Illumination correction for content analysis in uterine cervix images

Hila Dvir, Shiri Gordon and Hayit Greenspan
Tel-Aviv University
Tel-Aviv 69978, Israel
hayit@eng.tau.ac.il

Abstract

Illumination field inhomogeneity strongly affects the visual appearance of an image. It has a major influence on automatic information extraction within an image and its correction is critical for comparison or model learning across images. In this work a unique medical repository of cervicographic images ("cervigrams") collected by the National Center Institute (NCI), National Institute of Health (NIH) is being addressed. The large diversity of cervix shapes within this database, as well as the acquisition set-up, lead to varying illumination conditions among and within the cervigrams, which hamper their automatic analysis. Illumination correction is therefore one of the first preprocessing steps required prior to the image analysis task.

This paper presents a method for illumination correction in cervigrams based on a generalized expectation maximization (GEM) algorithm that interleaves pixels classification with estimation of class distribution and illumination field parameters. For cross-image analysis a normalization of the image dynamic range is conducted, using prior knowledge on cervix tissue intensity distribution.

Experimental results are provided and evaluated on a set of 110 cervigrams that were manually labeled by an NCI expert. Unsupervised segmentation as well as initial supervised tissue classification results are presented.

1. Introduction

Cervical cancer is the second most common cancer affecting women worldwide and the most common in developing countries. It can be cured in almost all patients, if detected by high quality repeated Pap screening, and treated. However, cervical cancer incidence and mortality remain high in resource-poor regions, where high-quality Pap screening programs often cannot be maintained because of inherent complexity and cost. An alternative method of cervical cancer screening, termed cervicography, uses visual testing based on color change of cervix tissues when exposed to acetic acid.

The National Cancer Institute (NCI) has collected a vast amount of biomedical information related to the occurrence and evolution of uterine cervical cancer in a longitudinal multi-year study carried out in Guanacaste, Costa Rica. The data collected includes among other medical detail, 60,000 cervicographic images ("cervigrams") in the form of 35mm color slides, as well as medical classifications for the cervigrams into diagnostic categories. NCI together with the National Library of Medicine (NLM) is developing a unique Web-based database of the digitized cervix images to study the evolution of lesions related to cervical cancer. This work is part of an on-going effort towards the generation of content-based image retrieval (CBIR) capabilities for this cervigrams database.
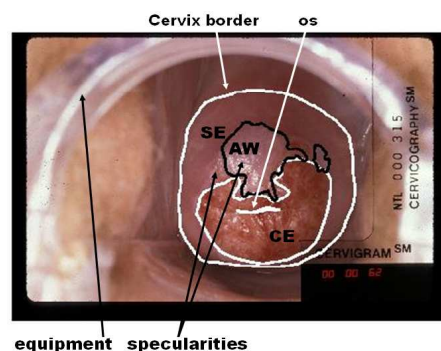

Figure 1. A typical cervigram with the tissues of interest marked.

Three main tissues may appear in cervicography images: The original squamous epithelium (SE), which is a smooth, pink epithelium; The columnar epithelium (CE) that appears red and irregular; and the acetowhite (AW) region which is a transient, white-appearing epithelium following the application of acetic acid. The area of acetowhite is of clinical significance. The three main tissues contain complex and confusing information. The narrow dynamic range of their colors and the lack of clear boundaries between tis-

sue regions hamper their automatic analysis (for an example cervigram see Figure 1)[1]. In addition, the large diversity of the cervigrams content within the database generates image-understanding challenges that require adequate computational analysis tools.

Initial studies can be found on the analysis of individual cervigram images, or the higher-resolution colposcopic images [1, 8, 4, 9, 5]. Preliminary segmentation efforts were recently introduced with respect to the NCI/NLM cervigrams database [3, 10]. In these efforts, it was concluded that the illumination field inhomogeneity is one of the major causes for difficulties in the analysis, thus it's correction ia a desirable preprocessing step.

Illumination inhomogeneity is generated during the cervigram acquisition process. In this process the neck of the uterus is photographed with a special 35mm camera with a ring flash that is used to provide enhanced illumination of the target region. The distance of the camera from the subject was kept constant for all the pictures within the NCI/NLM database. Due to the strong flash of the camera and convex shape of the cervix the image tends to be brighter around the cervix center and the illumination decreases gradually towards the cervix border. This non-uniform illumination confuses automatic segmentation algorithms since bright regions might be misclassified as AW lesions, while AW lesions, located in the shaded regions, won't be detected. Identifying the illumination component and removing it from the original cervigram is expected to reduce the non-uniform illumination effects and help to improve the segmentation results. Additional artifacts that interfere with the tissues segmentation are the specular reflections (SR) artifacts. These artifacts are small and bright regions on the cervix surface, which are generated during the image acquisition process due to the presence of fluids (Figure 1).

The large diversity of cervix shapes within the database, the uneven surface of the cervix and the alignment between the cervix center and the position of the camera lead to varying illumination conditions across the cervigrams in the database. Using a single model for the illumination field is thus not feasible; the illumination field should be learned for each cervigram separately. In this work a method adapted from the research field of bias correction in magnetic resonance (MR) brain images, is proposed. This method is based on the generalized EM algorithm (GEM) introduced by Van Leemput et al. [7]. In MR imaging the brain image is corrupted by a bias field which is an artifact of the MR acquisition protocol. The two mechanisms, the MRI bias and the illumination bias have a similar affect on the image content and as such can be solved in a similar way.

An additional task handled in the current work is a normalization of the intensity histograms across the images in the database in order to provide similar illumination conditions for all the images. This is important especially when trying to compare between the images in the database based on their intensity features.

The rest of the paper is organized as follows: A brief description of an existing cervigram analysis system and the main steps required prior to the illumination correction are presented in Section 2. The illumination correction algorithm is presented and explained in Section 3; Intensity normalization is presented in Section 4; Comparing segmentation results before and after the illumination correction, is exemplified in Section 5; The affect of the illumination correction on a tissue classification test using manual segmentations generated by a medical expert is also presented. A discussion concludes the paper in Section 6.

2. System Description

An initial automated cervigram segmentation system, based on a multi-step process, has been proposed in [3]. This system is extended here to include the steps of illumination correction and intensity normalization, as described in Figure 2.
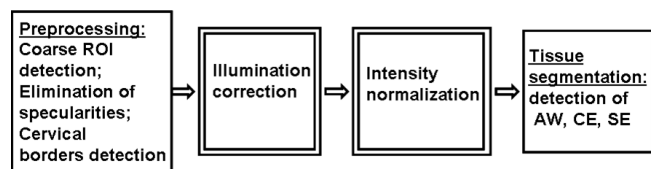

Figure 2. Block diagram of the multi-step analysis process

The first preprocessing step locates the approximate cervical region within the cervigram. This region of interest (ROI) excludes as much irrelevant information as possible, while making sure that the entire cervical area is included. The ROI is identified as a relatively pink and round region located around the image center [3]. Subsequent steps of the process are performed within this ROI.

The second preprocessing step resolves the problem of specular reflections which are generated during the image acquisition. A novel algorithm that detects and eliminates the specularities in cervigrams while preserving the original image features, was devised for this purpose [13].

Next the cervix boundaries are extracted more accurately using a new active contour functional that is based on a local convexity feature [12]. In this process the standard edge indicators, based on image gradients, are replaced by edge indicators that are based on the cervix convexity. This is motivated by the fact that most of the cervix boundaries are outlined by folds of skin that look as narrow valleys and are distinctively concave. At the end of this stage the image is cropped to include the detected cervix region.

The next two steps are the focus of the current work.

Illumination field is removed (Section 3), and the intensity histograms are normalized across the images in the database (Section 4). Tissue segmentation steps are performed next. Segmentation based on unsupervised clustering of the three tissues with no initial illumination correction were shown in [3]. The affect of correcting illumination prior to the segmentation step is shown in the current work.

3. Illumination Correction

Cervigram tissues in this work are mainly described via Lab color space. The illumination correction process is applied to the L color channel, which represents the intensity levels of each pixel and is the only channel affected from illumination changes [2]. Channels a and b remain unchanged. The process is performed on the cropped image, within the detected cervix boundaries. Specularities are initially detected as described in Section 2.

Digital images formation can be modeled as a product of a reflectance component and an illumination component [2]. A logarithmic transformation is therefore initially performed to make the illumination field additive. Next the image is down-sampled. This step slightly improves the results while considerably reducing the running time.

The illumination field is gradually changing within the image plane, thus it can be modeled by a polynomial:

$$I = \sum_{k=1}^{K} c_k \phi_k(x), \quad (1)$$

where c_k are the illumination field parameters, ϕ_k are 2D Lagrange's interpolation basis functions [11] and K is the number of the interpolation points. The interpolation points are located on an equally spaced grid of the image size with $N \times N = K$ points at positions (x_v, y_μ). The Lagrange's 2D interpolation-formula is:

$$Q(x) = (x - x_1)(x - x_2) \dots (x - x_N), Q_v(x) = \frac{Q(x)}{x - x_v}$$

$$\bar{Q}(y) = (y - y_1)(y - y_2) \dots (y - y_N), \bar{Q}_\mu(y) = \frac{\bar{Q}(y)}{y - y_\mu}$$

$$I = \sum_{v=1}^{N} \sum_{\mu=1}^{N} \frac{Q_v(x)}{Q_v(x_v)} \frac{\bar{Q}_\mu(y)}{\bar{Q}_\mu(y_\mu)} c(x_v, y_\mu), \quad (2)$$

where (x, y) are the spatial coordinates of the image pixels, $c(x_v, y_\mu)$ is the illumination field intensity in the position of interpolation point k and $\phi_k = \frac{Q_v(x)}{Q_v(x_v)} \frac{\bar{Q}_\mu(y)}{\bar{Q}_\mu(y_\mu)}$ is the associated interpolation basis function.

Using this model the illumination field can be removed from each pixel in the image using: $\hat{y}_i = y_i - \sum_k c_k \phi_k(x_i)$, where y_i is the pixel's intensity and x_i, is a vector of the

pixel's location. Equations 1 and 2 can be written in a matrix form, using A as the interpolation matrix:

$$I = A \cdot \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}, A = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_K(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \dots & \phi_K(x_2) \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}.$$

Once the illumination field influence is removed from the intensity values, the cervix tissue can be modeled by a mixture of Gaussians. Each tissue type is modeled by a single Gaussian , $G_{\sigma_j}(y - \mu_j)$, with mean μ_j and variance σ_j^2. The probability that the pixel value, $\hat{y}_i$, was generated by tissue type j is:

$$p(\hat{y}_i / T_i = j, \theta_j, c) = G_{\sigma_j}(y_i - \sum_k c_k \phi_k(x_i) - \mu_j) \quad (3)$$

and the overall probability density for $\hat{y}_i$ is:

$$p(\hat{y}_i / \theta, c) = \sum_j p(\hat{y}_i / T_i = j, \theta_j, c) p(T_i = j), \quad (4)$$

with T_i the tissue type at position i, $\theta_j = \{\mu_j, \sigma_j\}$ the distribution parameters for tissue type j, and c, the polynomial parameters. Given a set of n feature vectors: $\hat{y}_1, \dots, \hat{y}_n$, the maximum likelihood estimates of these parameters can be found by maximizing $p(\hat{y}|\theta, c) = \prod_i p(\hat{y}_i|\theta, c)$, using the GEM iterative algorithm [7].

Having the current estimation of the parameter set, each iteration of the GEM algorithm re-estimates the parameter set according to the following steps :

- Expectation step:

$$p(T_i = j / \hat{y}_i, \theta, c) = \frac{p(\hat{y}_i / T_i = j, \theta_j, c) p(T_i = j)}{\sum_j p(\hat{y}_i / T_i = j, \theta_j, c) p(T_i = j)}$$
(5)

- Maximization step of the GMM parameters:

$$p(T_i = j) = \frac{\sum_i p(T_i = j / \hat{y}_i, \theta, c)}{n},$$

$$\mu_j = \frac{\sum_i p(T_i = j / \hat{y}_i, \theta, c)(y_i - \sum_k c_k \phi_k(x_i))}{\sum_i p(T_i = j / \hat{y}_i, \theta, c)},$$

$$\sigma_j^2 = \frac{\sum_i p(T_i = j / \hat{y}_i, \theta, c)(y_i - \sum_k c_k \phi_k(x_i) - \mu_j)^2}{\sum_i p(T_i = j / \hat{y}_i, \theta, c)}.$$
(6)

- Maximization step of the polynomial coefficients:

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = (A^T W A)^{-1} A^T W R \quad (7)$$

$$W = diag(w_i); \qquad (8)$$

$$w_i = \sum_j w_{ij};$$

$$w_{ij} = \frac{p(T_i = j/\hat{y}_i, \theta, c)}{\sigma_j^2}$$

$$R = \begin{bmatrix} y_1 - \tilde{y}_1 \\ y_2 - \tilde{y}_2 \\ \vdots \end{bmatrix} \qquad (9)$$

$$\tilde{y}_i = \frac{\sum_j w_{ij}\mu_j}{\sum_j w_{ij}}, \qquad (10)$$

where A is the interpolation matrix representing the geometry of the illumination field. W is the diagonal matrix holding the sum over all weights, w_{ij}, assigned to each pixel. $\tilde{y}_i$ is the pixel's predicted intensity value, without the illumination field and R is a vector that represents a rough estimation of the illumination-field at every pixel.

When using the GEM for bias correction in MRI brain images [7], an a-priori anatomical atlas of the brain is used in order to initialize the GMM parameters. As no such atlas is available for the cervigram images, a standard K-means clustering algorithm is used to estimate the initial values of μ_j and σ_j. The c_k values are initially zeroed.

In equation 7 the polynomial parameters, c_k, are calculated using weighted least-squares, primarily from pixels that belong to tissue types with strong weights. The weights w_{ij} are chosen so they are inversely proportional to the tissue's intensity-variance (Equations 8), thus tissues with a narrow intensity distribution attain higher weights. Note that small intensity variance per tissue indicates reduced illumination artifacts, thus pixels that belong to these tissues with a high probability (Equation 5) attain more reliable $\tilde{y}_i$ and R_i values.

Applying the GEM algorithm to cervigrams requires further extensions. It was empirically found that the surface of the illumination field is best described as a fifth degree polynomial ($K = 36$ interpolation points). The number of tissues j was selected to be 4 (roughly representing the AW, SE and CE tissues and an additional "unknown" tissue), but experiments have shown that finding the illumination field is indifferent to this number. Adding channels a and b to the segmentations steps of the GEM algorithm, meaning: 3D segmentation in Equations 3-6 and performing the illumination corrections only on the L channel (Equations 7 - 10), had shown to have destructive affect on the segmentation, further encouraging the notion that this process should be performed on the intensity channel alone.

Specularities and non-cervix tissues (that remain outside the detected cervix boundaries within the cropped image) may interfere with the illumination field estimation.

Eliminating the specularities as part of the preprocessing steps ([13]) might generate artifacts of smooth bright regions that interfere with the correct estimation of the illumination field. The GEM algorithm is therefore modified to support such pixels that may interfere with the process. These pixels, which are identified in the preprocessing steps of SR and cervix boundary detection, are assigned w_{ij} values (Equation 8) of a small positive number, close to zero. This enables to perform the interpolation on the whole image plain while the contribution of these pixels to the illumination field calculation will be negligible. This modification is necessary since the interpolation can't ignore pixels, so the pixels can't be simply masked out.

Following the GEM process, the resulting illumination field is up-sampled to its original size using a bicubic interpolation and is extracted from the log of the original L channel. The corrected L channel is then transformed back to the intensity range using the inverse log transform.

The output of the GEM algorithm consists of the illumination field estimation along with the image segmentation map into 4 tissues of interest. Although the illumination field is well estimated, the segmentation is not accurate enough, as it is performed on the log of a down-sampled image using the intensity channel alone. Important information required for the correct cervigram segmentation is ignored in this scenario. An additional segmentation step is therefore required in order to detect the tissues within the cervix.

4. Intensity normalization

The illumination correction step thus described, supports the segmentation of a single image. Another major challenge that remains is the large diversity in the intensity range across the images within the database. Although this diversity might not interfere with the segmentation process of a single image it has a major affect when trying to compare between different images and when trying to generate global tissue models for tissue classification. Note that the dynamic range of the intensity channel is changed following the illumination correction step, thus further normalization and stretching of the intensity channel are required.

Histogram matching approaches are not suitable for the cervigram images because of their content diversity, for example: some of the cervigrams have no AW or CE tissues at all. Histogram matching to a specific histogram model that possess these tissues is therefore erroneous.

The normalization process suggested uses prior knowledge on the cervigram content. It is done on the basis of the following facts that relate to the SE tissue, (the original squamous epithelium): 1. The SE tissue always appears in the cervix image. 2. The SE tissue has the narrowest histogram of all tissues in the cervix (following illumination correction). 3. The SE tissue includes a large (or the

largest) amount of pixels within the cervix region.

These facts enable to model the intensity histogram of the cervix region as a diverse histogram with a maximum peak close to the average color of its SE region. Assuming that the SE tissue should possess a similar color across images the intensity histogram is shifted so its peak is set to a fixed value (for example: $L = 60$). This is done using the following equation:

$$L(x, y) = L(x, y) - \max(h) + 60,$$

where $\max(h)$ is the intensity value at the histogram's peak. The dynamic range of the histogram is linearly stretched to the original range of the image.

5. Experimental Results

In this section an experimental validation of the proposed method is provided. The algorithm was applied to a set of 110 cervigrams that were manually labeled by an NCI expert into the three tissues of interest. The number of patches extracted from the manual markings for each of the tissues is: 110 SE-patches, 70 AW-patches and 71 CE-patches. The illumination correction is evaluated via unsupervised cervix segmentation (Section 5.1), its affect on the intensity histograms of the corrected images (Section 5.2) and by its influence on a supervised tissue classification test (Section 5.3).

5.1. Illumination Correction Affect on Unsupervised Cervix Segmentation

In this experiment probabilistic unsupervised segmentation based on a mixture of Gaussians model [3] is used. Pixels within the cervix region are initially represented by a 3-dimensional feature vector in the Lab feature space and the feature vectors are clustered into a mixture of 4 Gaussians. Probabilistic segmentation is enabled by assigning each feature vector to the most probable Gaussian cluster in the learned mixture. Specularities and regions outside the detected cervix boundaries are masked out during the segmentation process and ignored, as they provide misleading tissue information.

Figure 4 exemplifies four segmentation results, before and after the illumination correction. The original cropped cervigram is shown at the top of each example (a). Manual markings of the expert are imposed on this image. AW regions are marked by a blue line, CE regions are marked by a purple line and the SE region is defined as the rest of the tissue within the cervix region, which is marked by a yellow line. In (b) the original cropped image is presented with the SR and non-cervix regions (outside the automatically detected boundaries) masked out (marked in black). Segmentation results into 4 tissues are presented in (c) with

		$Dice$	TP	FP
original	mean	0.13	0.53	0.22
	std	0.14	0.31	0.1
corrected	mean	0.10	0.42	0.14
	std	0.1	0.27	0.14

Table 1. Dice, true-positives (TP) and false-positives (FP) for detection of AW regions before and after the illumination correction.

different regions having different colors. The region corresponding to the Gaussian with the brightest intensity is associated with the candidate AW region and is the region who's quality is being examined. This region is marked in dark red. In (d) the estimated illumination field is displayed. The illumination corrected image is presented in (e) and the segmentation map generated for the corrected image is presented in (f). The candidate AW regions are marked in dark red.

Visual inspection of the results reveals the varying illumination field across the cervigrams (d). A correspondence can be detected between the estimated illumination field and the dark red region on the segmentation maps of the original images (c). The dark red regions are considerably reduced in the segmentation maps of the illumination corrected images (f). It is mostly the case that AW regions that were not detected before the illumination correction are now detected and a better correspondence to the medical expert's markings, now exist. In very few cases AW regions are missed (i.e. Example IV).

Quantitative evaluation is a known challenge in tissue characterization. Validation by expert markings suffers from the subjective expert's opinion and from the individual marking style of the experts, resulting in a large variability across observers. In order to numerically evaluate the segmentation results of the AW regions, the Dice metric, $\frac{S \cap R}{S \cup R}$, the true-positives metric (TP), $\frac{S \cap R}{R}$, and the false-positives metric (FP), $\frac{S \cap \hat{R}}{\hat{R}}$, were calculated for each of the detected candidate AW regions (S is the area of the automatically segmented AW region, R is the area of the expert's segmentation and $\hat{R}$ the complement of that area). The results of the different metrics before and after the illumination correction are summarized in Table 1. Checking the Dice and TP metrics it might seem that there is a small degradation in segmentation quality. This is not the case when visually inspecting the images. The segmentation of the marked AW lesions was a little damaged while the amount of false positives (FP) detected was considerably reduced (as can be seen from the table as well). The overall segmentation results are therefore more accurate.

5.2. Normalization of the Intensity Histogram

In this section the intensity histogram normalization process is demonstrated. Intensity histograms were calculated for the collections of patches marked on the 110 tested images for each of the three tissue types. Figure 3 presents these histograms, imposed on a common intensity scale, before (a), and after (b) illumination correction, as well as the following intensity normalization step (c). The dashed line represents the histogram of the SE tissue, the solid line the histogram of the AW tissue and the dashed-dot line the histogram of the CE tissue. Figure 3(b) motivates the intensity normalization step, it can be easily detected that a large overlap still exists between the AW and the SE tissues prior to this normalization. Following the histogram normalization process (Figure 3(c)) the peak of the SE tissue is located at $L = 60$. The variance of the SE and the AW histograms is much smaller and a better separation exists between the two tissue types. A single Gaussian can represent each of the tissues in this case.

5.3. Evaluating the Illumination Correction via a Supervised Classification Test

Using the manual markings of the medical expert, Gaussian models are learned for each of the cervix tissues in a supervised manner (using a training set which consists of patches from 35 cervigrams). The histogram presented in Figure 3(c) motivates this type of model. The rest of the patches (from the remaining 75 images) are defined as a test set. A single Gaussian is learned for each of these patches. Patch models and tissue models were learned in the $L - b$ feature space. Adding the a color channel reduced the overall results, it was therefore discarded.

The classification test was performed using the Kullback-Leibler (KL) divergence [6] as the similarity measure between two Gaussian distributions. Lower KL values indicate a larger similarity between two distributions. Each of the modelled patches within the test set was compared to the three tissue models using the KL divergence and was classified by the nearest model. The test was performed on the same patches before and after the illumination correction. Tables 2 and 3 are the class confusion matrices generated in each of the classification tests. Each row represents the classification distribution of the patches from a single tissue type to the different tissue models (columns). Inspecting these results it can be seen that the SE and the CE detection was considerably improved following the illumination correction. The AW classification was slightly damaged.

6. Conclusions

In this work a method for illumination correction and normalization in cervigrams is introduced. The method is
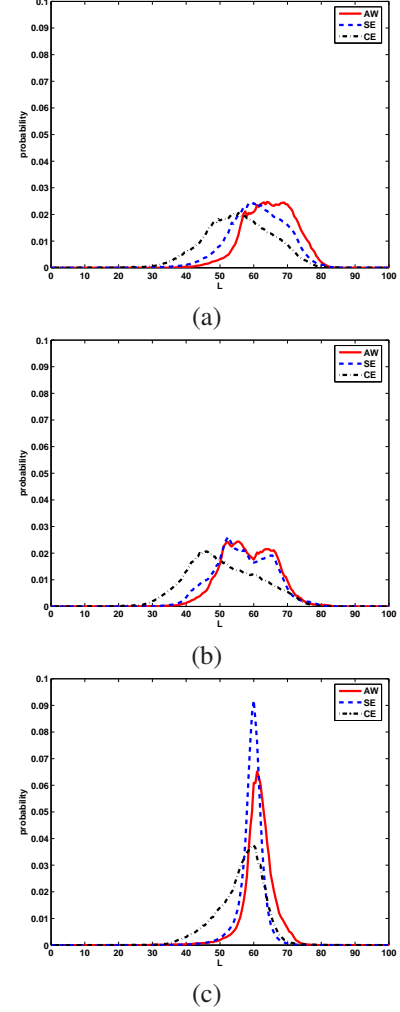

Figure 3. Histograms of the three cervix tissue types before (a) and after (b) the illumination correction, and after the intensity normalization (c).

tissue type / model	AW	SE	CE
AW	38	8	1
SE	30	41	4
CE	2	9	34

Table 2. Classification results before illumination correction. Tissue types - rows; Tissue models - columns

tissue type / model	AW	SE	CE
AW	35	9	3
SE	11	60	4
CE	2	4	39

Table 3. Classification results following illumination correction. Tissue types - rows; Tissue models - columns

evaluated in the context of cervix tissues automatic segmen-

tation (Section 5.1) and supervised classification (Section 5.3). The results of these experiment are consistent: following the illumination correction process, the SE and the CE tissues are more separated from the AW tissue. The SE tissue is not miss-classified as AW and the amount of false positives is considerably reduced. Prior to the illumination correction the SE and the AW tissues possessed very similar intensity properties and large parts of the cervix region were identified as AW. This result is important in the context of automatic cervigrams analysis where a small amount of false positives in the AW detection process is desirable.

The variance of the SE and the AW histograms presented in Section 5.2 is much smaller following the illumination correction and the histogram normalization steps, a better separation between the two tissue types exists. Such a separation is important when trying to model the tissues properties in order to enable their identification.

The overall AW detection quality introduced by the Dice and TP metrics in Section 5.1 is not impressive. This can be explained by the large similarity that exists between the AW and other cervix tissues like mucus and non-cervix regions that remains inside the automatically detected cervix boundaries. Substantial work is still required in order to improve the AW detection quality. This work includes the investigation of additional feature spaces and segmentation techniques. This investigation is part of the current work being conducted within the cervigrams analysis framework.

Acknowledgement

References

[1] P. M. Cristoforoni, D. Gerbaldo, A. Perino, R. Piccoli, F. J. Montz, and G. L. Captianio. Computerized colposcopy: Results of a pilot study and analysis of its clinical relevance. *Obstet. Gynecol*, 85:1011–1016, 1995.

[2] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prantice-Hall, Inc, 2002.

[3] S. Gordon, G. Zimmerman, R. Long, S. Antani, J. Jeronimo, and H. Greenspan. Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams. In *Proc. of SPIE Medical Imaging*, 2006.

[4] Q. Ji, J. Engel, and E. Craine. Texture analysis for classification of cervix lesions. *IEEE Transaction on Medical Imaging*, 19(11):1144–1149, 2000.

[5] P. King, S. Mitra, and B. Nutter. An automated, segmentation-based, rigid registration system for cervigram images utilizing simple clustering and active contour techniques. In *Proc. of 17th IEEE Symposium on Computer-Based Medical Systems*, pages 292–297, 2004.

[6] S. Kullback. *Information Theory and Statistics*. Dover, 1968.

[7] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model- based bias field correction of mr images of the brain. *IEEE Trans. on Medical Imaging*, 18:885–896, October 1999.

[8] B. W. Pogue, M. A. Mycek, and D. Harper. Image analysis for discrimination of cervical neoplasia. *Journal of Biomedical Optics*, 5(1):72–82, 2000.

[9] V. V. Raad. Frequency space analysis of cervical images using short time Fourier Transform. In *Proc. of the IASTED International Conference of Biomedical Engineering*, volume 1, pages 77–81, Salzburg, Austria, January 2003.

[10] Y. Srinivasan, D. Hernes, B. Tulpule, S. Yang, J. Guo, S. Mitra, S. Yagneswaran, B. Nutter, J. Jeronimo, B. Phillips, R. Long, and D. Ferris. A probabilistic approach to segmentation and classification of neoplasia in uterine cervix images using color and geometric features. In *Proc. of the SPIE Medical Imaging 2005*, volume 5747, pages 995–1003.

[11] J. F. Steffensen. *Interpolation*. Chelsea Publishing Company, 1950.

[12] G. Zimmerman, S. Gordon, and H. Greenspan. Automatic landmark detection in uterine cervix images for indexing in a content-retrieval system. In *Proc. of International Symposium on Biomedical Imaging*, 2006.

[13] G. Zimmerman and H. Greenspan. Automatic detection of specular reflections in uterine cervix images. In *Proc. of SPIE Medical Imaging*, 2006.
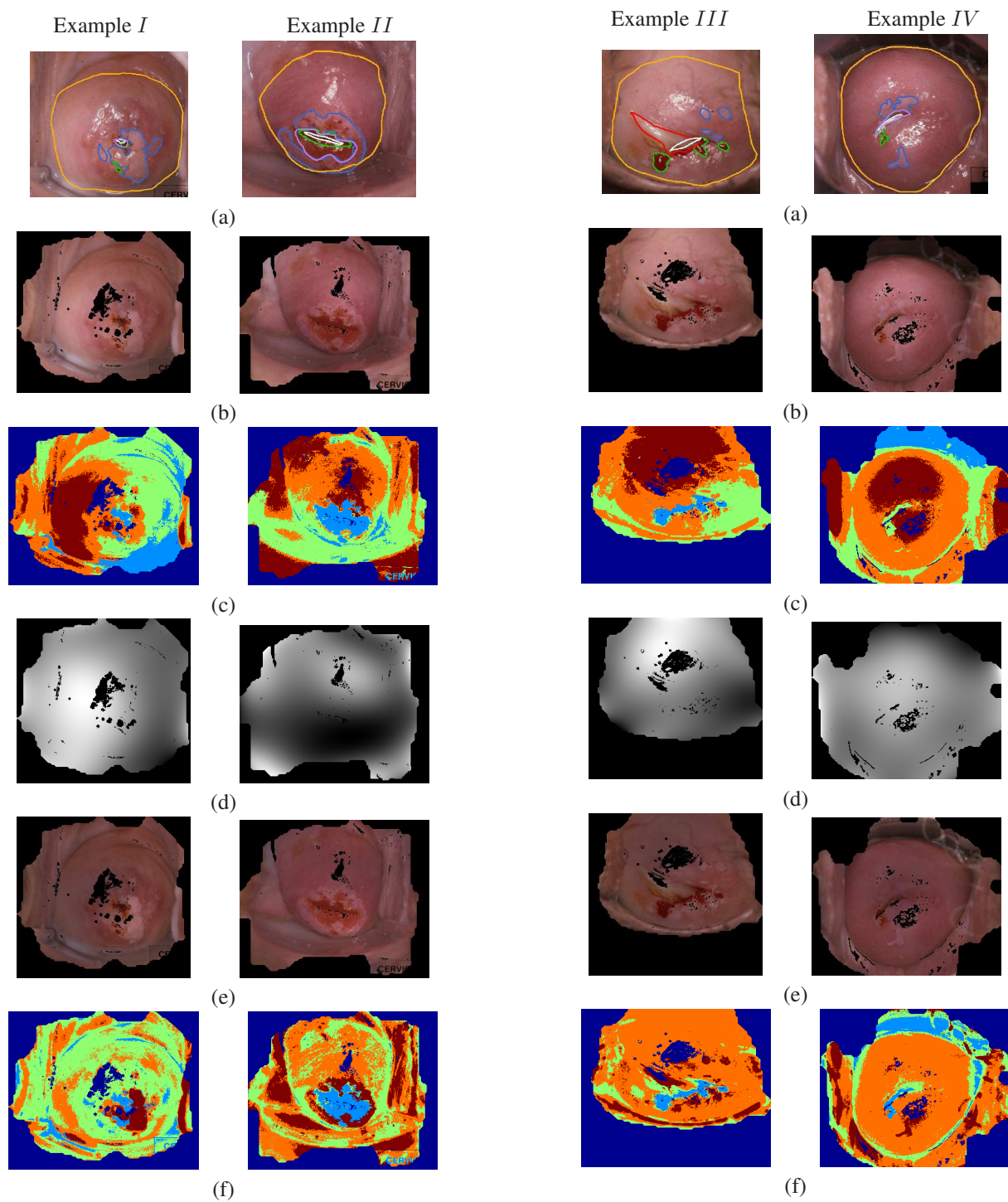
Figure 4. (a) original cervigram, manual markings by an expert imposed; (b) preprocessed cervigram, specularities and non cervix regions are masked out; (c) segmentation map of original image, brightest regions are marked in dark-red; (d) estimated illumination field; (e) illumination corrected cervigram; (f) segmentation map of the corrected cervigram, brightest regions are marked in dark-red.