

Research article

Objective and Subjective Assessment of Digital Pathology Image Quality

Prarthana Shrestha *, Rik Kneepkens, Gijs van Elswijk, Jeroen Vrijnsen, Roxana Ion, Dirk Verhagen, Esther Abels, Dirk Vossen, and Bas Hulsken

Philips Digital Pathology Solutions 5684 PC Best, The Netherlands

* **Correspondence:** Email: prarthana.shrestha@philips.com.

Abstract: The quality of an image produced by the Whole Slide Imaging (WSI) scanners is of critical importance for using the image in clinical diagnosis. Therefore, it is very important to monitor and ensure the quality of images. Since subjective image quality assessments by pathologists are very time-consuming, expensive and difficult to reproduce, we propose a method for objective assessment based on clinically relevant and perceptual image parameters: sharpness, contrast, brightness, uniform illumination and color separation; derived from a survey of pathologists. We developed techniques to quantify the parameters based on content-dependent absolute pixel performance and to manipulate the parameters in a predefined range resulting in images with content-independent relative quality measures. The method does not require a prior reference model. A subjective assessment of the image quality is performed involving 69 pathologists and 372 images (including 12 optimal quality images and their distorted versions per parameter at 6 different levels). To address the inter-reader variability, a representative rating is determined as a one-tailed 95% confidence interval of the mean rating. The results of the subjective assessment support the validity of the proposed objective image quality assessment method to model the readers' perception of image quality. The subjective assessment also provides thresholds for determining the acceptable level of objective quality per parameter. The images for both the subjective and objective quality assessment are based on the HercepTestTM slides scanned by the Philips Ultra Fast Scanners, developed at Philips Digital Pathology Solutions. However, the method is applicable also to other types of slides and scanners.

Keywords: subjective image quality; objective image quality; perceptual metrics

1. Introduction

The quality of the digital pathology images has a direct influence not only on the accuracy of the diagnosis but also on the readers' performance and reliability [1,2]. Additionally, the performance of the automated detection and diagnosis tools designed for clinical assistance are dependent on the information in the image content. Therefore, the images produced by the Whole Slide Imaging (WSI) scanners should represent an accurate and consistent quality.

The ideal judge of quality of digital pathology images would be the pathologists. However, in practice, such an assessment is very time-consuming, expensive and difficult to reproduce. A large number of pathologists are required to overcome the subjectivity involved in the assessment and to obtain a reliable result. Therefore, we propose a method to objectively assess the image quality, which is in close agreement with the pathologists' opinion. The proposed image quality assessment method is applicable to various purposes: (i) dynamically monitor the scanner performance (e.g., focus) and slide quality (e.g., color intensity of the stains); (ii) assist in the design and implementation of algorithms, which require iterative testing and optimization (e.g. data compression and computer aided detection/diagnosis); (iii) impartial evaluation and benchmarking of the imaging systems (e.g., scanners and microscopes), and software tools (e.g., codecs and filters).

In the existing literature regarding image quality analysis in pathology, objective and subjective image quality measures are used in comparing the microscopy and scanned images in [3]. The quality is defined as 'a good image should preserve those features which facilitate the pathological assessment of the tissues, i.e. whether the tissue is diseased or not. The objective quality assessment involved measuring the following metrics: color, contrast homogeneous luminance and texture. The subjective quality assessment involved psychophysical tests of the images involving pairwise comparison of the scanned and microscopic images. In both the objective and subjective assessments the quality of the scanned images is found to be slightly higher than the microscopic images. The most differentiating metric between the two systems were found to be focus and color. Additional work on the objective image quality analysis are aimed to improve or normalize the image quality [4]; to select optimal images for analysis; and to acquire the quality information as a criteria on making automated clinical decisions [5,6].

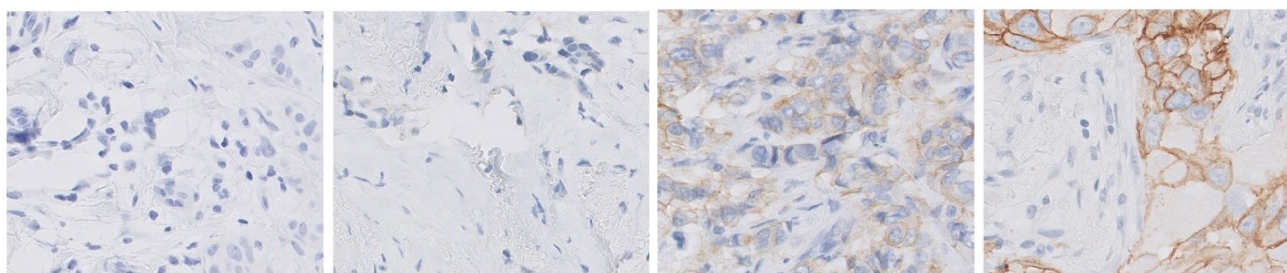


Figure 1. Example images cropped from four whole slide images. The HER2 scores of the images from left to right are: 0, 1+, 2+, and 3+ respectively.

The existing objective quality measurement involves different metrics and there is no universally accepted metric for defining image quality. The objective assessments are based on measuring multiple image related features e.g. light and color consistency [5]; intensity values such as brightness color analysis and histogram [7]; edge sharpness and noise [6]. In our evaluation measuring quality according to the image metrics [6,8] are found to be highly influenced by the

image content and therefore, not suitable for comparing image quality among different content. For example the brightness metric given by the mean luminance intensity of an image corresponding to a dark-color stain is more likely to be lower in value than another image corresponding to a light-color stain irrespective of the brightness quality in the images. Moreover, the existing works do not address the correlation between the subjective and objective measurement of image quality and their significance in clinical diagnosis.

In this paper, we propose a method to objectively assess image quality, which is independent of image content and corresponds to the perceptual or subjective image quality. We consider image quality as a measure of suitability of the image for clinical diagnosis. The image parameters for assessing quality are acquired from a survey of pathologists based on their perceptual and clinical significance. Then the parameters are formalized in terms of robust and quantifiable metrics; and corresponding image processing filters are developed to simulate controlled quality degradation over a range. The optimal quality images and their distorted versions are presented to pathologists for a subjective assessment of quality. The subjective assessment is used to validate our objective assessment and to determine threshold for an acceptable level of image quality. The assessment protocol is reviewed and approved by the IRB before the start of the study.

The images for both the subjective and objective quality assessments are based on the HercepTestTM [9] slides scanned by the Ultra Fast Scanners, developed at Philips Digital Pathology Solutions. The HercepTest is based on the immunohistochemistry, a special staining process performed on breast cancer tissues, where the amount of the HER2 receptor protein is represented by intensity and presence of blue and brown stains. The clinical diagnosis of the HercepTest involves providing scores: 0, 1+, 2+, and 3+ according to the presence of the HER2 receptor protein. Figure 1 shows four example HercepTest images, each representing a score category. The motivation for using the HercepTest slides for the proposed quality assessment is driven by another associated experiment designed to analyze diagnostic accuracy of pathologists using WSI. Since the diagnosis of the HercepTest slides are given in terms of quantitative scores, instead of qualitative or descriptive texts, it allows application of objective and statistical methods for analysis. The coverage on the diagnostic accuracy experiment is beyond the scope of this paper, however, as a next step we aim to correlate diagnostic accuracy with image quality. Therefore, we choose the HercepTest samples for image quality assessment. But the proposed assessment method is generic and also applicable to other types of tissues and stain types.

Table 1. Mean rankings of the image quality parameters given by pathologists in the order of importance (1 represents the most important parameter).

Image quality	Mean	Weight
Sharpness	1.25	0.377
Contrast	3	0.189
Brightness	3.4	0.166
Uniform illumination	4	0.141
Color separation	4.5	0.126
Image orientation	5.8	–
Color temperature	6	–

The top 5 parameters are selected for using in the objective image quality assessment. The weights represent the contribution of the individual parameters in the overall quality computation, where the sum of the weights equals to 1.

2. Materials and Method

2.1. Acquiring relevant image quality parameter

We start by conducting a survey involving pathologists, to acquire image quality parameters relevant for clinical diagnosis in HercepTest. The survey consisted of a brief explanation of the following 7 widely accepted image quality parameters: brightness, contrast, sharpness, color temperature, orientation, uniform illumination, and color separation followed by a questionnaire. The parameters represent image features with perceptual and diagnostic importance and are also sensitive to the performance of the WSI scanners. The survey was participated by eight pathologists, who were board certified members and working in two hospitals located in The Netherlands and Belgium. They had, at minimum, two years of experience with WSI. They were asked to rank the parameters according to the order of importance. We also asked an open question to suggest any other relevant parameters missing from the given list. Based on the results, we select the following 5 highest ranked parameters (given in the decreasing order of importance): sharpness, contrast, brightness, uniform illumination, and color separation. Table 1 shows the mean ranks given by the pathologists and the derived weight per parameter. The weights per parameter is used to compute a content-dependent overall image quality, described in Section 2.2.2. If q and w represent quality measure and weight, respectively, of a parameter represented by i , the overall quality Q of an image based on the selected five parameters is given by:

$$Q = \sum_{i=1}^5 w_i q_i . \quad (1)$$

2.2. Test-set generation

For the quality assessment, we use 12 HerepTest slides, uniformly representing the following HER2 score categories: 0, 1+, 2+, 3+. The slides were stained using Dako equipment and HercepTest kit. Each slide represents a unique individual case of invasive breast carcinoma, seen typically in a clinical practice. All the slides were pre-screened using an optical microscope by a Dako pathologist, who was not involved in this study, to ensure sufficient quality of the slide specimen and to obtain equal distribution of the score types.

The same slide scanned by a scanner at different times or by different scanners may appear slightly different, even when viewed on the same display, due to discrepancies in scanner characteristics or changes in configuration due to external influences like temperature. Therefore, to address the variations, which may occur during repetitive scans along time and across scanners, we scan the test slides 10 times in three different scanners. In the first whole slide image, a region of interest (ROI) representing an area meaningful for clinical diagnosis, is selected by a pathologist at 40× magnification. A ROI is given by an image dimension of 900 × 900 pixels. Next, the corresponding ROI in all the 30 whole slide images of a slide are searched and aligned. We use the deformable image registration based method to align the ROIs, as described in [10]. The alignment accuracy is verified visually and in case of a failure, the alignment method is repeated with manual inputs until an alignment of approximately 95% is achieved. The procedure is repeated for all the test

slides. Figure 1 shows example images of four ROIs, each representing a score category. For further analysis, the extracted images of the aligned ROI are used instead of the whole slide images.

The test-set images for the subjective quality assessment are prepared such that they represent a wide range of quality from being optimal to unacceptable and include a sufficient number of intermediate quality variations. We select an optimal quality image, given multiple images corresponding to the same ROI or content from different scans, in terms of absolute pixel performance, called metric values. Then to generate images with varying quality, the image with the highest metric value, called *reference image*, is subjected to a predefined six levels of image degradation, called distortion values. The distortion values correspond to a content independent and relative measure of image quality, which is normalized with respect to the reference image. The resulting images represent quality degradations due to scanner malfunction.

Figure 2 describes the schematic of the method used in generating test-images, in terms of data and process flow. Each data element is an input or output of the corresponding process step. The method is repeated for all of the 12 HercepTest slides to generate a complete test-set images for the subjective study. The following sections describe the methods used in computing metric values, finding reference image and distorting the reference image

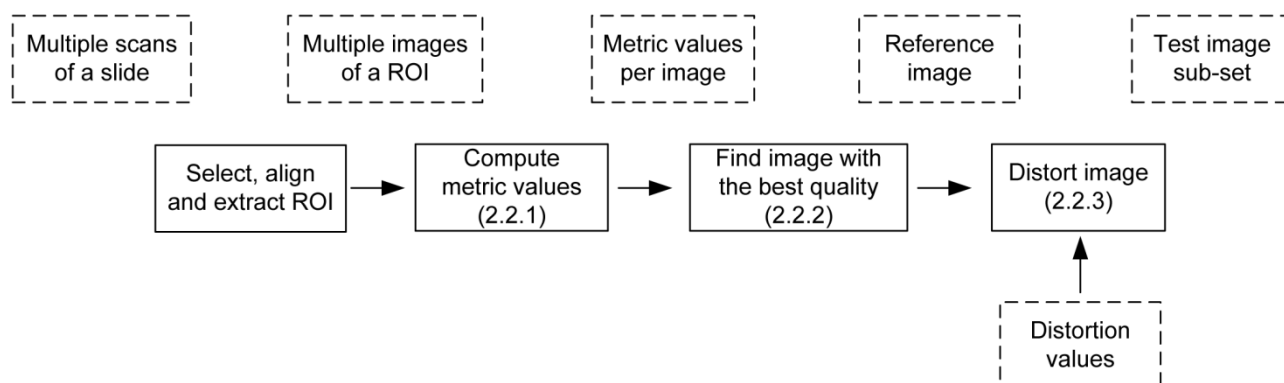


Figure 2. Schematic of the method used in generating test images for the subjective assessment, given a multiple number of whole slide images of a slide scanned by different systems. The method is repeated for all the test-slides to generate a complete set of test images. The blocks given by dotted and solid lines represent data and process flow, respectively. The sections containing the description of the processes are indicated by numbers in brackets.

2.2.1. Metric value computation

A metric value represents an absolute pixel performance of an image content, in terms of a quality parameter. The metric value computation requires an image to be transformed from its native sRGB color space [11], reproduced by a scanner, to the YCbCr color space [12], where the luminance and chrominance channels are separated. For the metric value computation of the different parameters described below, the luminance(Y) and chrominance (Cb and Cr) components are taken from the transformed YCbCr image. The metric values are represented by the numbers between 0 (low quality) and 1 (high quality).

Sharpness metric. Given two images with the same content, a sharper image contains more edges than a blurred image. We tested 18 different algorithms for measuring sharpness, presented in [8], and select the method described in [13] based on its performance results. In this method, edges

in an image are detected by using Sobel approximation of the luminance channel in the both horizontal and vertical directions. A pixel is considered as an edge-pixel if it is above a given threshold, which is set to be 0.005, based on our test results. The low threshold value supports the presence of very delicate edges present in pathology images. The sharpness metric value of an image is derived as the number of pixels classified as edge pixels over the total number of pixels

Brightness metric. Brightness corresponds to the visual perception of the amount of light coming from an image in display. We measure brightness metric of an image as the mean of luminance component, where values 0 and 1 correspond to an absolute black and white image, respectively.

Contrast metric. Contrast represents the difference between the darkest and the lightest pixels in an image [14]. The range of pixels present in an image is represented as a histogram of its luminance component in 256 bins. To discard any noise present in an image, the darkest and the brightest 1% of the pixels are ignored from the histogram. The contrast metric value of an image is computed as the width of the histogram over the 256 (total number of bins). The metric value 0 corresponds to no contrast and 1 corresponds to the maximum contrast.

Color separation metric. Color separation is given by the divergence between the blue and brown colors in the HercepTest images. We use the Principal Component Analysis (PCA), which de-correlates an image data (RGB) into components according the amount of information present in the data [15], resulting in different color planes. Since the second PCA component represents pixels in blue-brown color plane, the color separation metric is measured as a variance of the second PCA component.

Uniform illumination metric. Uniform illumination is perceived as the consistency in luminance across an image background. The luminance difference may be due to the image content itself and/or due to errors during scanning. To remove the content dependency, we first identify the background areas without any tissues in the luminance channel of an image. The background area is detected based on the absence of edges and color, using the method described in [16]. The luminance uniformity metric is measured using the contrast metric measurement, as described above, in the detected background area. The measurement is considered as unspecified if no background areas are detected and if the areas are limited only to a certain locations in an image.

2.2.2. Reference image selection

The overall image quality score is computed by adding the metric values of the five image quality parameters, according to the weights of the individual parameters obtained from the pathologists' survey, as given in Equation 1. Given a multiple scanned images of the same ROI, the highest scoring image is selected as a reference image. The image represents an optimal quality that is suitable to be used in clinical purposes

2.2.3. Image distortion

The metric values correspond to the absolute pixel performance and are dependent not only on the image quality but also on the image content. They do not allow fair comparison of quality among images with different content. Therefore, we use image filters corresponding to the quality parameters to distort a reference image in a controlled manner, according to a predefined set of

distortion values. The distortion values represent a relative measure of image quality with respect to its reference. Such a normalized quality representation allows comparison of images with different content. A distortion value of 0 refers to a reference image with an optimal quality, while the values diverging from 0 represent decreasing image quality compared to the reference. For each of the quality parameters, we apply six equally spaced distortion values in a predefined range resulting in the six test images per ROI to be used in the subjective assessment. The distortion range is selected based on the results of a pilot study such that it covers the perceptual limit from not-noticeable to unacceptable level of distortion. The distortion values for brightness are represented between -0.5 (darker) and 0.5 (brighter), and the rest of the parameters are represented between 0 (reference image) and -1 (most distorted). Figure 3 shows an example reference image after applying the different levels of distortion.

Sharpness distortion. Sharpness distortion is performed by applying a Gaussian filter [13] to an input image. The strength of the Gaussian window is determined by the input distortion value, where a more negative factor means a wider Gaussian filter resulting in a more blurred image. The minimum and maximum distortion is achieved by applying the Gaussian filters of pixel-size 3×3 and 16×16 , respectively.

Brightness distortion. The brightness of an image is distorted by multiplying its pixel values of the luminance (Y component), according to the given distortion values. The distortion values are in the range $[-0.5, 0.5]$, where 0 means no image distortion. If the value is greater than 0, the image is made brighter and if the value is less than 0, the image is made darker. The images with the maximum brightness distortion are 0.25 times the brightness of its reference (at distortion value -0.5) and 1.75 times the brightness of its reference (at distortion value 0.5).

Contrast distortion. Contrast distortion is applied by reducing the difference between the lightest and the darkest pixels in an image. The histogram range of an input image is reduced according to the distortion value. Although contrast is measured only in the luminance component, in order to prevent visual artifacts the distortion is applied to all channels. At the maximum distortion level of -1 , the luminance histogram is compressed by 75%, and the chrominance histograms are compressed by 45%.

Color separation distortion. The color separation distortion is applied to reduce distance between the brown and blue colors in an image. The input image is first transformed into PCA space. Then the spread of the values in the second PCA component is reduced according to the given distortion values. Subsequently, the PCA data are transformed back to the sRGB color space. At the maximum distortion level of -1 , the value of the second component of the PCA is reduced to 0.

Uniform illumination distortion. The uniform illumination distortion is performed by multiplying a mask image to the luminance channel of an input image. This mask contains spatially increasing darkness values starting from the image center, symmetrically in the horizontal direction. The distortion value determines the level of darkness. At the maximum distortion level of -1 , the outer most pixels at the left and right side of the image are multiplied by 0.10.

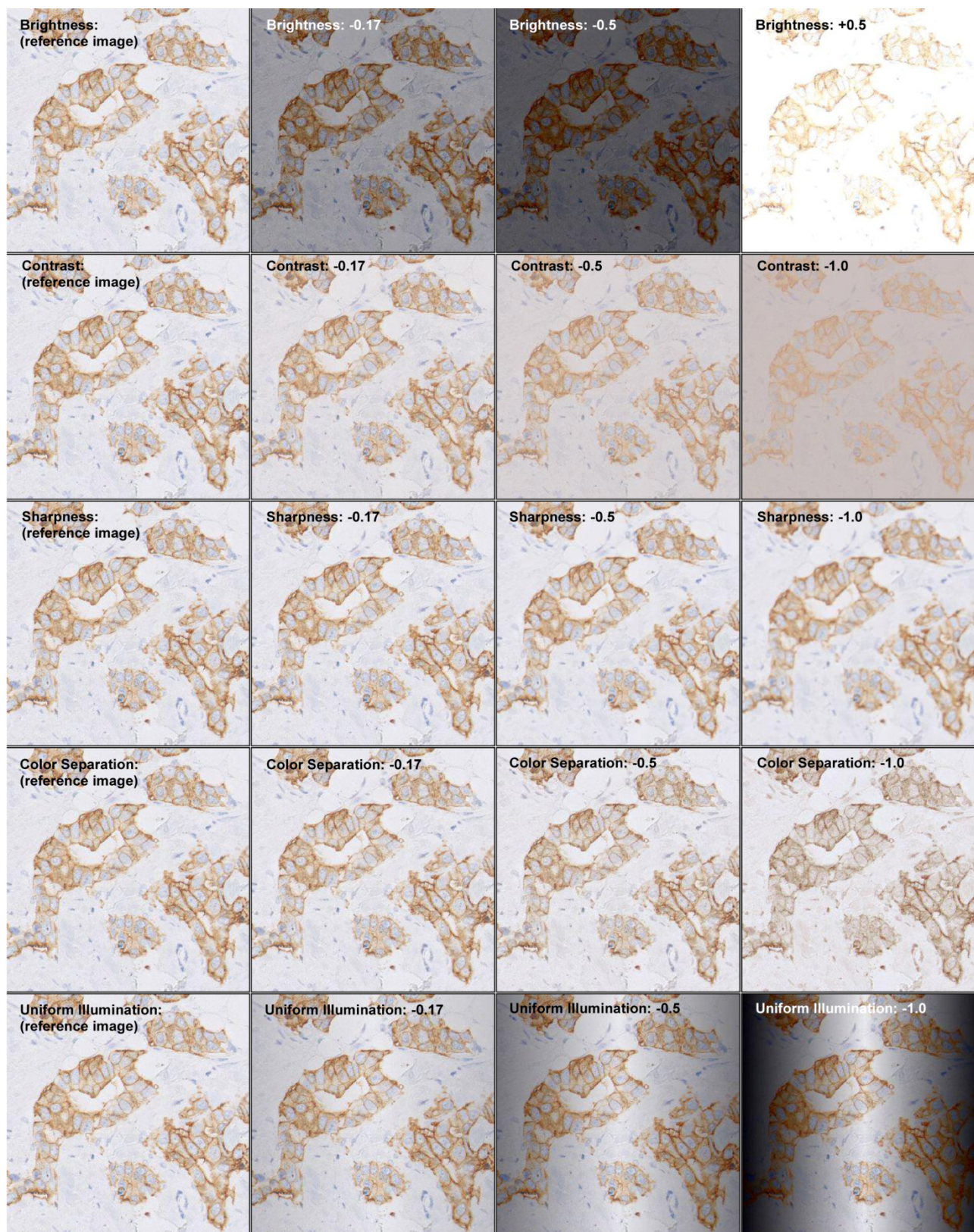


Figure 3. An example image with different level of distortion (given in column) per metric (given in row). The images in the first column are the unmodified, with the distortion value 0. The images have been cropped and downscaled for illustrative purpose.

2.2.4. Subjective test

The subjective assessment gathers quantitative rating of pathologists' opinion on image quality. The ratings provide a perceptual score to the distortion values. In addition, we can derive thresholds for level of distortion per quality parameter, where pathologists are no longer comfortable with providing a score.

During the assessment, pathologists were presented with test images and asked to answer the following question on a 1–5 scale:

Please rate: how comfortable are you with the quality of this image for providing a HercepTest score

1: Very Uncomfortable; 2: Uncomfortable; 3: Acceptable level of Comfort;
4: Comfortable; 5: Very Comfortable

The test was participated by 69 pathologists in total, where 22 were from Europe (EU) and 47 were from The USA (US), working in 7 different hospitals. All the pathologists were board certified members with experience in WSI. A software tool, called Presentation® available at [17], was used to visualize the images in random order to each of the participants and record the ratings. All the tests were performed using a single BARCO MDCC 2121 monitor, with resolution of 2MP (1600 × 1200) and pixel pitch of 0.27 mm. The luminance level of the display was set to be 240cd/m² and no settings could be changed by the participants. Similarly, no view manipulations such as panning or zooming of the images were possible.

At the start of the test, the readers were asked to rate a set of example images which were not used for the final analysis. This gave the participants a familiarity with the range of distortions they could expect during the test, and they could calibrate their ratings accordingly. The rating of 3 was explicitly defined as the level at which they are just comfortable enough for providing a HercepTest score. They were also asked explicitly to provide ratings regarding image quality only, regardless of whether the image itself is difficult to score or not. The test-set included one reference (undistorted) image and six distorted images of the predefined distortion levels per ROI, per quality parameter. In total, each pathologist rated a set containing $12 \text{ ROI} \times (5 \text{ parameters} \times 6 \text{ distortion level} + 1 \text{ undistorted})$ 372 images. Figure 3 shows example images of a ROI with different levels of distortion per quality parameter used in the subjective assessment.

The assessment protocol was reviewed and approved by the IRB (00007807) before the start of the study.

3. Results and Discussion

Figure 4 shows subjective ratings given by the readers and analysis results for the 5 image quality parameters: brightness, contrast, sharpness, color separation and uniform illumination. The ratings range between 1 to 5, where 1 represents low quality and 5 represents high quality image. The distortion values of the test set images range from -1 to 0 , where -1 represents unacceptable quality and 0 represents optimal quality. We expect that the perceived level of quality correlates with the level of distortion. The ratings from individual readers (in gray curves) show a large inter-reader variation across the whole rating and distortion ranges. As expected, the mean rating, obtained by combining ratings of all the individual pathologists, per distortion level (in red circles) decreases

with the distortion value resulting in a monotonous and continuous curve.

Given a large variation in rating among individual readers, the mean rating value does not provide a robust representation of the rating such that it can be applied to compute a distortion threshold that correspond to an acceptable level of quality. Therefore, we compute one-tailed 95% confidence interval of the mean rating and use the lower interval value as a representative rating value. The confidence interval is derived from the results of a two-way ANOVA with slides and readers as independent random factors and distortion values as dependent parameter. Figure 4 shows the representative rating curve, as the lower limit of the 95% confidence interval per distortion value in the solid red lines. The fact that each of these representative curves are smooth, monotonically decreasing with distortion values, and cover a significant range of the rating axis (including the acceptable threshold of 3) support the validity of the proposed distortion algorithm, and suitability of the proposed objective image quality assessment method to model the readers perception of image quality.

The distortion level threshold, computed as the intersection between the representative reader curve and the rating equal to 3 are: sharpness -0.43 , contrast -0.48 , brightness -0.26 – -0.24 , uniform illumination -0.41 , and color separation -0.78 . It implies that there is a 95% confidence that readers find the image quality above the thresholds as acceptable for clinical diagnosis.

We also analyze differences in the ratings among images with different HER2 score types, which are distinguished by image content such as cellular structure and color. The analysis provides a qualitative indication of the influence of content in pathologists' rating. If the ratings would be based entirely on quality, and not on content, the different score type images with the same distortion level would be rated similar. However, our analysis shows that the mean ratings corresponding to the 3+ images are significantly higher at every distortion level than the images of the other score types in parameters: sharpness, contrast, uniform illumination and brightness. In the color separation parameter, the images with 3+ are rated significantly higher than the other score type images at the lower distortion levels, while the ratings become similar at the higher distortion level. Figure 5 shows the mean ratings of the parameters: sharpness, contrast and color separation for the test-images with different HER2 score types. It means that quality is less important to diagnose a 3+ image and the ratings are not entirely dependent in image quality but also in content, at least in the given range of distortion values. As seen in the color separation parameter, at very low distortion values all images irrespective of the score types, become ambiguous, saturated with either brown or blue color, resulting in equally poor ratings.

The subjective assessment was conducted in two phases. In the first phase, there were 31 readers out of which 22 readers were from EU and 9 were from US. During the analysis of the variability of the individual reader curves, it was found that the US readers rated the image quality significantly different from the EU readers. The number of US readers was relatively small and they were instructed by a different trainer than the EU readers. Upon discussion with the US readers, it became clear that they felt uncomfortable providing a HER2 score for the whole slide based on a limited ROI. As this was not our research question, we decided to run a second assessment with a larger number of US readers with an improved training procedure. In the second phase, 38 additional US readers participated in the assessment. The analysis results of the second phase of assessment were very similar

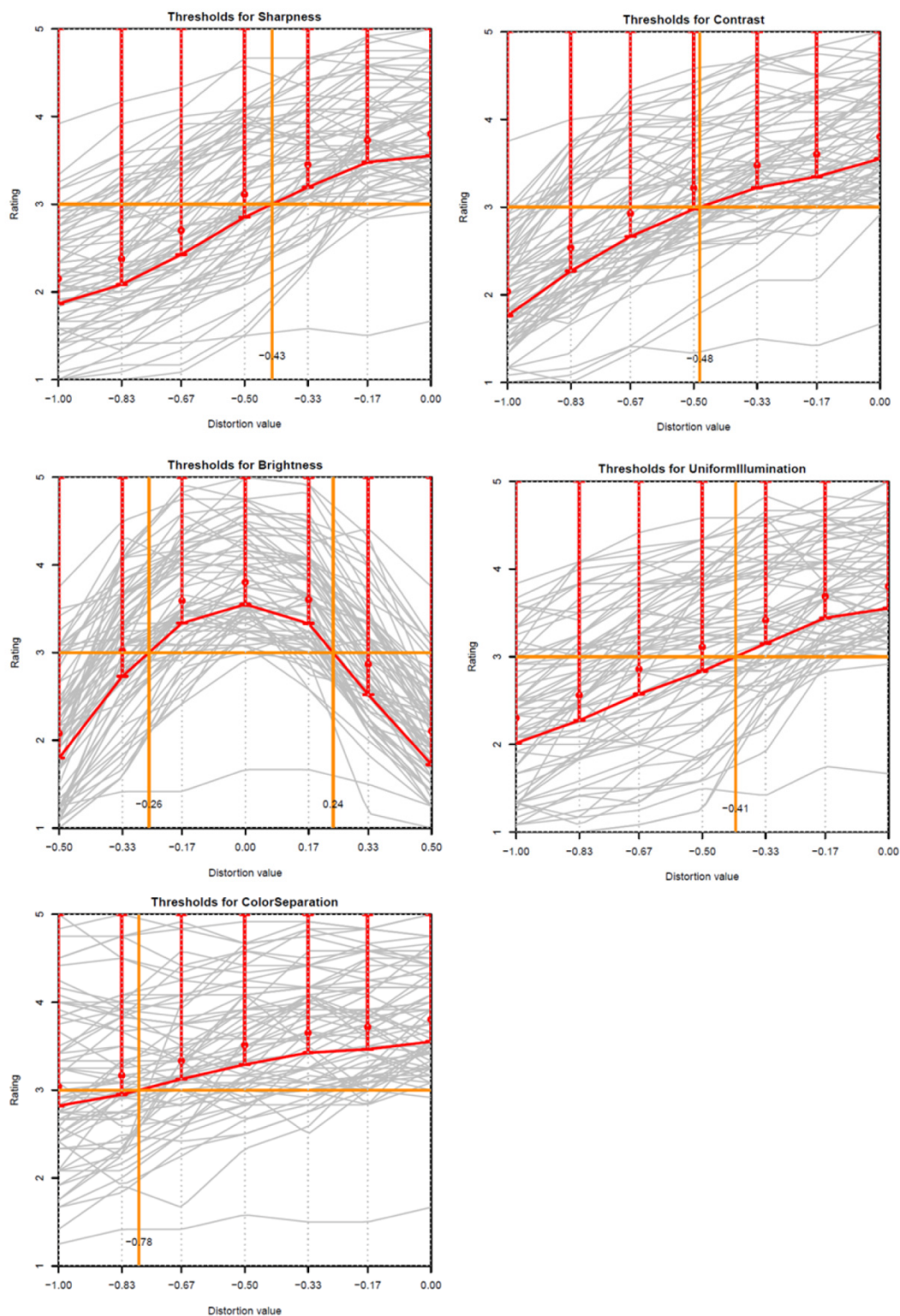


Figure 4. Pathologists' perception rating as a function of distortion value for the five image quality parameters (from left to right and top to bottom): Sharpness, Contrast, Brightness, Uniform illumination, and Color separation. Individual reader perception curves (gray), mean rating of pathologists (red circle), representative curve (red solid line) corresponding to 95% one-tailed confidence interval (red dashed line) and threshold for an acceptable image quality (intersection of rating value 3 and representative curve given by orange lines).

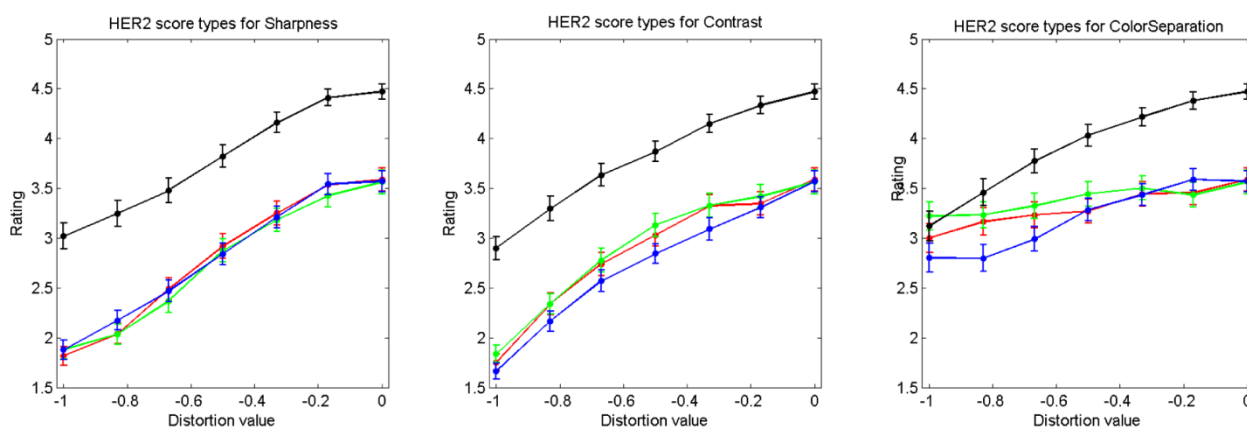


Figure 5. Mean pathologists' rating for sharpness (left), contrast (mid) and color separation (right) parameters corresponding to images with HER2 scores 0, 1+, 2+ and 3+ shown in red, green, blue and black curves, respectively. The error-bars represent 95% confidence interval of the mean value.

4. Conclusion

In this paper, we assess image quality in terms of suitability of the images for clinical diagnosis by means of objective and subjective analysis. The main contributions of the paper can be summarized as:

- Identification of the five most relevant image quality parameters for the HercepTest diagnosis: brightness, contrast, sharpness, color separation and uniform illumination, acquired in a survey of pathologists.
- Method to quantify the image parameters based on absolute pixel performance (content dependent), called metric values and relative degradation of image quality (content independent), called distortion values.
- Application of a range of distortion values to optimal quality images such that a substantial variation in perceptual quality including an acceptable distortion level is covered.
- Model for image quality perception by pathologists' such that the subjective quality rating correlates with the objective image distortion values.
- Thresholds in terms of distortion values corresponding to an acceptable level of comfort, with single-tailed 95% confidence, for clinical diagnosis.

In the existing literature, the objective image quality of digital pathology images is evaluated using visual parameters, such as color and sharpness, based on pixel performance. This evaluation method is comparable to our metric value computation method, however, its performance is content dependent and cannot be generalized to compare quality among images with different content. The novelty in our approach is in translating metric values to distortion values, which provide content independent measure of quality. Furthermore, the performance of our objective evaluation is verified with the diagnostic or subjective quality.

The described subjective and objective quality assessment methods are generic. In this paper, we use the methods to assess the quality of HercepTest images, however, they are likely to be applicable also for other tissues or stains with an updated list of image quality parameters and

distortion range.

A limitation of the described objective assessment method is that it requires an optimal quality image to use as reference. We select a reference image based on the highest scoring metric values among multiple scans of a slide. In case the reference image does not correspond to an optimal quality, the method may not correlate with the subjective assessment. Similarly, if the choice of an optimal quality image is not unique, for example when images are obtained using scanners from different manufacturers, the method becomes inapplicable. Another limitation of the current assessment methods is that they are based on a localized ROI and lack the quality overview of a whole slide image.

In the current subjective assessment, our hypothesis is that the perceived level of comfort by a pathologist in performing a diagnosis is dependent on image quality. Therefore, we presented a pathologist with a single image at a time for evaluation. However, the perceived image quality is found also to be influenced by content. For example, in our analysis the images with score type 3+ are found to be rated significantly higher than other score type images. It means that quality is less important to diagnose a 3+ image. To minimize such a bias due to image content, pathologists can be asked for a pairwise comparison between an image with its reference.

The subjective assessment results in a large inter-reader variation, as commonly seen in the field of perception research. The ratings are found to be influenced by variations in the training procedure performed before the actual assessment. A similar variation can be expected in case the question to the readers is phrased differently. Therefore, for future tests the training procedure and the question are required to be standardized.

As a next step, we employ the tools developed for computing the objective image quality in measuring reproducibility of the WSI scanners. The thresholds acquired from the subjective assessment are used to set a limit such that the image quality parameter variations exceeding the thresholds are considered unacceptable. Furthermore, based on the existing framework, we plan to design experiments to evaluate image quality including different tissues, stains and image parameters; and also to derive relationships between image quality and diagnostic accuracy/performance.

Conflict of Interest

The authors declare no conflict of interests regarding this paper.

References

1. Pantanowitz L (2010) Digital images and the future of digital pathology. *J Pathol Inform* 1: 15.
2. Krupinski EA, Silverstein LD, Hashmi SF, et al. (2012) Observer performance using virtual pathology slides: Impact of LCD color reproduction accuracy. *J Digit Imaging* 25: 738-743.
3. Redondo R, Bueno G, Cristbal G, et al. (2012). Quality evaluation of microscopy and scanned histological images for diagnostic purposes. *Micron* 43: 334-343.
4. Murakami Y, Gunji H, Kimura F, et al. (2012) Color correction in whole slide digital pathology. *Twentieth Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*. pp. 253-258.
5. Korzyska A, Neuman U, Lopez C, et al. (2010) The method of immunohistochemical images standardization. *Image Processing and Communications Challenges 2, Advances in Intelligent*

- and Soft Computing*. 84: 213-221.
6. Hashimoto N, Ohyama N, Bautista PA, et al. (2012) Referenceless image quality evaluation for whole slide imaging. *J Pathol Inform* 3: 9.
 7. Kayser K, Grtler J, Metze K, et al. (2008) How to measure image quality in tissue-based diagnosis (diagnostic surgical pathology). *Diagn Pathol* 3 (Suppl 1): S11.
 8. Duthaler S, Sun Y, Nelson BJ (2005) Autofocusing algorithm selection in computer microscopy. *Intelligent Robots and Systems, 2005. 2005 IEEE/RSJ International Conference on IEEE*: 419-425.
 9. Dako HercepTest™ Product information. Available from: http://www.dako.com/28633_04may10_herceptest_brochure-9086.pdf.
 10. Vossen D, Mueller D and Hulsken B (2011) Real-time deformable registration of multi-modal whole slides for digital pathology. *Comput Med Imaging Graph* 35: 542-556.
 11. IEC 61966-2-1: 1999 Color management default RGB colour space - sRGB.
 12. ITU-R BT.601: Definition for RGB to YCbCr color space conversion.
 13. Yeo TTE, Ong SH, Jayasooriah and Sinniah R (1993) Autofocusing for tissue microscopy. *Image Vision Comput* 11: 629-639.
 14. Thompson W, Fleming R, Creem-Regehr S and Stefanucci JK (2011) Visual perception from a computer graphics perspective. A. K. Peters Ltd..
 15. Tzeng D (1999) Spectral-based color separation algorithm development for multiple-ink color reproduction. Ph. D. the- sis, R.I.T., Rochester, NY.
 16. Javed O, Sha_que K and Shah M (2002) A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information. *IEEE Workshop on Motion and Video Computing*.
 17. Presentation ® (2014) Available from: http://www.neurobs.com/menu_presentation/menu_features/features_overview.

© 2015, Prarthana Shrestha, et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)