Medical Engineering and Physics 37 (2015) 309–314

Contents lists available at ScienceDirect

Medical Engineering and Physics

journal homepage: www.elsevier.com/locate/medengphy

Technical note

Colourimetric image analysis as a diagnostic tool in female genital schistosomiasis

Sigve Dhondup Holmen[a,b,*], Eyrun Floerecke Kjetland[a], Myra Taylor[c], Elisabeth Kleppa[a,b], Kristine Lillebø[a], Svein Gunnar Gundersen[a,d,e], Mathias Onsrud[b], Fritz Albregtsen[f,g]

[a] Centre for Imported and Tropical Diseases, Oslo University Hospital, Oslo, Norway
[b] Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway
[c] School of Public Health, Nelson Mandela School of Medicine, University of KwaZulu-Natal, South Africa
[d] Research Department, Sørlandet Hospital HF, Kristiansand, Norway
[e] Institute of Development Studies, University of Agder, Kristiansand, Norway
[f] Department of Informatics, University of Oslo, Oslo, Norway
[g] Institute for Cancer Genetics and Informatics, Oslo University Hospital, Oslo, Norway

ARTICLE INFO

Article history:
Received 27 May 2014
Revised 21 November 2014
Accepted 21 December 2014

Keywords:
Female genital schistosomiasis
Diagnosis
Colposcopy
Cervix
Lesion
Colour

ABSTRACT

Female genital schistosomiasis (FGS) is a highly prevalent waterborne disease in some of the poorest areas of sub-Saharan Africa. Reliable and affordable diagnostics are unavailable. We explored colourimetric image analysis to identify the characteristic, yellow lesions caused by FGS. We found that the method may yield a sensitivity of 83% and a specificity of 73% in colposcopic images. The accuracy was also explored in images of simulated inferior quality, to assess the possibility of implementing such a method in simple, electronic devices. This represents the first step towards developing a safe and affordable aid in clinical diagnosis, allowing for a point-of-care approach.

Abbreviations

FGS	female genital schistosomiasis
HSV	Hue, saturation, value
RGB	red, green, blue
Lab	lightness, a and b
SNR	signal to noise ratio

1. Introduction

Female genital schistosomiasis (FGS) is a poverty-related disease caused by the water-transmitted parasite *Schistosoma haematobium*. The disease may create lesions and inflammation in the genital tract [1]. It is highly prevalent in certain rural areas of Africa, which have limited access to health care and diagnostics. The disease may cause vaginal bleeding, discomfort, dyspareunia, infertility and it might increase the risk of acquiring HIV [1–3].

The lesions associated with FGS consist of changes in the genital mucosa, the most characteristic being the sandy patches (Supplementary Figure 1). These may be classified in two subtypes: First, the sandy patches appearing as grains, which are pathognomonic of FGS [4]. They are characterised by yellow grains that appear clustered or singularly, both superficial and deep in the mucosa. Secondly, the sandy patches appearing as homogenous yellow areas, which are strongly associated with FGS [4].

Detection of ova in direct, bedside microscopy of crushed biopsies from lesions has been considered the gold standard in diagnosing FGS [1,5]. In an area endemic of HIV, biopsy may be considered a safe approach if sexual abstinence or condom protection can be assured one day prior to, and fourteen days following the procedure [6]. However, it is not possible to ensure that all patients follow these precautions, making biopsies an inappropriate diagnostic approach. Furthermore, pathology services are scarce and the transport of specimens (and patients for return visits) is often impossible. It is necessary to develop alternative methods for non-invasive, objective diagnosis of FGS that can be performed at the point of care [1].

Computerised image analysis has been applied in the classification of precancerous cervical lesions using both colour [7] and

* Corresponding author at: Institute of Clinical Medicine, University of Oslo, Postbox 1171 Blindern, 0316 Oslo, Norway. Tel.: +47 900 67 924.
E-mail address: sigve.holmen@medisin.uio.no (S.D. Holmen).

http://dx.doi.org/10.1016/j.medengphy.2014.12.007

Table 1
Exploring colour channels for differentiating the sandy patches from the adjacent mucosa.

Channel	$\bar{c}_{sandy\ patch}$	$\bar{c}_{mucosa}$	$\Delta_{\bar{c}}$	$\sigma\,\Delta_{\bar{c}}$	*t*-value	*p*-value	Reason for exclusion
Red	182.0	148.0	34.0	17.4	10.5	<0.001	
Green	111.4	82.0	29.3	16.0	9.9	<0.001	
Blue	107.3	89.7	17.7	17.0	5.6	<0.001	
Hue	114.6	177.3	7.4	5.6	7.2	<0.001	
Saturation	110.9	118.7	7.8	14.9	2.8	0.009	
Value	182.0	148.0	34.0	17.4	10.5	<0.001	
L	54.3	41.6	12.7	6.6	10.3	<0.001	Similar to *Value*
a	27.6	31.2	3.6	6.9	2.8	0.010	Significance level < 99.9%
b	14.8	7.5	7.30	7.9	5.0	<0.001	

$\bar{c}_{sandy\ patch}$: mean colour value of sandy patch;
$\bar{c}_{mucosa}$: mean colour value of adjacent mucosa;
$\Delta_{\bar{c}}$: difference of means between colours of sandy patches and adjacent mucosa;
$\sigma\,\Delta_{\bar{c}}$: standard deviation of $\Delta_{\bar{c}}$.
The *t*- and *p*-values were calculated using the Student's *t*-test.

texture analysis [8]. Automation of such algorithms has also been described [9]. This paper explores the possibility of exploiting the characteristic colour qualities of the sandy patches as a first step towards developing a computerised diagnostic tool for FGS. Camera equipped cell phones are widely available in developing countries [10], and this computerised image analysis could be implemented as a novel tool using that technology in the clinical diagnosis of FGS.

2. Material and methods

2.1. Image material

The images used to develop this method were acquired in a cross-sectional field study of young women (age 16–22 years) in Southern KwaZulu-Natal, South Africa. The area is endemic for urogenital schistosomiasis. The images were captured using an Olympus OCS-500 colposcope with a mounted Olympus E-420 10 megapixel (Mpx) single lens reflex device (SLR) or a Leisegang colposcope with a Canon EOS 40D 10 Mpx SLR. The image files were stored using high-quality JPEG compression along with data from the clinical investigation regarding lesion size, type, demarcation and location.

Thirty images with clearly visible sandy patches (as diagnosed by the investigating clinician at the time of the examination) were selected for training the computer algorithm. For the experimental validation, 60 images were randomly selected from other patients from the same geographical area: Half of the images from positive cases and half from negative cases (cases and controls). The controls were selected from those patients where sandy patches had not been found by colposcopy, ova had not been detected by microscopy of urine, schistosomal DNA had not been detected in neither cervico-vaginal lavage nor urine (by real time PCR), and no other pathology had been indicated. All the images included in this study fulfilled the following criteria: no foreign elements occluding the field of view (spatula, swab etc.), the lesion was visible within the field of view, there was no other dominant pathology.

2.2. Statistics

The difference in the mean colour values in each colour channel was evaluated using the Student's *t*-test.

The algorithm's read-out in pixels was recorded and used to plot a receiver operating characteristics (ROC) curve. The area under the curve (AUC) was calculated using SPSS Statistics v19 (SPSS Inc., NY, USA) and used as an indicator of the algorithm's predictive quality. The optimal cut-off level for defining a positive case was found by identifying the point on the ROC closest to the upper, left corner.

The significance relating to which imaging device and colposcope were used for capturing the images on the algorithm's output was evaluated using logistic regression using the state of the image (positive or negative) as the dependent variable.

2.3. Colour channels

There are a number of different colour-spaces and it may be difficult to predict which are most useful for a specific analysis [11]. In addition to the *RGB* colour-space, which was used when recording the images, we transformed the colours to the colour spaces called *HSV* and *CIE-Lab* using the open source image analysis software ImageJ (U.S. National Institutes of Health).

In order to train the algorithm in identifying the lesions based on colourimetric properties, the mean colour values were measured for each colour channel of the three selected colour-spaces in superficial grainy sandy patches and the adjacent, normal mucosa (see Table 1).

2.4. Identifying the region of interest

Colposcopic images may contain non-mucosal elements such as parts of the speculum, medical instruments and skin. This may complicate the image analysis. Automated detection of the region of interest (ROI) was performed, drawing from a previously described method [9]. The method exploits the fact that the cervix is characterised by colours in the tones of magenta and that it usually occupies the centre of the image. The colour channel *a* of the *Lab* colour space, represents the increasing intensity of magenta for increasing positive values. The ectocervical mucosa can therefore be characterised by high values of *a* in combination with a short distance from the centre of the image.

By creating a Cartesian coordinate system where the origin corresponds to the centre of the image (x = width/2 and y = height/2), a pixel's distance from the centre could be calculated using the Pythagorean Theorem. A two-dimensional representation of the image was then created using a pixel's distance from the centre of the image as the first dimension and the *a*-channel as the second dimension. The algorithm then identifies the cluster of pixels with the highest values of *a* combined with having the most central position (Fig. 1). The *a*-channel was first smoothed using a Gaussian blur filter with a radius of 35px to eliminate minor details. Clustering was done using the *k*-means method initialised with two points (seeds); closest to the upper left and lower right corners [9,12]. A binary mask was made from the largest group of connected pixels (4-neighbourhood) within the optimal cluster. Empirically, we set a lower threshold of the *a*-channel values corresponding to 65% of the range in order to not oversize the mask. This mask was represented on an empty canvas with the same dimensions as the original image and a 1:1 relationship of pixels. It served as a coarse indication of the ectocervix that had to be expanded in order to reach the ectocervical boundaries (Fig. 1).