



## Quality evaluation of microscopy and scanned histological images for diagnostic purposes

R. Redondo<sup>a,b</sup>, G. Bueno<sup>a,\*</sup>, G. Cristóbal<sup>b</sup>, J. Vidal<sup>a</sup>, O. Déniz<sup>a</sup>, M. García-Rojo<sup>c</sup>,  
C. Murillo<sup>c</sup>, F. Relea<sup>c</sup>, J. González<sup>c</sup>

<sup>a</sup> Visilab, E.T.S. Ingenieros Industriales, Universidad de Castilla-La Mancha, Spain

<sup>b</sup> Instituto de Óptica, Consejo Superior de Investigaciones Científicas, Spain

<sup>c</sup> Dpto. Anatomía Patológica, Hospital Universitario General de Ciudad Real, Spain

### ARTICLE INFO

#### Article history:

Received 1 June 2011

Received in revised form

21 September 2011

Accepted 23 September 2011

#### Keywords:

Histological image processing

Microscopic image quality

Whole slide imaging (WSI)

Psychophysical tests

Objective Perceptual Contrast

### ABSTRACT

In this work we present a study for assessing and comparing the fidelity of biopsy and cytology images captured with two different devices, that is optical microscopes and scanners, at 40× magnification in bright field. The devices use different ways to magnify images. Microscopes use a set of lenses while scanners capture light through arrays of micro-photoreceptors. The objective is to carry out a quantitative evaluation to discern which of the two devices provides better image quality in terms of contrast, colour and stain. Since there is no unanimous consensus on quality metrics, we will make use of both an objective metric based on perceptual features, together with a subjective psychophysical test as the International Telecommunications Union (ITU) recommends in ITU-R BT.500 for such type of tests. Both techniques indicate a slight preference for the scanner over the microscope in terms of better image quality, considering defocus as the main problem followed by colour distortions. However, the image quality of both devices is suitable for clinical, educational and research purposes.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

Advances in new technologies for complete slide digitization in pathology have led to a wide spectrum of technological solutions for whole-slide scanning. They may be classified into two main categories, motorized microscopes and scanners (García-Rojo et al., 2006). The first one is made of a set of lenses which properly aligned magnify the light reflected by the sample into the eye or another light detector. For focusing in the proper focal plane, the slice is often held by a motorized platform controlled via software. The second one captures the light reflected by a glass slice with the sample inside through a co-aligned array of photodetectors. A comparison of 31 commercially available digital slide systems in pathology, describing the most relevant characteristics of the scanning devices, was carried out by the authors in García-Rojo et al. (2006), where most of the systems evaluated allow a high-resolution digitization of the whole slide. Fig. 1 shows an example of a tissue sample acquired with both devices.

Table 1 summarizes the main features of these devices. Motorized microscopes are more flexible and customizable than the scanners. They have the same functionality as traditional ones, but

featuring motorized components. Thus they have a wide range of objective lenses, that is 2.5×, 5×, 10×, 20×, 40×, and 60× and they are able to digitize in bright and fluorescence field. This may help different diagnostic process. However, the digitization speed at 40× (mm<sup>2</sup>/s) is lower than in scanners.

The previously cited study García-Rojo et al. (2006) concludes that the image quality of current virtual microscopy systems is suitable for clinical, educational and research purposes. However, the study lacked a quantitative comparison between the image quality provided by the motorized microscopes and the scanners. This article provides this quantitative assessment. Thus, the image quality is evaluated in terms of focusing, contrast, stain, colour, brightness, overexposure and sharpness for biopsies and cytology tissue samples digitized at 40× bright field.

The first question that arises is what should be the criterion for assessing the fidelity of both types of images. The first obvious answer is: a good image should preserve those features which facilitate the pathological assessment of the tissues, i.e. whether the tissue is diseased or not. From the point of view of image processing, this criteria is translated into preservation of features such as matching colour, high contrast, homogeneous luminance, contour or texture.

During the last decades many image quality metrics have been proposed in the literature. Initially the metrics only considered mathematical discrepancies with a so-called reference image.

\* Corresponding author.

E-mail address: [gloria.bueno@uclm.es](mailto:gloria.bueno@uclm.es) (G. Bueno).

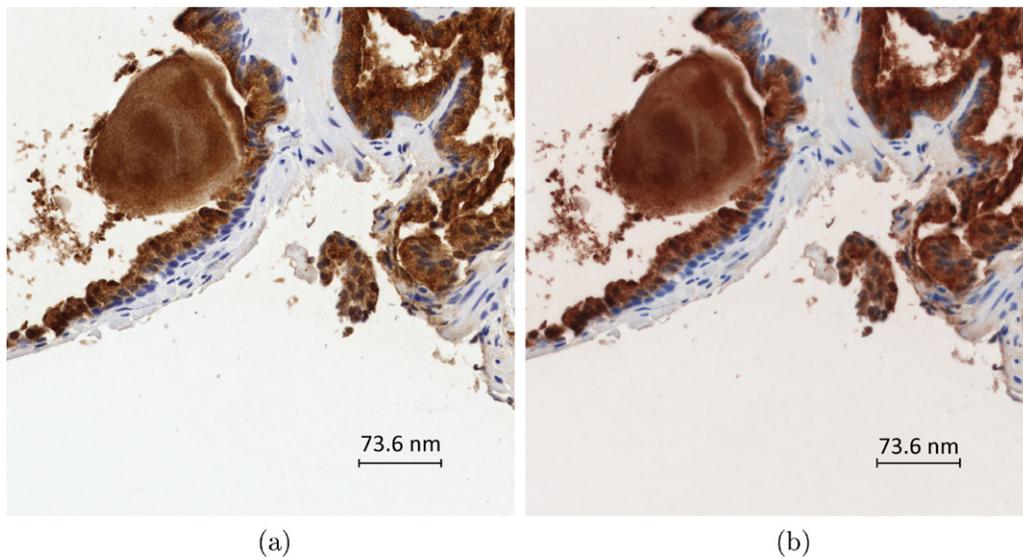


Fig. 1. Example of two captured histological images of size 1800 × 1800 pixels. (a) Scanner. (b) Microscope. Note differences could be masked by low-quality printing.

Among these metrics, also named objective metrics, one can find a simple sum of cumulative errors such as the Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) (González and Wintz, 1987) or more complex measures such as Structural SIMilarity (SSIM) (Wang and Bovik, 2002), which correlates information of luminance and contrast. These objective metrics are still an important tool for quality assessment of images, however they can suffer notable bias in terms of the overall final perceived image quality. For such reason many efforts have been devoted to developing perceptual metrics which involve an increasing number of stages of the complex Human Visual System (HVS).

It is beyond the scope of this paper to make a thorough description of the HVS, but the main ideas will be addressed here. The visual pathways are composed of multiple stages, from the retinal photoreceptors to the upper neuronal layers, which are highly non-linear (Wurtz and Kandle, 2000). At the first stages the neuronal response acts like a non-linear normalization of the dynamic range

of the luminance. Other upper cortical areas present spatial connections for neuronal excitation and inhibition mechanisms. These intrinsic non-linearities provoke that the final perceived distortion of a given feature between two images depends not only on the mathematical difference but also other contextual factors like local luminance, local contrast, texture information and other contextual information abstracted by the highest layers of the HVS. The state of the art in perceptual image metrics considers mainly the early stages, i.e. visual areas V1 and partially V2. This is due to its complexity on the one hand and on the other side due to the still unclear neuronal mechanisms of the highest stages which has to do with abstraction and memory manipulation of the visual information. However these perceptual metrics have shown more correlation with the overall perceived image quality than those purely statistical metrics.

Several authors have incorporated perceptual aspects to objective metrics in a bottom-top architecture, but only a few perceptual metrics have properly defined modules of the early HVS. The two most relevant proposals in this respect are the Visual Discrimination Model proposed by Lubin (1993) and the Visible Difference Predictor proposed by Daly (1993). In this study we consider Daly's contrast measure because it provided better results in previous studies by the authors Gallego (2006), Chalmers et al. (2000). It also operates in the spatial domain and visual discrepancies can be easily observed (Li et al., 2000). However, such a metric has been designed for reference image-problems and in this framework it is not available, therefore we will operate with its notion of contrast. In perceptual quality the contrast is the key concept because visual information is perceived as difference of luminance instead of absolute level. We will refer to Objective Perceptual Contrast (OPC) as the output of the contrast modules of this metric for measuring perceptual discrepancies of local contrast between microscopy and scanned pairs of images.

The 'no-reference' problem is challenging and frequent in real scenarios. Several no-reference metrics have been proposed in the literature, see Ferzli and Karam (2009) for a review. We evaluate here additionally a non-reference metric named Cumulative Probability of Blur Detection (CPBD). It was proposed by the previously cited authors using a top-bottom architecture which incorporates perceptual models of blurring/contrast adjusted to the evaluation of a training database through psychophysical experiments.

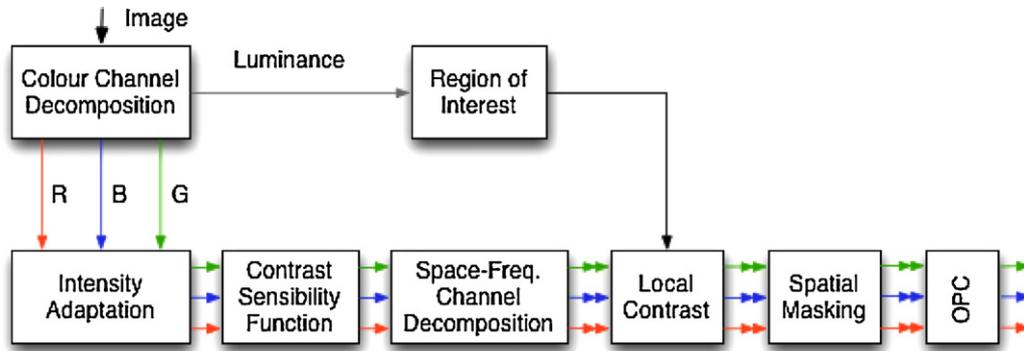
In spite of the utility of these objective metrics for evaluating image quality, none of these metrics, neither statistical nor

Table 1  
Main characteristics of whole slide scanning systems.

	Scanners	Microscopes
CCD RGB	Yes	Yes
CCD monochromatic	No	Yes
CMOS size <sup>a</sup>	3 × 2098 pixels	2048 × 2048 pixels
Sensor pixel size <sup>a</sup> (μm)	14 × 14	7.4 × 7.4
Resolution (μm/pixel)		
20× objective	0.47	0.37
40× objective	0.23	0.185
Slide feeder	up to 160 (BCR) <sup>+</sup>	up to 300 (BCR) <sup>+</sup>
2.5× objective lens	No	Yes
10× objective lens	No	Yes
20× objective lens	Yes	Yes
40× objective lens	Yes	Yes
60× objective lens	No	Yes
Bright field digitalization	Yes	Yes
Fluorescence field digitalization	No	Yes
Digitization speed 40× (mm <sup>2</sup> /s)	0.35–0.66	0.05–0.18
Digitization time at 40×		
Area 10 × 10 mm	4–60 min	11–35 min
Area 15 × 15 mm	9–150 min	25–80 min
Whole area 25 × 50 mm	40–720 min	107–448 min
JPEG compression method	Yes	Yes
JPEG2000 compression method	No	Yes
Compressed file size 40×	1.5 G–2 GB	1.5 G–2 GB

(BCR)<sup>+</sup> = includes a bar code reader.

<sup>a</sup> The devices tested in this paper.



**Fig. 2.** Block diagram of the modules of the Objective Perceptual Contrast (OPC). First, the image is decomposed in a luminance channel and three colour channels (red, blue and green), then an intensity adaptation, contrast sensibility function and space-frequency decomposition is performed where the local contrast is finally calculated. Background areas are discarded by using a region of interest mask calculated in the luminance channel. The local contrast is used for calculating the visibility masking thresholds for all bands and finally they are summed for obtaining the OPC value for each colour channel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

perceptual, is a universally accepted metric for image quality. Therefore we will make use of subjective techniques for assessing visual discrepancies. In particular psychophysical tests (ITU, 2000) provide a complete perceived quality which involves the overall HVS, from the early to the complex and later neuronal stages. However, psychophysical tests have their own drawback since they are also subject to scoring bias depending on the type of instructions given to the observers, the order of the evaluated images, luminance conditions or even the inherent visual adaptation to distortions. Another associated problem with these type of tests is the losing of spatial or temporal information where the discrepancies take place because distortion is scored globally.

Since each of the techniques presented above has its own useful properties but also its own limitations for the purpose at hand, we present here the evaluation of microscope/scanner images with the three techniques as necessarily complementary studies. A description of the materials used for this work can be found in Section 2. Section 3 presents the notion and experiments with OPC. Section 4 shows the results for CPBD and Section 5 presents psychophysical evaluation. Finally, conclusions are drawn in Section 6.

## 2. Material preparation

Tissue samples from biopsies and cytologies, prepared with different stains, were digitalized with a motorized microscope ALIAS II and with a scanner Aperio Scanscope XT at 40 $\times$  with no compression. The samples were extracted from 8 breast biopsies and 8 prostate biopsies, with the diagnosis of carcinoma. Specimens fixed in 4% buffered formalin were selected to prepare 4 mm thickness histological slides deparaffinized in xylene. Both conventional haematoxylin–eosin stain and immunohistochemical (IHC) techniques were performed. Immunohistochemical detection in 4 nm section of paraffin embedded prostate and breast biopsies was performed using monoclonal mouse anti-human Ki-67 antigen (clone MIB-1, DAKO, Denmark) in 12 breast slides, and polyclonal rabbit anti-human antibodies for Prostate-Specific Antigen, PSA (DAKO, Denmark) in 16 prostate slides. The immunocytochemical detection in cytology from pleural effusions was performed using monoclonal mouse anti-human calretinin (clone DAK-Calret 1, DAKO, Denmark) in 2 slides, and monoclonal mouse anti-thyroid transcription factor, TTF-1 (clone 8G7G3/1, DAKO, Denmark) in 6 slides. In all tissue cases, target retrieval was performed with a pre-treatment module for tissue specimens, PT Link (DAKO, Denmark). Ready to use primary antibodies were incubated for 1 h at room temperature, the detection was performed using the EnVision

FLEX+(DAKO, Denmark) visualization system in an Autostainer Link 48 (DAKO, Denmark).

The autofocus algorithm used in the microscope uses three iterations. The user provides three initial coordinates ( $Z_0, Z_{max}, Z_{min}$ ) and the step is set to  $\Delta Z = \Delta Z_{max}$ . The algorithm obtains a focus measure  $F_0$  at initial coordinate  $Z_0$ . Then, the z-position of the stage is incremented by  $\Delta Z$  stepwise until  $F_{pos} > F_{pos+\Delta Z}$  in two consecutive steps. The same procedure is done with the stage but decrementing from  $Z_0$ . This finally delimits a range where the best focused z-position named  $Z_1$  is saved. From that point a second iteration is performed with new initial coordinates ( $Z_1, Z_{1+\Delta Z_{max}}, Z_{1-\Delta Z_{min}}$ ) and a new step  $\Delta Z = \Delta Z_{max}/2$ . The last iteration narrows even more the search area, setting the initial coordinates to ( $Z_2, Z_{2+\Delta Z_{max}}, Z_{2-\Delta Z_{min}}$ ) and a more exigent step  $\Delta Z = \Delta Z_{max}/8$ . The image corresponding to the best in focus position in this last range is finally captured. The Aperio Scanscope XT autofocus algorithm uses a pre-focus calibration where the best-focus lens-height is measured at a fixed number of points on the slide. During scanning, the focus profile is calculated using interpolation based upon Delaunay triangulation.

The images finally captured with the microscope and the scanner were registered with an affine rigid registration to avoid misinterpretation due to possible rotation and translations produced in the digitalization process (Thévenaz et al., 1998). Then they were cropped in paired tiles of size 800  $\times$  800 to speed up the algorithm and to ensure the whole image is displayed entirely at full resolution in the monitor for psychophysical tests and a number of 34 random paired tiles were finally selected.

## 3. Quality evaluation through Objective Perceptual Contrast (OPC)

The perceptual contrast measurement used for assessing the quality of the images is made of several stages corresponding to the responses of the early stages of the HVS. Some models of the perceptual stages correlate better than others with psychophysical evaluation performed by real observers. Those modules which delivered the best correlation (Gallego, 2006), which mainly follows the implementation proposed by Daly (1993), were employed here. The modules, plotted in Fig. 2, are briefly described afterward.

### 3.1. Perceptual contrast modules

#### 3.1.1. Colour decomposition

In the same way as the perceptual metrics dedicated to measure luminance distortions have complications to imitate the HVS, there is not consensus between colour metrics. The majority of colour

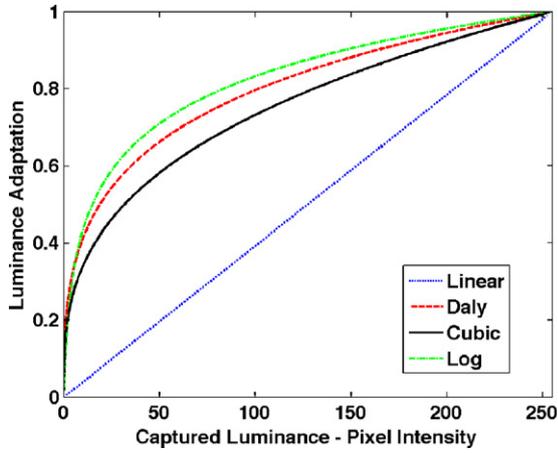


Fig. 3. Cubic root luminance adaptation compared to other often used non-linearities.

metrics propose a transformation of the RGB channels to a luminance channel and two decorrelated colour channels like S-CIELAB Zhang and Wandell (1998) or  $l\alpha\beta$  Toet and Lucassen (2003). In our case, due to the nature of the images at hand, which present clearly two main colours, brown (closely to red) and blue, we decided to assess the quality of YCbCr transformation. The resulting evaluation of the three channels could be joined in a single and global measure, however we preferred to keep them independent to evaluate different possible distortions in different luminance/colour channels.

3.1.2. Intensity adaptation

Intensity differences are dependent on local luminance. As the Weber–Fechner’s law describes, the more the local luminance is, the more contrast is required for perceiving the same intensity difference. This behaviour is typically modeled as a non-linearity of the luminance intensity. We used a cubic-root exponential law for that, which is similar to Daly’s model Daly (1993), see Fig. 3.

$$i_{\text{adapted}}(x, y) = i(x, y)^{1/3} \tag{1}$$

The input image  $i$  should be normalized, adapted and then multiplied by the maximum intensity level of the luminance displayed

by the monitors, which is often around  $250 \text{ cd/m}^2$ . The variables  $x$  and  $y$  are the cartesian image coordinates.

3.1.3. Contrast Sensitivity Function (CSF)

It has been already demonstrated in many vision research fields that the HVS has less sensitivity to very high and very low frequency signals (Wurtz and Kandle, 2000). The importance lays in the sense that a certain distortion in a signal of very high or very low frequency are less perceived than the same distortion in the mid-frequency range. The CSF is multiplied in the Fourier domain as follows:

$$i_{\text{CSF}}(x, y) = F^{-1}\{F\{i_{\text{adapted}}(x, y)\} \cdot \text{CSF}(u, v)\}, \tag{2}$$

where  $F$  refers to the Fourier operator and  $u$  and  $v$  the cartesian frequency coordinates. This can be seen as a frequency normalization or equalization according to visual sensitivity of the stimuli according to their frequency components. The higher the sensitivity in a certain range, the more visible the stimuli is in that frequency range. Another special feature of the HVS is anisotropy, which means that horizontal and vertical stimuli are more visible than oblique stimuli for a same given intensity. This is reflected in the 2D Fourier shape of the CSF. We finally adopted again the CSF proposed by Daly (1993) modelled according to their psychophysical experiments (see Fig. 4).

3.1.4. Space-frequency decomposition

The HVS in the area V1 is made of multiple neurons oriented to detect explicit oriented and scaled features. This is often modeled as a decomposition of the Fourier domain by means of filters tuned in a certain frequency range. We used Gabor filters because they present high fidelity with the neuronal V1 cells response (Daugman, 1980). Let  $G_{s_o}$  be the filter of the scale  $s$  and the orientation  $o$  defined in the polar Fourier domain  $(\rho, \theta)$  as:

$$G_{s_o} = \exp\left(-\frac{1}{2}\left(\frac{\rho - \rho_o}{\sigma_\rho}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{\theta - \theta_{s_o}}{\sigma_\theta}\right)^2\right), \tag{3}$$

where  $\rho_o$  and  $\theta_{s_o}$  are the polar coordinates of the filter centers and  $(\sigma_\rho, \sigma_\theta)$  their variance. The tuning parameters are not in the scope of this paper, but can be found in a publication of the same authors (Fischer et al., 2007). We used for this study a decomposition of 5

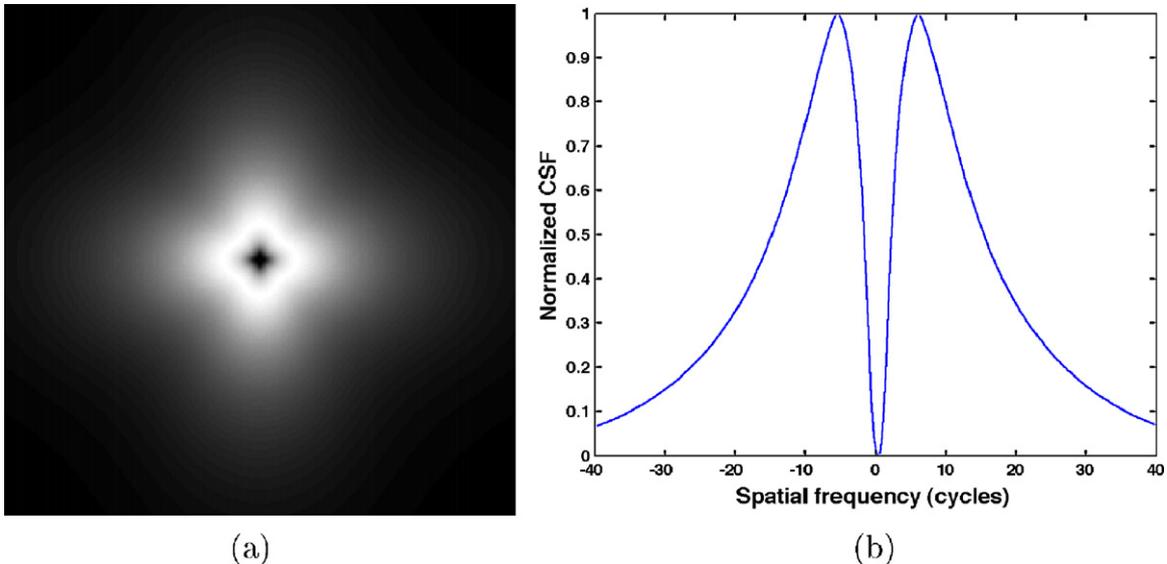
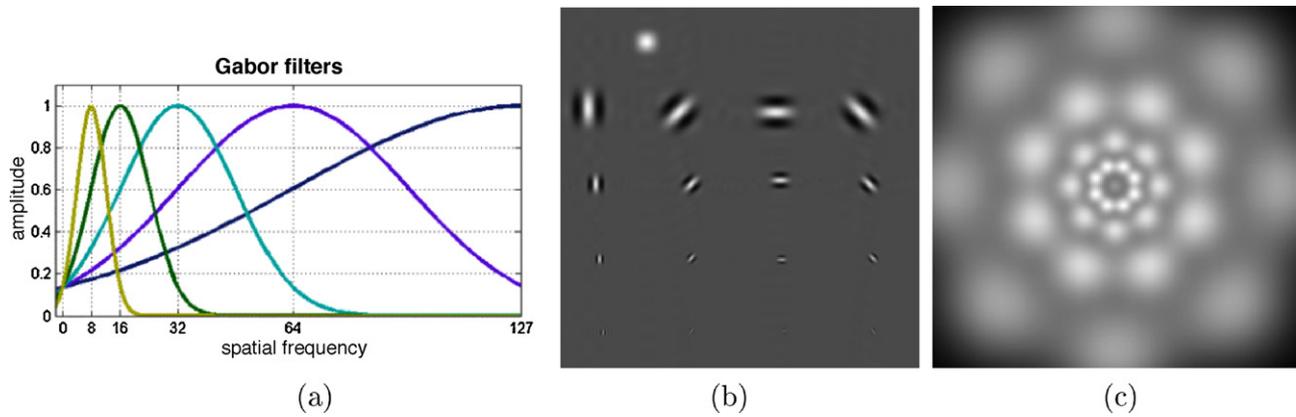


Fig. 4. Normalized contrast sensitivity function proposed by Daly for default values of resolution 32 pixel/cm and viewing distance 40 cm. (a) 2D view of the Fourier domain where the frequency decay and the anisotropy are the main features. (b) Central profile of the 2D view.



**Fig. 5.** Gabor filters used for space-frequency decomposition. (a) Profiles of the filters across scales. (b) Filters in space domain: 4 orientations, 4 scales and low-pass filter. (c) Same 4 scales and orientations in the Fourier domain.

scales and 4 orientations. A Fourier covering of such Gabor filters can be seen in Fig. 5.

### 3.1.5. Local contrast

Once the image is decomposed into normalized cortical channels, each one is transformed into the space domain which gives the contrast information. The final contrast is calculated dividing each channel in the space domain by the local luminance provided by the low-pass channel as follows:

$$c_{so}(x, y) = \frac{i_{so}(x, y)}{i_0(x, y)}, \quad (4)$$

where  $c_{so}(x, y)$  represents the local contrast for a given channel,  $i_{so}(x, y) = i_{CSF}(x, y) \times G_{so}$  refers to the response of the Gabor filter and  $i_0(x, y) = i_{CSF}(x, y) \times G_0$  is the low-pass response. Other contrast calculation like global contrast or Lubin's contrast (Lubin, 1993) have been proposed, though we finally adopted the local contrast following the results of Gallego (2006).

### 3.1.6. Spatial masking

One of the last characteristics of the early stages of the HVS has to do with the observed behaviour between neurons for exciting and inhibiting their surrounding cells when a pronounced stimuli pass through. This can be modeled as a masking effect where a given contrast difference is hardly perceived in regions with high local contrast, like object contours or textures, and more easily perceived in regions where signals of low-frequency predominate. From

psychophysical experiments (Daly, 1993), the visibility threshold can be modeled following the Minkowski metric as:

$$t_{so}(x, y) = [1 + [K_1 [K_2 |c_{so}(x, y)|]^\alpha]^\beta]^{1/\beta}, \quad (5)$$

where  $K_1 = 0.0153$ ,  $K_2 = 392.4980$ ,  $\alpha = 1$  (for band-pass channels) and  $\beta = 4$ . Note that ideally the threshold varies from 1 (visibility is unchanged) to  $+\infty$  (no distortion can be perceived).

### 3.1.7. Objective Perceptual Contrast (OPC)

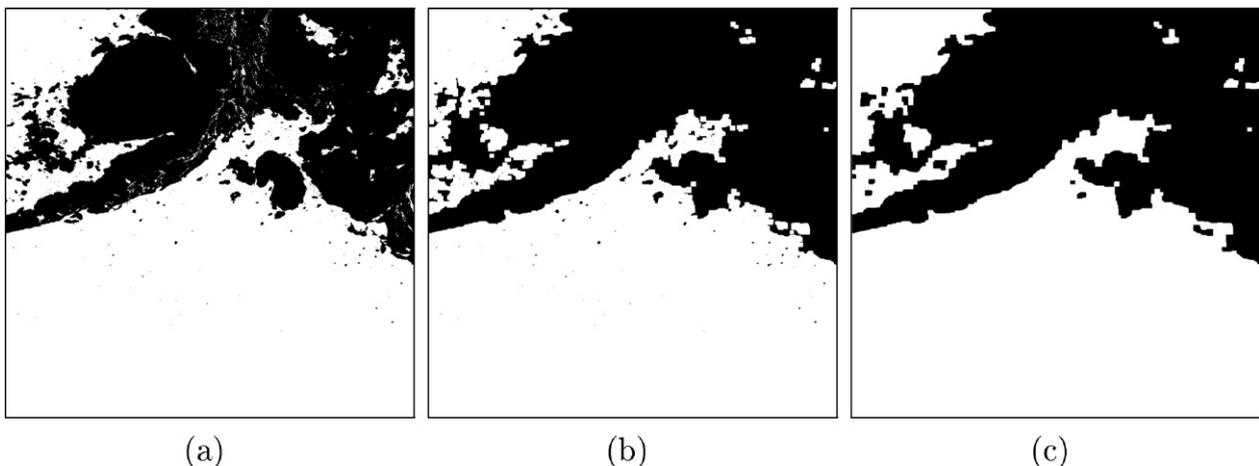
We named Objective Perceptual Contrast to the masked local contrast which has been attenuated by the visibility threshold:

$$p_{so}(x, y) = \frac{c_{so}(x, y)}{t_{so}(x, y)}. \quad (6)$$

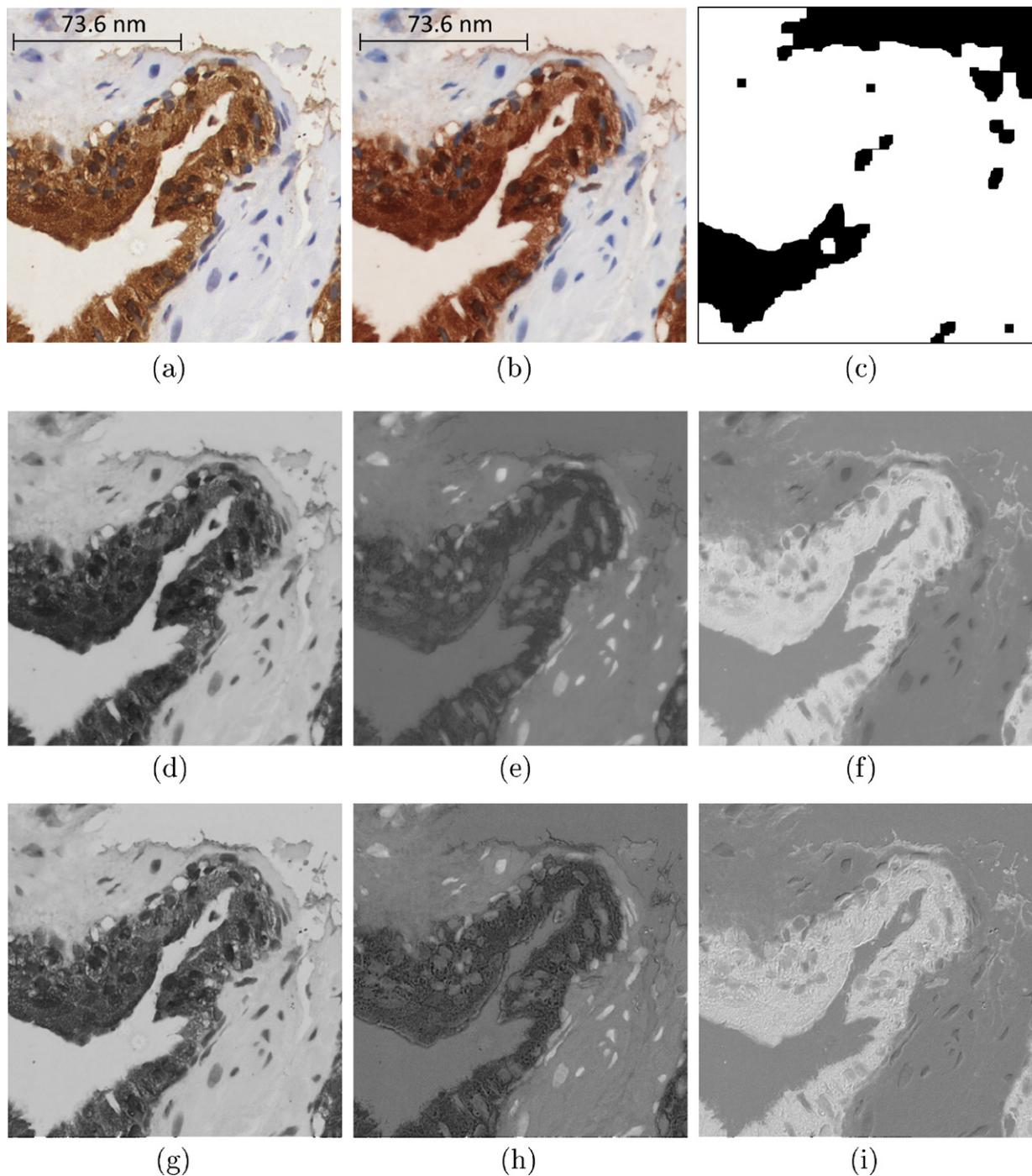
This equation gives the modeled intensity of contrast which will be finally perceived in a local region (with a given scale and orientation). Because every pair of images that we want to compare corresponds to the same view, the sum of their local contrast across scales, orientations and spatial coordinates provides an overall energy measurement of perceived contrast and therefore image quality.

$$OPC = \sum_{x, y, s, o} p_{so}(x, y). \quad (7)$$

In this respect a contrast measure used for evaluating image quality was also proposed in Wang et al. (2004). Although such a measurement reveals which image of the same view has better



**Fig. 6.** Example of Region of Interest extraction from images in Fig. 1. (a) Threshold binarization. (b) Closing operation for filling small holes. (c) Opening operation for removing small particles.



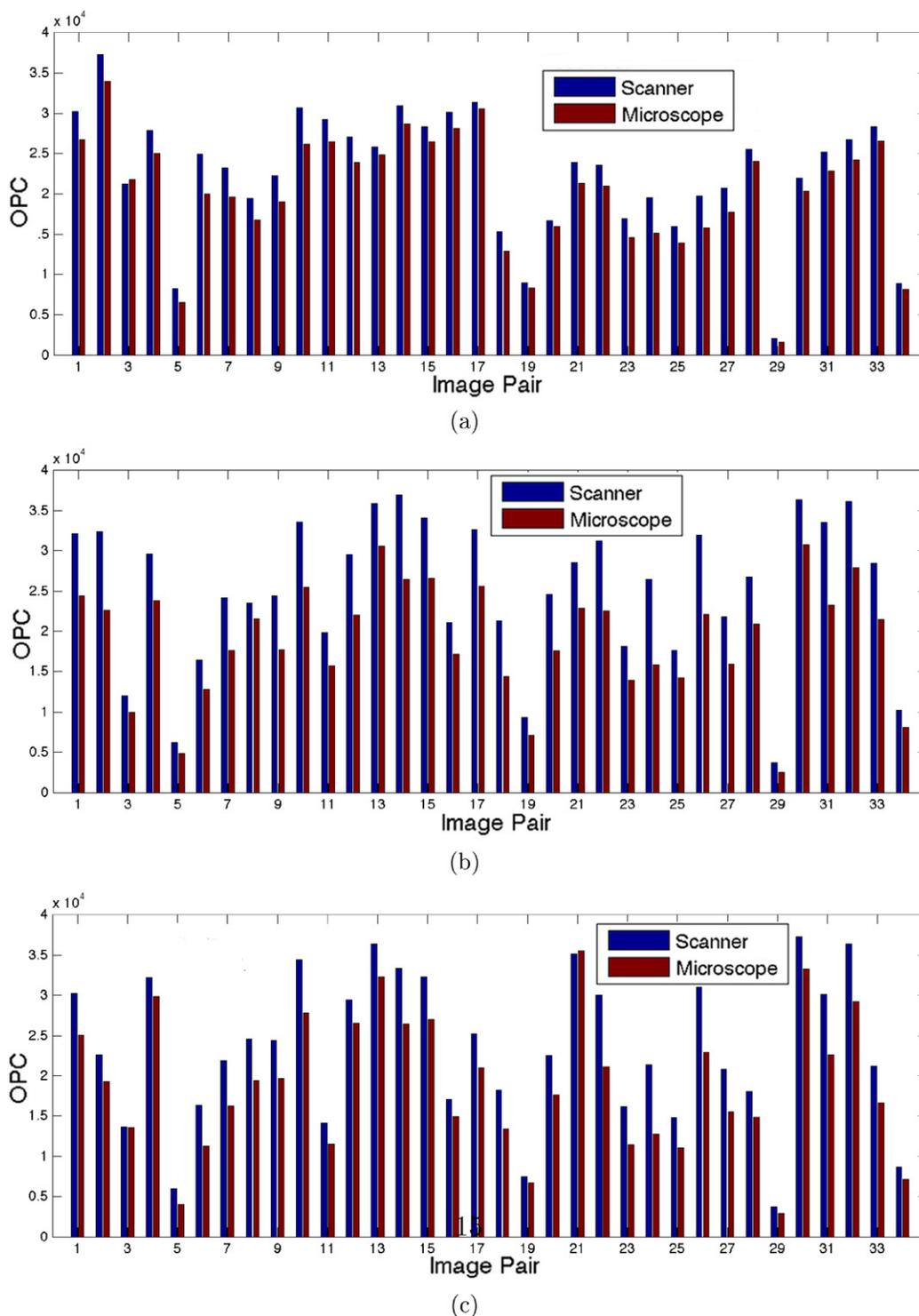
**Fig. 7.** Example of assessment of the 18th pair of (a) scanner; (b) microscopy images; (c) ROI; (d)–(f) luminance – Y, blue – Cb and red – Cr channel decomposition of the scanned image, respectively; (g)–(i) channel decomposition of the microscopy image.

quality in terms of contrast, sharpness or contour preservation, it is not able to provide an absolute significance of quality between two images.

### 3.1.8. Region of Interest (ROI)

Some of the images evaluated present large background regions which are not relevant for diagnosis. Therefore they should not interfere with the perceptual quality. For estimating the ROI where the metric should be computed, we adopted a simple and effective method. Firstly, the image is binarized by taking the peak value of the histogram as a threshold. This value corresponds closely to the grey level of the background since it is the most predominant

value. Then two consecutive closing and opening morphological operations are computed with a mask of size  $21 \times 21$  (González and Wintz, 1987). The closing operation removes holes in the tissue regions and the opening operation removes small particles in background regions. The size of the mask was adjusted empirically depending on the original size of the captured images. Actually, this method makes use of well-known strategies for such a particle screening task. Although other strategies could be used, we finally adopted this method because it fulfilled our requirements. Finally, the ROI is calculated for the luminance channel of the microscopy or scanned images, the result was similar, and applied to both colour channels (red and blue). See an example in Fig. 6.



**Fig. 8.** Objective Perceptual Contrast (OPC) evaluated for 34 pairs of microscopy and scanned images of the (a) luminance – Y, (b) blue – Cb and (c) red – Cr colour channels.

### 3.2. OPC results

Fig. 8 shows an example of a pair of images captured by a scanner and a microscope. Fig. 7 presents the OPC output plots for the 34 examined pairs. The measurement of perceptual contrast reveals an evident superiority of the scanned images over the microscopy images (see Fig. 7). The number of scanner images with more contrast than the microscope images was 94%, 100% and 97% for Y, Cb and Cr, respectively. The ratio between the scanner OPC divided by the microscope OPC, averaged for all the pairs, was 1.13, 1.34 and

1.30 for Y, Cb and Cr, respectively, which means that the scanner provides more contrast in general. Moreover, although there is not any colour channel predominance, the luminance channel present slightly less impairment compared to the colour channels. In any case, because the three colour channels present similar ratios one can argue that defocus is responsible for the main impairment cause. Although this impairments are measured quantitatively in terms of perceptual contrast, the measurement does not reflect the real significance in terms of visual discrepancies perceived by an observer. To that end we will use the following quality metrics.

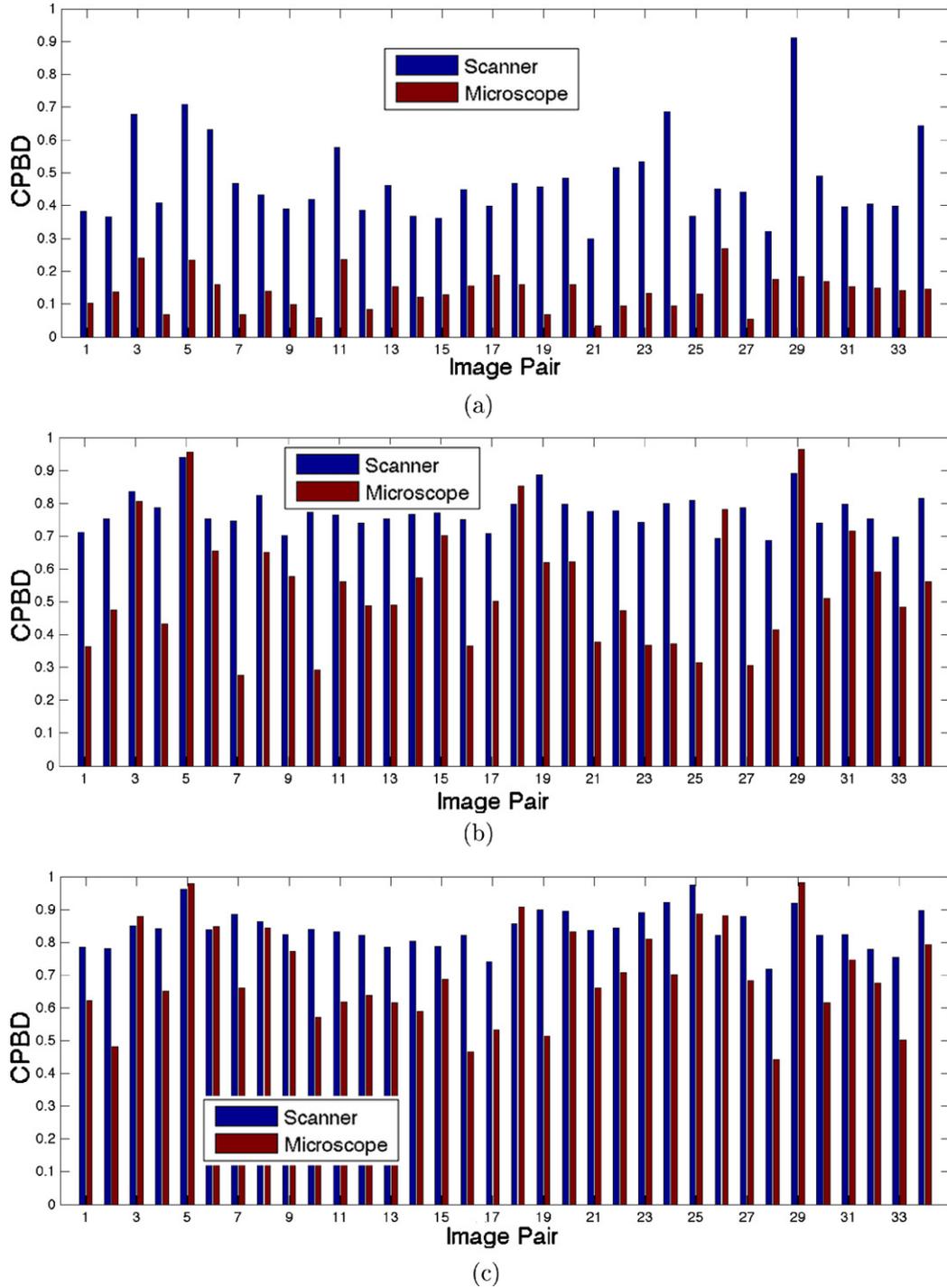


Fig. 9. Cumulative Perceptual Blurring Detection (CPBD) evaluated for 34 pairs of microscopy and scanned images of the (a) luminance – Y, (b) blue – Cb and (c) red – Cr channels.

**4. Quality evaluation through Cumulative Perceptual Blurring Detection (CPBD)**

The CPBD quality metric measures the contrast of edge contours to estimate the degree of blurriness present in images. First, the image is split in  $64 \times 64$  blocks which are evaluated only if they have a significant number of edge pixels in it. Similarly to the notion of luminance adaptation introduced before in Section 3.1.2, the edge contrast is calculated following the psychophysical concept of “Just Noticeable Blur” (Ferzli and Karam, 2009) which models the

contrast Sensitivity against a given background luminance. Such JNB works as a visibility threshold which is also related to the notion of masking defined previously in Section 3.1.6. Then the model estimates the probability of detecting blur in a certain area as:

$$P(i) = 1 - \exp\left(-\left|\frac{w(i)}{w_{JNB}(i)}\right|^\beta\right), \tag{8}$$

where  $i$  means the index of the edge,  $w(i)$  the width and  $w_{JNB}(i)$  the visibility threshold. The probabilities are finally grouped together in a single value for scoring the whole image as:

$$P_{blur} = 1 - \prod_i (1 - P(i)). \quad (9)$$

#### 4.1. CPBD results

In Fig. 9 the CPBD output is plotted for the 34 examined pairs. As it was explained in Section 3.1.1 the colour image is decomposed similarly to YCbCr colour channels. The quality metric reveals again an evident superiority of the scanned images over the microscopy ones. Now the number of scanner images with more contrast than the microscope images was 100%, 88% and 82% for Y, Cb and Cr, respectively, and the ratios between scanner and microscope are 4.05, 1.57 and 1.24 for Y, Cb and Cr, respectively. On the other hand, the luminance channel of the microscope images is now more penalized by the metric than what we found in OPC. Nonetheless, this reinforces the idea that defocus is still the main responsible for impairment. Although the slight discrepancy between both colour channels, Cb and Cr, evidences some colour alterations due to the image capture process. Such clues given by objective evaluation will be also corroborated afterwards through subjective experiments.

### 5. Quality evaluation through subjective psychophysical tests

We completed the quality assessment with subjective quality metrics because they can give a complementary analysis of the global perceived image quality. In this study 6 experts on the diagnosis of this type of tissues from the Hospital Universitario General de Ciudad Real (Spain) were subject to psychophysical tests. For such purpose we followed the recommendation ITU-R (ITU, 2000) where the main guidelines are: (1) the experts are exposed to an automated test consisting of a random series of the same 34 pairs of images introduced in the previous section. (2) The pairs are displayed twice and sequentially, ABAB (they are never displayed at the same time), in an interval of 10 s each image with a gap of 2 s with a grey level luminance between each one. (3) At the end of each pair displaying there exists a final period of time of 10 s dedicated to scoring from 0 to 5, in correspondence with PQJ, representing how much similar are the observed pair of images. We called this measure the Mean Opinion Score (MOS) in accordance with ITU-R. (4) The order of A and B pair is also random for avoiding a possible bias. This means that A sometimes corresponds to the scanned image and sometimes corresponds to microscopy image. (5) The first case is taken as an example, where appropriate instructions are given, and the second case is avoided in order to stabilize the scoring.

We consider that a preference among the pair of images exists if at least 4 experts consider the same image as better, and only their scores are taken into account. Consequently if only three experts prefer one image and the other three prefer the other one, then no single image is considered better than the other. For 66% of the cases the experts considered that the scanned images had better quality and scored a MOS difference of 2.75 averaged value. For 14% of the cases the microscope images were preferred with a 2.57 MOS score, and the remaining 20% percentage both types of images were equally preferred. This means that the experts consider that generally the scanner provides an overall better image quality. Furthermore in the cases where microscope provided better images, the scanned images were a bit closer than the opposite cases.

The experts were also asked for the main causes of discrepancies between focusing, contrast, stain, colour, brightness, overexposure and sharpness. When the scanner was preferred, the main cause

addressed was focusing with 76% of the cases and colour discrepancies for the remaining 24%. Stain was also addressed as a third cause for 22% of the cases. When microscope was preferred then focusing was the main cause for 100% of the cases, colour and stain the second cause and with contrast and sharpness the third cause. This results places defocus as the main addressed cause of impairment of the image quality.

Note that no correlation between the objective metric like CPBD and psychophysical scoring like MOS was calculated since the data set does not present a wide range of degradations to accommodate a logistic function. In any case, both assessments released conclusive results on their own.

### 6. Conclusions

The aim of this paper was to carry out a quantitative quality assessment of biopsy and cytology images provided by two different devices for clinical diagnosis used nowadays: microscope and scanner. To that end we decided to apply three different strategies. First, an Objective Perceptual Contrast measure was used for assessing the magnitude of the possible impairments, which revealed blurring as the main problem with no significant colour distortions in microscopy images. Second, a non-reference quality image based on blur estimation of edge areas from a top-down architecture corroborated the previous results. Finally, psychophysical tests were performed with expert observers, which confirmed again from a global perception point of view that in general the scanner provides slightly better image quality, addressing focusing as the main problem and colour distortions as the second. However, for some of the evaluated cases, the microscope was able to provide better in-focus images, which encourages us to develop new strategies to optimize and automate the focusing process as the next research line.

Since motorized microscopes are more flexible and customizable than scanners, it is possible to improve focusing and contrast, and therefore image quality. Moreover, motorized microscopes are more suitable for applications where other objective lenses different than 20× and 40× are required; as well as for fluorescence field applications. In summary, in this paper it has been shown that the image quality provided by scanners is slightly better than in motorized microscopes (when digitizing at 40× in bright field). In any case, the image quality of both devices is suitable for clinical, educational and research purposes.

### Acknowledgements

This work has been carried out with the support of COST Action IC0604, and the research projects DPI2008-06071, TEC2010-20307, TEC2010-09834-E of the Spanish Research Ministry, PI-2010/040 of the FISCAM and PAI08-0283-9663 of JCCM.

### References

- Chalmers, A., Daly, S., McNamara, A., Myszkowski, K., Troscianko, T., 2000. Image quality metrics. In: SIGGRAPH Tutorials.
- Daly, S., 1993. The visible difference predictor: an algorithm for the assessment of image fidelity. In: Watson, A.B. (Ed.), *Digital Images and Human Vision*. MIT Press, Cambridge, MA, pp. 179–206, Ch. 5.
- Daugman, J.G., 1980. Two dimensional spectral analysis of cortical receptive fields profiles. *Vision Research* 20, 847–856.
- Ferzli, R., Karam, L., 2009. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE Transactions on Image Processing* 18 (4), 717–728.
- Fischer, S., Sroubek, F., Perrinet, L., Redondo, R., Cristóbal, G., 2007. Self invertible gabor wavelets. *International Journal of Computer Vision* 75, 231–246.
- Gallego, A.R., 2006. Modelos visuales en el análisis de la calidad de imagen. Master's Thesis. ETS Ing. de Telecomunicación, Universidad Politécnica, Madrid.

- García-Rojo, M., Bueno, G., Peces, C., González, J., Carbajo, M., 2006. Critical comparison of 31 commercially available digital slide systems in pathology. *International Journal of Surgical Pathology* 14 (4), 285–305.
- González, R.C., Wintz, P., 1987. *Digital Image Processing*. Ed. Addison Wesley.
- ITU, 2000. Methodology for the subjective assessment of the quality of television pictures. Recommendation ITU-R BT.500-10.
- Li, B., Meyer, G.W., Klassen, R.V., 2000. A Comparison of Two Image Quality Models. Computer Science Department, University of Oregon, Eugene, OR 97403 Color and Digital Imaging Systems, Xerox Corporation, Webster NY 14580.
- Lubin, J., 1993. The use of psychophysical data and models in the analysis of display performance. In: Watson, A.B. (Ed.), *Digital Images and Human Vision*. MIT Press, Cambridge, MA, pp. 163–178, Ch. 5.
- Thévenaz, P., Ruttimann, U., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing* 7 (1), 27–41.
- Toet, A., Lucassen, M.P., 2003. A new universal colour image fidelity metric. *Displays* 24, 197–207.
- Wang, Z., Bovik, A.C., 2002. A universal image quality index. *IEEE Signal Processing Letters*, 81–84.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13 (4), 600–612.
- Wurtz, R.H., Kandel, E., 2000. Central visual pathways. In: Kandel, E., Schwartz, T.H., Jessell, T.M. (Eds.), *Principles of Neural Science*. Appleton & Lange, New York, pp. 523–547.
- Zhang, X., Wandell, B.A., 1998. Color image fidelity metrics evaluated using image distortion maps. *Signal Processing* 70, 201–214.