A Comparative Study of Color Spaces in Skin-Based Face Segmentation

J. Montenegro and W. Gómez
Laboratorio de Tecnologías de Información
CINVESTAV-IPN
Ciudad Victoria, Tamaulipas
Email: jmontenegro@tamps.cinvestav.mx

P. Sánchez-Orellana
División de Estudios de Posgrado e Investigación
Instituto Tecnológico de Ciudad Victoria
Ciudad Victoria, Tamaulipas
Email: mcpedrosanchez@gmail.com

***Abstract*—This paper presents a comparative study of five color spaces commonly used for detecting human skin. The evaluated models were: normalized RGB, HSV, YCbCr, CIE Lab, and CIE Luv. These color spaces attempt to separate the luminance from chrominance components, which is useful to make the face skin detection illumination independent. We used the Microsoft Kinect®sensor for acquiring 705 RGB images from 47 subjects in the age range from 18 to 45 years and distinct skin tones. Besides, each image was segmented manually to define true skin pixels. A probabilistic classifier was built for each tested colorspace to classify a pixel color into skin class or non-skin class. The Matthews correlation coefficient (MCC) was used to evaluate the quality of the computerized skin classification. The results pointed out that the CIE Lab colorspace reached the best MCC performance with median value equal to 0.779 and Q_n estimator equal to 0.074. The worst performance was attached by normalized RGB with with median value equal to 0.606 and Q_n estimator equal to 0.143.**

I. Introduction

Many biometric applications require a preliminary process to detect human skin. These applications could involve tasks such as face recognition or face analysis. The former attempts to identify or verify a person from a digital image or a video frame, whereas the latter commonly analyses facial expressions to recognize distinct emotional states such as sadness, anger, happiness, etc.

In these applications it is convenient to detect first the face region for facilitating further feature description and recognition. In this context, face detection could be viewed as a face segmentation problem whereby the face region is separated from the background [1], [2]. Thus, by detecting image pixels within a range of "skin-like" colors, it is feasible to isolate the face region by some thresholding stage.

However, when designing a skin segmentation approach it is important to considerate the skin tone variability among different subjects, the characteristics of the environment, and the amount of illumination [2], [3]. To cope with these conditions, a controlled environment for acquiring images could be used for increasing the performance of the segmentation technique. Hence, the scene complexity is reduced by controlling variables as the amount of illumination and the background color.

Frequently, skin segmentation is performed in four steps [4], [5]: (i) transforming the input image (commonly in RGB colorspace) to some colorspace with single luminance channel; (ii) eliminating the luminance component of the colorspace and using only the chrominance channels in the classification process; (iii) classifying the image pixels by modeling the probability distribution of skin color; and (iv) thresholding the classified image to obtain exclusively the skin region.

Ideally, when mapping a large variability of skin-like tonalities to some specific chromatic space, the samples should be concentrated in a very narrow-band [6]. Thus, when classifying certain skin pixels, they should have high probability to be skin-like colors. However, in real applications this goal is difficult to attach and it is necessary to determine which colorspace provides an adequate skin segmentation performance.

Concerning our application, we are interested in acquiring automatically 3D face sequences from the Microsoft's Kinect®sensor for face analysis. This device is considered as a natural user interface that involves several advantages such as low-cost (about $150 dollars), portability, and the 3D scanning and motion capture is marker-less. Moreover, there exists diverse open source frameworks, such as Open NI or Open Kinect, to acquire and process easily Kinect's data. This technology projects infrared (IR) patterns on the scene and estimates the depth map by measuring the distortion of IR patterns using an IR sensor and structured light algorithms. Also, a RGB camera with VGA resolution is used to obtain the natural color of the reconstructed objects [7].

For achieving 3D face reconstruction, the depth map information, acquired from the IR sensor, should be employed. Thus, the RGB data could be used to detect skin-like colors for segmenting automatically the face region in the corresponding depth map. However, it is necessary to determine which colorspace is the most adequate for increasing the segmentation performance of our approach. Therefore, in this paper we compare five distinct color spaces commonly used in skin-based segmentation: normalized RGB, HSV, YCbCr, CIE Lab, and CIE Luv. Also, we propose a methodology to evaluate objectively which color-space is the most adequate for skin segmentation applications.

II. Materials and Methods

A. Evaluated color spaces

Let the tuple $[C_0, C_1, C_2]$ the color of a pixel in image. Then, the colorspace transformation converts $[C_0, C_1, C_2]$ to $[C_0', C_1', C_2']$. Herein, images in RGB colorspace were captured

IEEE Catalog Number CFP13827-ART
ISBN 978-1-4799-1461-6

using the Kinect device, and they were converted to the following non-RGB color spaces [8]:

- **RGB** model corresponds to the three primary colors: Red, Green and Blue, respectively. To reduce the dependence on lighting, the RGB color components are normalized so that sum of the resultant components is unity, i.e., r + g + b = 1. We denote to this normalized colorspace as **rgb** (lower case).
- **HSV** defines color as **H**ue, **S**aturation and **V**alue (also called Lightness or Intensity). It was designed to approximate the way that humans perceive and interpret color. The main advantage of HSV model lies in the extremely intuitive manner of specifying color.
- **YCbCr** separates explicitly both the luminance and the chrominance components. The luminance (**Y**) is computed as a weighted sum of RGB values, whereas chrominance (**Cb** and **Cr**) is calculated by subtracting the luminance component from **B** and **R** channels.
- **CIE Lab** and **CIE Luv** color spaces are nearly linear with visual perception. Its components of lightness (**L**) and chromaticity (**ab** and **uv** channels) represent how two colors differ in appearance to a human observer. CIE Luv has an associated two-dimensional chromaticity chart which is useful for showing additive color mixtures. On the other hand, CIE Lab has no associated two-dimensional chromaticity diagram and no correlate of saturation.

B. Recording environment

We built a controlled recording environment for reducing the scene complexity to attempt improving the automatic face segmentation stage.

Basically, the recording environment is an enclosed cabin, with dimensions $1.9\times1.3\times1.7$ meters, covered by a black vinyl fabric and held by a PVC pipe structure as shown in Figure 1. The goal is to avoid uncontrolled lighting from outside within the cabin. Besides, two high intensity LED lamps were used to provide constant illumination on subject's head. Also, an homogeneous black matte background was placed behind the subject to contrast the face skin. Finally, the Kinect device was mounted on a tripod while the subject remained sat for acquiring a set of RGB images.

C. Image dataset

The dataset comprises 705 RGB images (640×480 pixels) of 47 Mexican people (students and workers from Cinvestav-Tamaulipas) in the age range from 18 to 45 years. These images were acquired with the RGB camera of the Kinect device. For each subject, 15 pictures were taken within the recording environment shown in Figure 1.

Besides, regarding the entire image dataset, binary masks were generated manually to identify pixels corresponding to skin. The objective was to obtain certain skin color samples for designing the classifier. Thus, about 12.5×10^6 skin color pixels were considered for our experiments and they were converted to the five color spaces aforementioned.

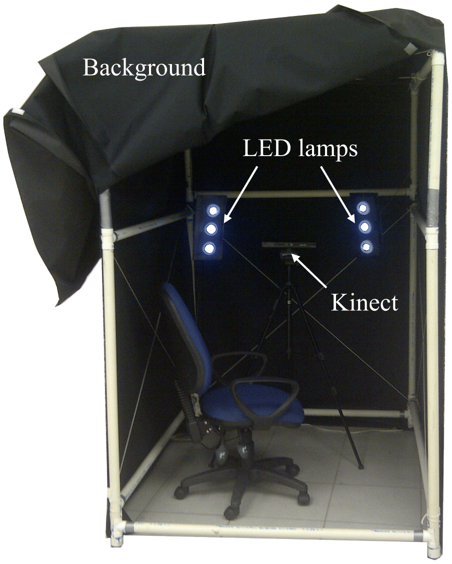

Fig. 1. Recording environment used for image acquisition.

D. Skin color classification

Once a skin color pixel was converted from RGB to some non-RGB colorspace, the luminance channel was dropped to make the skin detection illumination independent. Thus, the following luminance components were removed: r for normalized rgb, V for HSV, Y for YCbCr, L for CIE Lab and CIE Luv.

Thereafter, a Bayesian classifier was built for each colorspace to classify a pixel color into skin class (ω_{s}) or non-skin class (ω_{ns}). Moreover, for each pixel a feature column vector $\mathbf{x}$=$(C'_1, C'_2)^T$ was created by using two chrominance components after colorspace transformation.

The class conditional probability of the Bayesian rule could be expressed by the multi-variate normal density function as [9]:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right) \quad (1)$$

where μ is the d-component mean vector (here $d = 2$), Σ is $d \times d$ covariance matrix, and $|\cdot|$ denotes the determinant.

Once the probability in Equation 1 was computed on every pixel in the image, each one should be labeled as a skin pixel if such probability is larger than a threshold and non-skin otherwise.

To cope with this requirement, the probability image was normalized to the range [0, 255]. Next, the Otsu method [10] was applied to define a gray-level threshold for binarizing the normalized image. This thresholding approach attempts to group the two classes, skin and non-skin, by maximizing the interclass variance. Hence, white pixels correspond to skin, whereas black pixels to non-skin. It is worth mentioning that interclass separation could be improved by guaranteeing bimodality of the image gray-level histogram. Therefore, the output of the probabilistic model in Equation 1 should assign

IEEE Catalog Number CFP13827-ART
ISBN 978-1-4799-1461-6

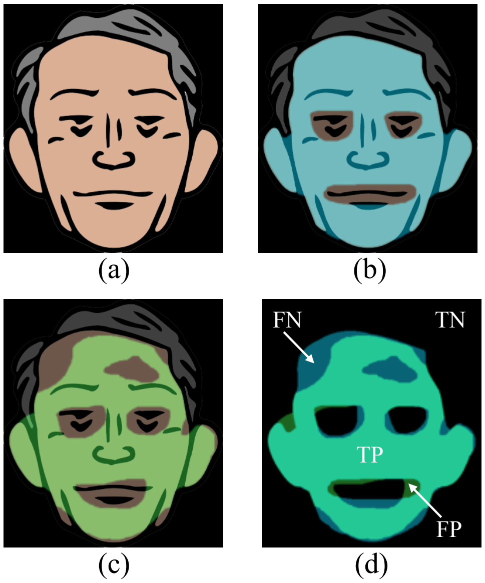

Fig. 2. (a) Original face image; (b) and (c) superimposed masks of manual, S_m, and computerized skin classification, S_c, respectively; (d) differences on binary masks in (b) and (c) in terms of area error metrics.

larger probability values to skin pixels than background pixels for contrasting both regions.

E. Performance evaluation

The objective of the performance evaluation is to define the degree to which the computerized skin classification, S_c, agrees with manual classification (i.e., binary mask), S_m, regarding a specific colorspace. It could be assessed in terms of area error in pixels, which involves four basic metrics, as illustrated in Figure 2 [11]: false-positive (FP) denotes the area falsely identified by S_c when compared with S_m; false-negative (FN) expresses the area in S_m that was missed by S_c; true-positive (TP) indicates the total area of S_m that was actually covered by S_c; and true-negative (TN) denotes the total area in S_m that is truly not in the lesion that was also excluded by method S_c. These metrics could be computed as:

$$\begin{aligned}\mathrm{FP} &= \left|(S_m \cup S_c) - S_m\right|,\\ \mathrm{FN} &= \left|(S_m \cup S_c) - S_c\right|,\\ \mathrm{VP} &= \left|S_m \cap S_c\right|,\\ \mathrm{VN} &= \left|\overline{S_m \cup S_c}\right|,\end{aligned} \tag{2}$$

where $\cup$ and $\cap$ are union and intersection, respectively, $|\bullet|$ is number of pixels with value "1" and $\bar{\bullet}$ denotes the complement.

We used the Matthews correlation coefficient [12] to measure the quality of the computerized skin classification. This metric takes into account the four expressions in Equation 2 to make a balanced measure when the classes are of very different sizes. Generally, in our image dataset the amount of skin and background pixels is uneven. The Matthews correlation coefficient is computed as:

$$\mathrm{C} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}} \tag{3}$$

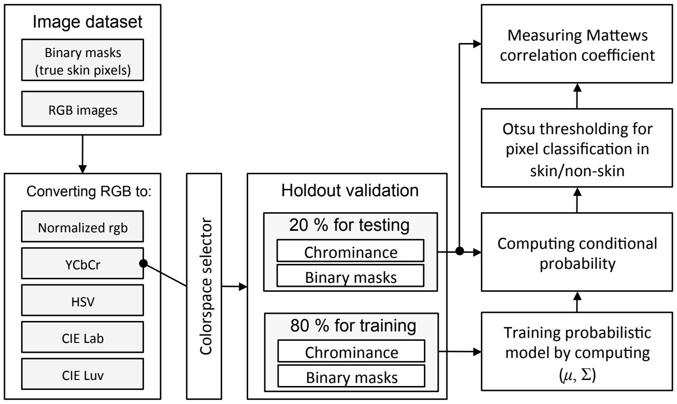
Image dataset
Binary masks (true skin pixels)
RGB images
Converting RGB to:
Normalized rgb
YCbCr
HSV
CIE Lab
CIE Luv
Colorspace selector
Holdout validation
20 % for testing
Chrominance
Binary masks
80 % for training
Chrominance
Binary masks
Measuring Mattews correlation coefficient
Otsu thresholding for pixel classification in skin/non-skin
Computing conditional probability
Training probabilistic model by computing (μ, Σ)

Fig. 3. Block diagram of the experimental setup.

Equation 3 returns a value in the range $[-1, +1]$, where a coefficient of "$+1$" represents a perfect matching between S_m and S_c, and "-1" indicates total disagreement. Hence, adequate computerized skin classification should tend toward "$+1$".

III. Results

A. Experimental setup

The entire image dataset (i.e., 705 RGB images) was converted to the five color spaces in Section II.A. Thereafter, a specific colorspace was selected and the holdout validation was used to built two image subsets chosen randomly: 80 % for training (564 images) and 20 % for testing (141 images).

Next, the parameters of the probabilistic model in Equation 1 were computed, that is, the mean vector (μ) and the covariance matrix (Σ). Recall that only the chrominance components of true skin pixels (marked by binary masks) were used to computed these parameters.

Finally, each image of the testing subset was evaluated in the probabilistic model and the Otsu method classified the pixels as skin or non-skin. Then, the resultant binary image was compared with its respective binary mask (i.e., true skin pixels) by using Equation 3. Figure 3 presents a block diagram of the steps adopted to evaluate which colorspace is the most adequate for segmenting skin face.

All the algorithms were developed in C/C++ language. The Open Natural Interaction (OpenNI) framework was used for accessing the Kinect device, whereas the Open Source Computer Vision (OpenCV) library was employed for colorspace transformations.

B. Statistical analysis

The experiment depicted in the Section III.A was executed 100 times for performing statistical analysis, in terms of Matthews correlation coefficient, and for each execution two robust statistics for estimating location (sample median, MED) and scale (Q_n estimator) were used.

The sample median is the middle value in an ordered sequence of data values, that is, the 50th percentile. On the other hand, the Q_n estimator computes all possible distances between pairs of observations, sorts them, and picks out the

IEEE Catalog Number CFP13827-ART
ISBN 978-1-4799-1461-6

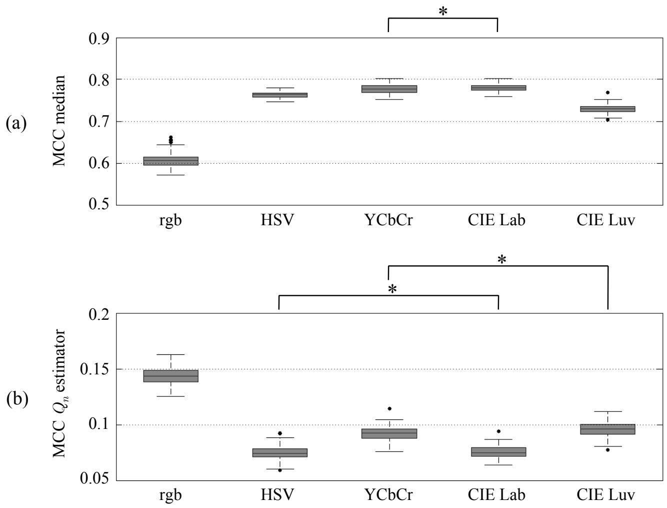

Fig. 4. Matthews correlation coefficient (MCC) results for each evaluated colorspace: (a) median and (b) Q_n estimator. The asterisk indicates no statistical difference ($p > 0.05$) between two groups.

smallest distance, corresponding to the first quartile of all interpoint distances [13].

Furthermore, a statistical analysis was conducted by using the Kruskal-Wallis test ($\alpha = 0.05$) whose null hypothesis was formulated as follows: the mean ranks samples from the groups are expected to be the same. Besides, the correction for multiple testing on the basis of the same data was made by the Bonferroni method.

We decide to use these non-parametric statistics since some groups presented asymmetrical distribution, which was confirmed by Shapiro-Wilk test ($\alpha = 0.05$). In Figure 4, the Matthews correlation coefficient (MCC) results for each colorspace are illustrated.

We found that the null hypothesis is accepted ($p > 0.05$) regarding the pairwise comparison "YCbCr/CIE Lab", that is, the groups come from populations with identical MCC median values (Figure 4a). Besides, both color spaces presented the best MCC performance in relation to the remaining color models.

From Q_n estimator results, it is possible to determine which colorspace is most stable (Figure 4b). The null hypothesis was accepted ($p > 0.05$) for the pairwise comparisons "HSV/CIE Lab" and "YCbCr/CIE Luv". However, the comparison "YCbCr/CIE Lab" was significant different ($p < 0.05$). This finding pointed out that CIE Lab colorspace is most stable than YCbCr model, since reached the lowest Q_n estimator value.

Table I summarizes the robust statistics of MCC results for each evaluated colorspace. Figure 5 illustrates an example of skin detection by using the CIE Lab colorspace. Due to space limitations in this manuscript, we just show the region of the face in this example. However, the entire image was processed for performance analysis.

TABLE I. ROBUST STATISTICS OF COLOR SPACES IN TERMS OF MCC.

	Median value	**Q_n estimator**
rgb	0.606	0.143
HSV	0.763	0.074
YCbCr	0.776	0.092
CIE Lab	0.779	0.074
CIE Luv	0.729	0.095

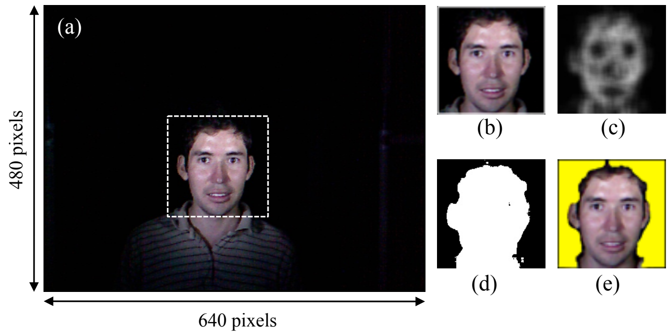

Fig. 5. (a) Original RGB image acquired with Kinect sensor. (b) Region of interest corresponding to subject's face. (c) Outcome of the probabilistic model after converting (b) to CIE Lab colorspace. (d) Pixel classification in skin (white) or non-skin (black) performed by Otsu method. (e) Final skin detection.

IV. CONCLUSION

We are interested in segmenting automatically 3D face sequences acquired with the Microsoft Kinect sensor for further expression analysis. However, first it is necessary to find accurately the face region in the image and this task could be simplified by detecting "skin-like" colors. Hence, the detected skin region could be used to crop the depth map provided by the Kinect's IR sensor.

For this study, we chose the Kinect device due to the following advantages: (i) low-cost (about \$150 dollars), (ii) it is portable and easy to handle, and (iii) there exists open source frameworks to access and process its data.

It is well known that colorspace transformations could improve skin detection performance. Therefore, in this paper we present a comparative study of five color spaces commonly used for detecting human skin. The evaluated models were: normalized RGB (rgb), HSV, YCbCr, CIE Lab, and CIE Luv. These color spaces attempt to separate the luminance from chrominance components, which is useful to make the skin detection illumination independent.

We built a controlled recording environment for acquiring color images by using the Kinect's RGB camera. The goal is to reduce the scene complexity to attempt improving the automatic face segmentation. Besides, it was acquired an image dataset composed of 705 RGB images from 47 subjects. Additionally, each image in dataset was segmented manually for identifying true skin pixels for the classifier's supervised training.

A probabilistic model (based on Bayesian rule) was trained for each color space to classify a pixel into skin class or non-skin class. Moreover, only the chromatic components were used to compute the parameters of the classifier. Since the probabilistic classifier assigns a probability value to each pixel, we used the Otsu method to establish a threshold to label each pixel as skin or non-skin.

The output of the classifier was a binary image, where the white regions correspond to skin pixels and black regions otherwise. We used the Matthews correlation coefficient (MCC) to measure the quality of the computerized segmentation in relation to its respective binary mask (i.e., true skin pixels) when using a specific colorspace. In this sense, the MCC should tend toward "$+1$" to indicate an adequate computerized skin classification.

The statistical analysis of the MCC results suggested that CIE Lab colorspace reached the best performance for our application, with median value equal to 0.779 and Q_n estimator equal to 0.074. Besides, the YCbCr performed similarly than CIE Lab with median value equal to 0.776; however, it was less stable since Q_n=0.092. On the other hand, the normalized RGB model (denoted rgb) reached the worst MCC performance with median value equal to 0.606 and Q_n=0.143. Note that rgb colorspace did not separate explicitly the luminance component. Therefore, from these statistical results we can conclude that dropping the luminance channel improves the performance of skin detection.

CIE Lab colorspace was designed to approximate human vision and the colors are defined independent of their nature of creation or the device they are displayed on. Hence, these attributes could make the CIE Lab model adequately for detecting human skin.

Finally, we believe that the main contributions of this paper are:

- An experimental methodology for evaluating statistically distinct color spaces for skin segmentation applications.
- A study to compare and determine objectively which color-space is the most adequate for automatic face segmentation.
- The construction of a recording environment for reducing the scene complexity to attempt improving the automatic face segmentation stage.

Future work considers the construction of an automatic system for acquiring 3D face sequences by using the Kinect sensor, where the CIE Lab colorspace will be used for detecting skin face.

Acknowledgment

The authors would like to thank the National Council on Science and Technology (CONACYT, Mexico) for the financial support.

References

[1] S. D. F. Hilado, E. P.Dadios, and R. C. Gustilo,"Face detection using neural networks with skin segmentation," in IEEE 5th International Conference on Cybernetics and Intelligent Systems, 2011, pp. 261-265.

[2] Y. Wang and L. Xia, "Skin color and feature-based face detection in complicated backgrounds," in 2011 International Conference on Image Analysis and Signal Processing, 2011, pp. 78-83.

[3] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods", Pattern Recognition, vol. 40, no. 3, pp. 11061122, 2007.

[4] D. Petrisor, C. Fosalau, M. Avila, and F. Mariut, "Algorithm for face and eye detection using colour segmentation and invariant features," in 34th International Conference on Telecommunications and Signal Processing, 2011, pp. 564-569.

[5] J. S. Schmugge, S. Jayaram, M. C. Shin, and L. V. Tsap, "Objective evaluation of approaches of skin detection using ROC analysis," Comput. Vis. Image Underst., vol. 108, pp. 41-51, 2007.

[6] X. Yang and F. lv, "The design of a face recognition system based on skin color and geometrical characteristics," in 2011 International Conference on Image Analysis and Signal Processing, 2011, pp. 276-279.

[7] "Microsoft Xbox 360", online, accessed September 2012, retrieved by www.xbox.com/en-US/xbox360.

[8] J. D. Foley, A. v. Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice.* New York: Addison Wesley, 1990.

[9] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed., New York: Academic Press, 2009.

[10] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst., Man, Cybern., vol. 9, pp. 62-66, 1979.

[11] J. K. Udupa, V. R. LeBlanc, Y. Zhuge, et al., "A framework for evaluating image segmentation algorithms," Computerized Medical imaging and Graphics, vol. 30, pp. 75-87, 2006.

[12] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," Bioinformatics, vol. 16, no. 5, pp. 412-424, 2000.

[13] P. J. Rousseeuwand and C. Croux, "Alternatives to the median absolute deviation," Journal of the American Statistical Association, vol. 88, no. 424, pp. 1273-1283, 1993.

IEEE Catalog Number CFP13827-ART
ISBN 978-1-4799-1461-6
978-1-4799-1461-6/13/$31.00 ©2013 IEEE