fuzzy classifier. Alush and Goldberger detect cervical lesion by extracting features from the intensity of the edges of the regions obtained by the Watershed transform [5]. After that, they created a dictionary of instances using a clustering algorithm. A different approach for lesion detection was carried out by Park *et al.* [6] who extracted features from the relation between the values of each RGB channel before and after the application of acetic acid. In the classification stage, they used an ensemble of classifiers. Finally, Acosta *et al.* built time series using the intensity change after the application of acetic acid [7]. They fitted a parabola to the time series and, using a Naive Bayes classifier, determined the lesions according their severity degree.

Although these works report good results in the lesion detection, they focus only in the Hinselmann step of the WHO protocol. To the best of our knowledge, there is no previous work on the development of a platform that covers the whole protocol. Given that our final goal is to detect lesions in any stage of the protocol, this work focuses on the temporal segmentation of the different aforementioned steps.

3 System Overview

An automated system is proposed in this paper to segment the different steps of the colposcopic assessment. Our system can be splitted in three stages: transition removal, diagnosis-step classification (frame labeling) and temporal segmentation. Figure 2 illustrates this process.

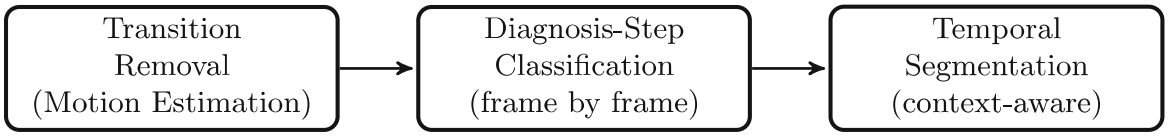

Fig. 2. Flow chart describing the proposed framework.

In general, the temporal segmentation problem implies finding simultaneously the segments and the labels. It is a hard problem which can be addressed sequentially. In this work, we take advantage of the knowledge domain by labeling the frames without considering the context and then, minimizing the temporal inconsistencies using the WHO protocol definition. The labeling is done by classification using templates from previous colposcopies and the temporal segmentation is done by translating the colposcopic procedure to a non-deterministic weighted automaton, which can be implemented using Dynamic Programming (DP). The temporal boundaries optimization tries to reach maximal consistency with the preassigned labels. The remainder of this section details the proposed system.

3.1 Transition Removal

In order to remove transition scenes, we adopted a motion-based approach. We assume that transitions correspond to frames with high motion. First, we apply a Gaussian blur to atenuate noisy pixels. Then, motion is estimated using the Euclidean pixel-wise distance between a frame and its neighborhood (W frames before and after the given frame). Finally, we apply a thresholding operator to differentiate between transition and non-transition frames. Equation (1) shows the formula that determines if a given frame belongs to a transition. Therein, I_i stands for the $i-th$ frame of the sequence I. Although more advanced approaches could be implemented, this standard procedure attained already a very good performance.

$$Transition(i) = \left(\frac{1}{2W} \sum_{w \in [-W..W]} \| \mathcal{G}(I_{i+w}) - \mathcal{G}(I_i) \|_2 > \ threshold \right) \quad (1)$$

3.2 Diagnosis Step Classification

Each colposcopic image is represented by its one-dimensional hue histogram and saturation histogram. In order to efficiently reduce the presence of noisy objects in the boundaries of the image, we masked the region of interest by removing everything outside a image-centered circle (with diameter equal to 0.75 of the image side). This approach considers that the cervix region occupies more than half of the cervigram image [2] and that it is approximately centered.

The diagnosis step classification is done in a per-frame basis. We propose a classification method based on K-Nearest Neighbors (KNN). The similarity between two images is defined by the average distance between their histograms. We compared three histogram distances. The bin-to-bin Minkowski distance of order 1 (L_1), which is equivalent to the Histogram Intersection distance [8] and the cross-bin distances: Earth Mover's distance (EMD) [8] and Circular Earth Mover's distance ($CEMD$) [9]. Given the huge amount of images and the low intra-variance between image within the same video, we indexed an equally spaced subset of images in the KNN knowledge base. Each video contains the same number of images per phase in order to avoid oversampling and bias. We smooth the labels by selecting the mode of a local window.

3.3 Temporal Segmentation

Finally, we have to decide the temporal boundaries between the diagnosis steps. For this purpose, let's generalize the problem of temporal segmentation as the problem of *universality* in a Weighted Finite Automaton (WFA) [10], whereby we aim to accept a word (sequence of predicted labels) with minimal accumulated value. The WFA is derived from the domain-dependant protocol. Furthermore, the transition weights are related to the presence of mislabeling. Although

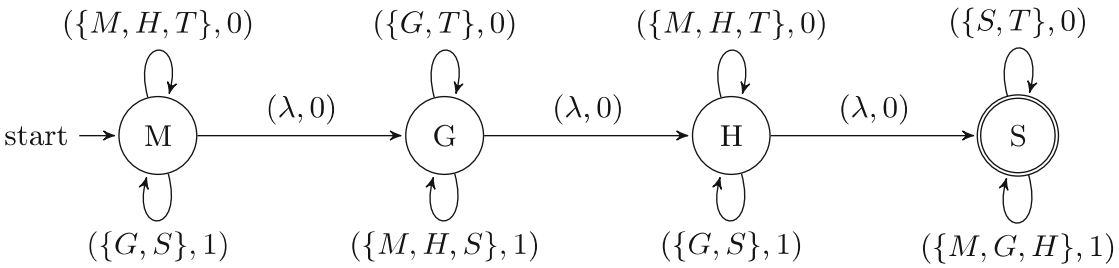

Fig. 3. Weighted Finite Automaton that recognizes the temporal segmentation of colposcopies (Transition - T, Macroscopic view - M, Green - G, Hinselmann - H and Schiller - S).

the problem of universality is PSPACE-complete [10], if any transition in our policy either consumes an input character or moves "forward" to another state in a directed acyclic graph, the recognition problem holds the conditions to formulate a DP implementation. Figure 3 shows a graphical representation of the automaton. We denote each phase by its first letter. The automaton represented in Fig. 3 is formally defined as $A =< \Sigma, Q, \Delta, c, \{M\}, \{S\}, 0, v >$, where

- $\Sigma = \{T, M, G, H, S, \lambda\}$.
- $Q = \{M, G, H, S\}$.
- $\Delta \subseteq Q \times \Sigma \times Q$ is the transition relation defined below, together with the cost function $c : \Delta \longrightarrow \{0, 1\}$. The accepted labels of each state are defined by the top-loop transitions shown in Fig. 3.
  - $(s, q, s) \longrightarrow 0$, if $q \in accepted\_labels(s)$.
  - $(s, q, s) \longrightarrow 1$, if $q \notin accepted\_labels(s)$.
  - $(s, \lambda, s') \longrightarrow 0$, if $s' \neq s$ and s' follows s in the protocol.
  - $v \in \mathbb{N}$, the minimal threshold that accepts the word.

Using the same reasoning we could instantiate any other policy in a straightforward manner. As we said before, given that after each transition the recognition problem is smaller, we can implement this automaton using the DP function defined in the Eq. (2), where *seq* stands for the sequence of labels predicted by the step classifier, next(s) returns the protocol step that follows s and has_next$(s) \equiv (s \neq S)$. It is assumed that the preconditions are evaluated in the same order they are shown. The optimal boundaries can be retrieved from the DP matrix. Since the number of steps in the colposcopic procedure is constant, the performance of the algorithm is linear in the length of the sequence.

$$B[i,s] = \begin{cases} 0, & \text{if } i = length(seq) \\ B[i+1,s], & \text{if } seq[i] \in \{\text{transition, s}\} \\ min(B[i+1,M], B[i,H]), & \text{if } s = M \ \wedge seq[i] = H \\ B[i+1,H], & \text{if } s = H \ \wedge seq[i] = M \\ min(B[i,\text{next}(s)], 1 + B[i+1,s]), & \text{if has\_next(s)} \\ 1 + B[i+1,s]), & \text{otherwise} \end{cases} \quad (2)$$

4 Experiments

In this section we describe the experiments that we performed in this work. We gathered a dataset of 56 colposcopies from different patients that cover a total of 143640 colposcopic images, with every image resized to 64×64 pixels. Sequences were manually annotated by a specialist. The videos and annotations are public on request. Table 1 shows the number of frames per video in each phase. In order to avoid biased results due to differences in the length of the procedures, every patient was equally weighted in the compilation of the results.

Table 1. Statistics of the class distribution per video

Class	Number of frames			Percentage	
	Min	Max	Avg.	Max	Avg.
Transition	0	6488	1071	59.76	39.53
Macroscopic	66	1380	313	100.00	13.88
Green	0	767	187	31.32	7.72
Hinselmann	0	2104	688	52.60	28.45
Schiller	0	2752	304	32.33	10.42
Video	200	11998	2565	–	–

For the assessment of every step of the proposed framework we used a *leave-one-patient-out cross-validation* approach (LOOCV). In this sense, each colposcopy is entirely new for the system at the evaluation stage.

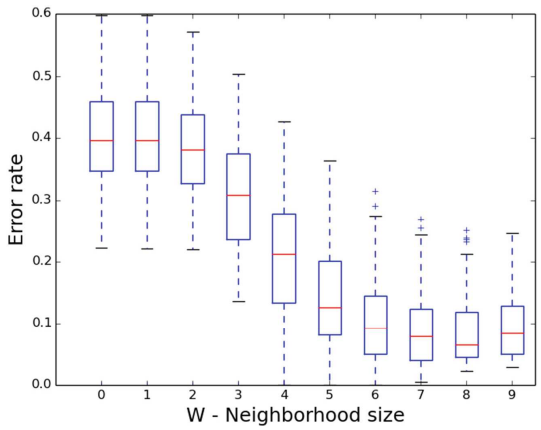

Fig. 4. Error rate of the transition removal method using different neighborhood sizes

For the classification of the transition frames we performed several experiments varying the number of neighbors. This parameter is internally learned using LOOCV. Figure 4 shows the performance of the algorithm at different

Table 2. Transition classifier results

Class	Precision	Recall	F-measure
Non-transition	0.9325	0.9129	0.9206
Transition	0.8610	0.8820	0.8672
Weighted Avg.	0.9146	0.9087	0.9087

values of W. Table 2 shows the classification results for this stage. The average accuracy of the transition recognition is 90.86 %. Furthermore, 93.25 % of the frames that pass to the next stage (colposcopic-step classification) belong to non-transition interval. There is room for human error in the decision of the boundaries of the transition intervals, i.e., it is difficult for a trained human to decide where is the beginning of a transition interval and where it ends. This artifact is also common between the different steps of the colposcopic evaluation. Therefore, the errors shown in these experiments are prone to small human inaccuracies.

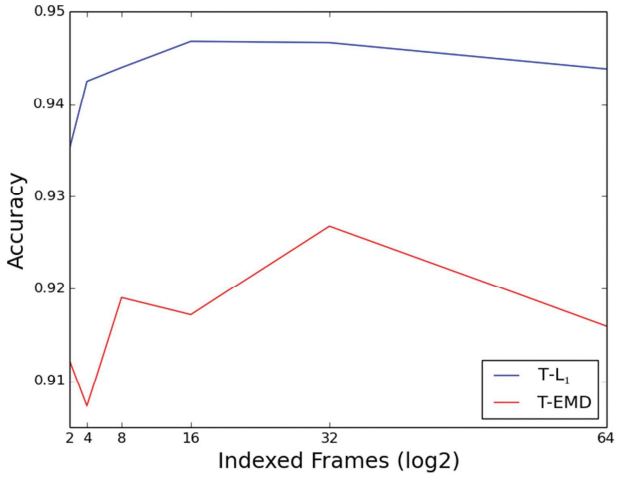

Fig. 5. Colposcopic step accuracy varying the number of indexed frames

For the assessment of the step classification, the number of neighbors in the KNN was set to 5 and the hue and saturation histograms had 180 and 256 bins respectively. We performed experiments varying the number of indexed frames in the KNN database. The results of this experiments can be seen in Fig. 5. The highest accuracy was achieved with 16 indexed frames per phase per video. These results include the temporal segmentation.

Table 3 shows the classification metrics for each colposcopic step using two distance functions: L_1 (equivalent to Histogram Intersection) and *EMD*. We compare each distance before and after temporal segmentation. Contrary to what we thought, *CEMD* did not improve the accuracy but a obtained a significant performance impact. Therefore, we only show the results related to the first two

Table 3. Average classification metrics per class: Macroscopic, Green, Hinselmann and Schiller. Results with 16 indexed frames per video. The results denoted by T-d, where d is the similarity distance, include the temporal segmentation step.

Phase	Distance	Transition				Non-transition			
		Acc.	Prec.	Rec.	F	Acc.	Prec.	Rec.	F
Macro	L_1	0.82	0.36	0.28	0.52	0.70	0.38	0.31	0.52
	T-L_1	**0.96**	**0.99**	**0.78**	**0.84**	**0.98**	**1.00**	**0.95**	**0.95**
	EMD	0.80	0.32	0.28	0.48	0.65	0.33	0.31	0.49
	T-EMD	0.95	**0.99**	0.74	0.80	0.96	**1.00**	0.89	0.89
Green	L_1	**0.97**	0.97	**0.67**	**0.75**	**1.00**	**1.00**	**0.98**	**0.98**
	T-L_1	**0.97**	**0.98**	0.66	0.74	0.99	**1.00**	0.96	0.96
	EMD	**0.97**	0.96	**0.67**	**0.75**	**1.00**	**1.00**	**0.98**	**0.98**
	T-EMD	**0.97**	0.97	0.63	0.70	0.99	0.99	0.91	0.90
Hins	L_1	0.80	0.75	0.55	0.56	0.67	0.76	0.62	0.60
	T-L_1	**0.92**	**0.96**	**0.79**	**0.81**	**0.92**	**0.98**	**0.89**	**0.88**
	EMD	0.80	0.76	0.47	0.54	0.65	0.76	0.53	0.58
	T-EMD	0.91	0.93	0.76	0.77	0.89	0.95	0.86	0.83
Sch	L_1	0.90	0.74	0.60	0.60	0.88	0.79	**0.93**	0.79
	T-L_1	**0.91**	**0.83**	**0.61**	**0.65**	**0.89**	**0.89**	**0.93**	**0.82**
	EMD	0.88	0.67	0.54	0.52	0.82	0.70	0.82	0.62
	T-EMD	0.89	0.77	0.55	0.55	0.84	0.84	0.83	0.71
Avg.	L_1	0.87	0.70	0.52	0.61	0.81	0.73	0.71	0.72
	T-L_1	**0.94**	**0.94**	**0.71**	**0.76**	**0.95**	**0.97**	**0.93**	**0.90**
	EMD	0.86	0.68	0.49	0.57	0.78	0.70	0.66	0.67
	T-EMD	0.93	0.91	0.67	0.70	0.92	0.94	0.87	0.83

distances. As can be seen in the results, the selection of the decision boundaries using the proposed DP algorithm improves the detection of almost every stage. On average, the temporal segmentation algorithm improved the accuracy in 14 % and 28 % in the Macroscopic view phase. In general, the L_1 distance achieved better performance than the *EMD*. Figure 6 shows an example of the step detection results before and after the temporal decision.

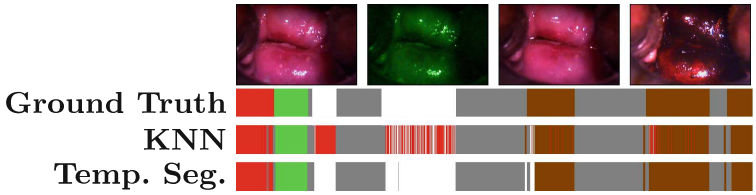

Fig. 6. Timeline with the steps represented by colors: Transition (gray), Macroscopic View (red), Green (green), Hinselmann (white) and Schiller (brown) (Color figure online).

5 Conclusions

In this work we provided a framework to temporarily segment a colposcopic assessment according to its different steps. To assess the quality of the proposed framework we gathered and annotated an open dataset of 56 colposcopies. The proposed framework achieved a precision of 91.46 % in the transition detection using an efficient threshold on motion estimation. Using chromacity information (hue and saturation histograms), we achieved a precision of 96.65 % in the step classification. As we observed in the experiments, for this problem the L_1 distance behaved better than the *EMD*, because histograms from different stages are near, and noisy pixels have a high weight in the resulting *EMD*. Contextual information provided valuable information to smooth and improve the classification results obtained by the per-frame KNN classifier.

Acknowledgement. This work is financed by the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

References

1. Guía global para la prevención y control del cáncer cervicouterino. Technical report, International Federation of Gynecology and Obstetrics, October 2009
2. Das, A., Kar, A., Bhattacharyya, D.: Elimination of specular reflection and identification of ROI: the first step in automated detection of cervical cancer using digital colposcopy. In: 2011 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 237–241. IEEE (2011)
3. Rouhbakhsh, F., Farokhi, F., Kangarloo, K.: Effective feature selection for precancerous cervix lesions using artificial neural networks (2012)
4. Gordon, S., Zimmerman, G., Long, R., Antani, S., Jeronimo, J., Greenspan, H.: Content analysis of uterine cervix images: initial steps towards content based indexing and retrieval of cervigrams. In: Medical Imaging, International Society for Optics and Photonics, 61444U–61444U (2006)
5. Alush, A., Greenspan, H., Goldberger, J.: Lesion detection and segmentation in uterine cervix images using an arc-level MRF. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp. 474–477. IEEE (2009)
6. Park, S.Y., Follen, M., Milbourne, A., Malpica, A., MacKinnon, N., Markey, M.K., Richards-Kortum, R., MacAulay, C., Rhodes, H.: Automated image analysis of digital colposcopy for the detection of cervical neoplasia. J. Biomed. Opt. **13**(1), 014029–014029 (2008)
7. Acosta-Mesa, H.G., Zitova, B., Rios-Figueroa, H., Cruz-Ramirez, N., Marin-Hernandez, A., Hernandez-Jimenez, R., Cocotle-Ronzon, B.E., Hernandez-Galicia, E.: Cervical cancer detection using colposcopic images: a temporal approach. In: 2005 Sixth Mexican International Conference on Computer Science, ENC 2005, pp. 158–164. IEEE (2005)
8. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**(2), 99–121 (2000)

9. Rabin, J., Delon, J., Gousseau, Y.: Circular earth mover's distance for the comparison of local features. In: 2008 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4. IEEE (2008)
10. Almagor, S., Boker, U., Kupferman, O.: What's decidable about weighted automata? In: Bultan, T., Hsiung, P.-A. (eds.) ATVA 2011. LNCS, vol. 6996, pp. 482–491. Springer, Heidelberg (2011)