3D SIMILARITY INDEX FOR EVALUATING QUALITY OF LOSSY COMPRESSED SPECTRAL IMAGES

Dobromir Dochev, Vladimir Bochko, Diana Kalenova, Pekka Toivanen*, Arto Kaarna; Lappeenranta University of Technology, Department of Information Technology P.O. BOX 20, FI-53851 Lappeenranta, Finland; * Tampere University of Technology, Digital Media Institute/EPANET, Frami, Kampusranta 9 C, FI-60320 Seinäjoki, Finland

Abstract

This study describes the development of a similarity index for comparing the quality of a lossy compressed spectral image. The index is based on the structural errors that are computed between the original spectral image and the spectral image compressed in a lossy manner. The index is an average of the local indices obtained from using a three-dimensional sliding window over the images. We compare the performance of the new index to the two-dimensional index. The main advantage of the new index is that it can detect both structural changes from spatial domain and color changes from spectral domain. It has several parameters and they can be used to adjust the sensitivity of the index.

Introduction

Multispectral images are available nowadays for different purposes. Their applications increase due to the development of in the spectral imaging systems. Usually geoscience, remote sensing and quality control systems have been the main source of multispectral images, since RGB space does not provide the sufficient color information needed for their proper exploitation. The raw format of these images requires large amounts of storage space, and therefore a lot of efforts are focused to achieve optimal compression techniques for them [1].

Even though the capacities of storage media and communication channels grow the research concentrates on providing better and better quality. Digital images are subject to a wide variety of distortions during acquisition, compression, transmission, processing and reproduction. These may result in degradation of visual quality. If the observer of images is a human being, the only trustworthy method of estimating the image quality is through visual assessment. However, in practice subjective quantification is mostly inconvenient, time-taking and expensive. Finding an objective image quality assessment that can automatically estimate image quality has been the goal of many researchers [1-7].

There are a number of qualitative measures for gray scale images and RGB color images, but for multispectral images such measures are rare. Such an objective quality metric can play a variety of roles in image processing systems. First application is to dynamically monitor and adjust image quality during transmission of visual data in dependence of the conditions that a communication channel provides. Second implementation is to optimize algorithms and parameters of different compression techniques. Third, it can be used to benchmark image processing systems or algorithms, helping in decision making about the usefulness of compression models.

Image quality metrics can be classified according to the presence of a reference image, with which the observed image is to be compared. In first place the reference image is known and the metrics are working comparing both images. This is used in most approaches and they are known as full-reference. In many practical cases, however, the original picture is not known. This situation is known as no-reference, or "blind" quality assessment approach. In the third type of methods the reference picture as partially available in the form of extracted features. They are available as side information to help estimate the quality of the distorted picture. This case is popular as reduced-reference quality assessment. This study focuses on full-reference image quality approach.

It is a widely adopted assumption, that an image whose quality is being evaluated can be presented as a sum of an undistorted reference signal and an error signal. Every compression algorithm leads to certain types of errors – color shifts, blurring, blocking, and noise. The loss of perceptual quality is usually described as the level of visibility of these errors. The simplest and most widely used full-reference metric implementing this concept are the mean squared error (MSE), computed by averaging the squared intensity differences of original and distorted image pixels, followed by the related quantity of peak signal-to-noise ratio (PSNR), which objectively quantifies the strength of the error signal. They are simple to calculate and have clear physical meaning. But two distorted images with very different types of errors may have the same MSE and it is well known that some errors are much more visible or irritating than others. Most assessment approaches based on perceptual image quality attempt to weight different aspects of these errors according to their visibility. They rely on determined human visual system (HVS) models stated on psychophysical measurements of humans and physiological measurements of animals [4].

Most perceptual quality assessment models include similar stages, although they may differ in details – pre-processing, CSF filtering, channel decomposition, error normalization and error poling. In the first stage (pre-processing) a variety of basic operations are performed to eliminate known distortions from the images compared. These include scaling and aligning of original and distorted signals; transformation from one color space to another if needed; transformation of digital pixel-values of the image to luminance values of pixels on the display device; low-pass filtering, simulating the point spread function of the eye optics; some modification simulating light adaptation by the eye. The second stage (CSF Filtering) implements weighting the signal to the contrast sensitivity function (CSF) that describes the sensitivity of HVS to different spatial and temporal frequencies. In the third stage (channel decomposition) images are usually divided into subbands that are sensitive to spatial and temporal frequency and orientation. This is believed to be related to the neural response in the primary visual cortex. During the forth stage (error normalization) the difference between the reference and distorted signals in each channel is obtained and normalized according to some masking model. These try to implement the decrease of visibility of one kind of image components in dependence of the visibility of other components presented using certain visibility thresholds. The final stage (error pooling) is combining the normalized error signals from different channels over the spatial dimension of the image.

Although, this approach to the problem has found universal acceptance, it is known that there are limitations, mainly coming from the complexity and high nonlinearity of HVS, leading to a significant number of assumptions and generalizations. The most fundamental problem is the definition of image quality. In particular it is questionable whether error visibility is equal to loss of quality, as some distortions may lead even to better quality. The second is the suprathreshold problem, connected with the levels at which a stimulus is just barely visible. Very few psychophysical studies have satisfyingly proved that such near-threshold models can be generalized to characterize perceptual distortions. The third problem is known as natural image complexity. It is based on the fact, that most psychophysical experiments are conducted using relatively simple patterns. These are much simpler than real world images, which can be described as a superposition of large number of simple patterns. Next there is the decorrelation problem, which comes from the assumption that errors at different locations in an image are statistically independent when using a metric for spatial errors. This would be true if processing prior to pooling eliminated dependencies between the input signals. It has been shown that for natural images after channel decomposition the intra- and inter- channel coefficients are highly dependent. And at last this is the cognitive interaction problem. It is widely known that interactive visual processing influences the perceptual quality of images. Prior information regarding the image content, different instructions and attention or fixation also may affect the evaluation of the image quality. These metrics are not well understood and are difficult to quantify.

SSIM Index

The signals of natural images are highly structured, their pixels exhibit strong dependencies and these carry important information about the structure of the objects in the visual scene. In [2] a new approach is proposed, the motivation of which is to find a more direct way to compare the structures of the reference and distorted signals. It is based on the assumption that the human visual system is specialized to extract structural information from the scenes observed. Therefore a measure of structural information change can provide approximation to perceived image distortions. The new philosophy in [4] has three main distinguishing features. First, it considers image degradation as perceived changes in structural information variation. Second, unlike the error-sensitivity approach, this new paradigm is a top-down approach in correspondence with the hypothesized functionality of the overall HVS. This avoids the suprathreshold problem and the cognitive interaction problem is also reduced. Third, it proposes to evaluate the structural changes between two signals directly. This way, the problems of natural image complexity and decorrelation is avoided.

The luminance of an object in the picture being observed is the product of the illumination and the reflectance. The structure of the objects in the scene should be independent of the illumination; consequently to obtain the structural information in an image the influence of the illumination must be separated. Since luminance and contrast vary across a scene the local brightness and contrast are used in definitions.

First the luminance of each signal is compared. The mean intensity for signal **x** is:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

The luminance comparison function is then $l(\mathbf{x},\mathbf{y})$ and it is a function of mean intensities μ_x and μ_y.

Second, the mean intensity must be removed from the signal, resulting in the signal $\mathbf{x}-\mu_x$. The standard deviation is used as an estimate of the contrast of the signal:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \mu_x\right)^2} \qquad (2)$$

The contrast comparison function $c(\mathbf{x},\mathbf{y})$ is then based on the comparison of σ_x and σ_y.

Third, the signal is normalized by its own standard deviation, so that the two signals to be compared have unit standard deviation. The structure comparison $s(\mathbf{x},\mathbf{y})$ is conducted on these normalized signals $(\mathbf{x}-\mu_x)/\sigma_x$ and $(\mathbf{y}-\mu_y)/\sigma_y$.

Finally, the three components are combined to obtain an overall similarity measure:

$$S(\mathbf{x},\mathbf{y}) = f\,(l(\mathbf{x},\mathbf{y}),\, c(\mathbf{x},\mathbf{y}),\, s(\mathbf{x},\mathbf{y})) \qquad (3)$$

An important point is that the three components are relatively independent. It is needed that these functions satisfy some conditions:

1. Symmetry: $S(\mathbf{x},\mathbf{y}) = S(\mathbf{y},\mathbf{x})$.
2. Boundedness: $S(\mathbf{x},\mathbf{y}) \leq 1$.
3. Unique maximum: $S(\mathbf{x},\mathbf{y}) = 1$ only and if only $\mathbf{x} = \mathbf{y}$.

Luminance comparison in [2] is defined:

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \qquad (4)$$

where the constant C_1 is included to avoid instability when $\mu_x^2 + \mu_y^2$ is close to zero. Typically $C_1 << 1$ and it is seen that equation (4) obeys the three properties above. This expression is qualitatively consistent with Weber's law – the HVS is sensitive to the relative luminance change, not the absolute.

For the contrast comparison [2] a similar function is defined:

$$c(\mathbf{x},\mathbf{y}) = \frac{2c_x c_y - C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \qquad (5)$$

where C_2 is a similar small constant like C_1. This definition again satisfies the three conditions listed above. Also it is consistent with the contrast-masking feature of the HVS, because it is less sensitive to contrast change in the case of high base contrast than in the case of the low.

Structure comparison is conducted after luminance subtraction and variance normalization. It has been proved that the correlation between $(\mathbf{x}-\mu_x)/\sigma_x$ and $(\mathbf{y}-\mu_y)/\sigma_y$ is equal to the correlation coefficient between **x** and **y**:

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \qquad (6)$$

where C_3 is a small constant presented to avoid instability when denominator is close to zero. Geometrically the result is given as the cosine of the angle between the vectors $\mathbf{x}-\mu_x$ and $\mathbf{y}-\mu_y$.

Finally, these three comparisons are combined to obtain the resulting similarity measure, called in [4] SSIM index:

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha}\,[c(\mathbf{x},\mathbf{y})]^{\beta}\,[s(\mathbf{x},\mathbf{y})]^{\gamma} \qquad (7)$$

where $\alpha > 0$, $\beta > 0$, $\gamma > 0$ are parameters used to adjust the relative importance of the three components. This definition satisfies the conditions stated above.

Multispectral Image Assessment Using SSIM Index

In image quality assessment it is known to be useful to apply the SSIM index locally rather than globally, because of several reasons: the statistical features if images are spatially nonstationary; image distortions may also be space-variant and varying in dependence of image statistics; because of HVS features, only a local area in the image can be perceived with high resolution; localized measures can bring to more informative spatially varying quality map of the image.

A spectral image consists of many grayscale images of the same scene taken at different wavelengths. Attention should be paid to the difference in spatial and spectral dimensions, since they provide totally different information contents. In this study two ways of image quality assessment using SSIM are implemented. In the first case we apply the original SSIM algorithm to every band of the image consequently and average the result. In the second case we extend the SSIM algorithm to three dimensional SSIM index and apply it. The common approach for both cases is the use of Gaussian weighting windows function $\mathbf{w} = \{\omega_i | i = 1,2,\ldots,N\}$, normalized to unit sum ($\Sigma\omega_i = 1$) for estimating the local SSIM. The aim of this is to escape the undesirable "blocking" artifacts which appear when calculating SSIM within local window. The estimates of local statistics are then modified accordingly as:

$$\mu_x = \sum_{i=1}^{N} \omega_i x_i \tag{8}$$

$$\sigma_i = \sqrt{\sum_{i=1}^{N} \omega_i \left(x_i - \mu_x\right)^2} \tag{9}$$

$$\sigma_{xy} = \sum_{i=1}^{N} \omega_i \left(x_i - \mu_x\right)\left(y_i - \mu_y\right) \tag{10}$$

For the experiments with two-dimensional SSIM we used 11 x 11 circular-symmetric Gaussian weighting function. The coefficients in the comparison functions are determined as

$$C_1 = \left(K_1 L\right)^2,\ \ C_2 = \left(K_2 L\right)^2 \tag{11}$$

where L is the dynamic range of pixel values (255 for 8-bit grayscale 'images) and K_1, K_2 << 1 are small constants. Assuming that $C_3 = C_2/2$ and $\alpha = \beta = \gamma = 1$ the result in the specific form for the SSIM is:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{12}$$

For the second case we have to define three-dimensional Gaussian weighting function. According to [12] it is given by

$$h(x, y, z) = \sqrt{2\pi}\sigma A e^{-2\pi^2\sigma^2\left(x^2+y^2+z^2\right)} \tag{13}$$

This way a sliding cube of size 11 x 11 x 11 is used to process the spectral image and the three components of the quality measure are computed for each pixel: the brightness, contrast and structure. The small constants are chosen so to satisfy Eq. 11, but we find that in our current experiments, the performance of the SSIM index algorithm is fairly sensitive of these values, so some optimization there is needed. For the calculation of local SSIM indices we use the general form of the expression Eq. 7.

In practice, a single overall quality measure of the entire image is required. Thus, a mean SSIM (MSSIM) index can be used to evaluate the overall image quality defined as

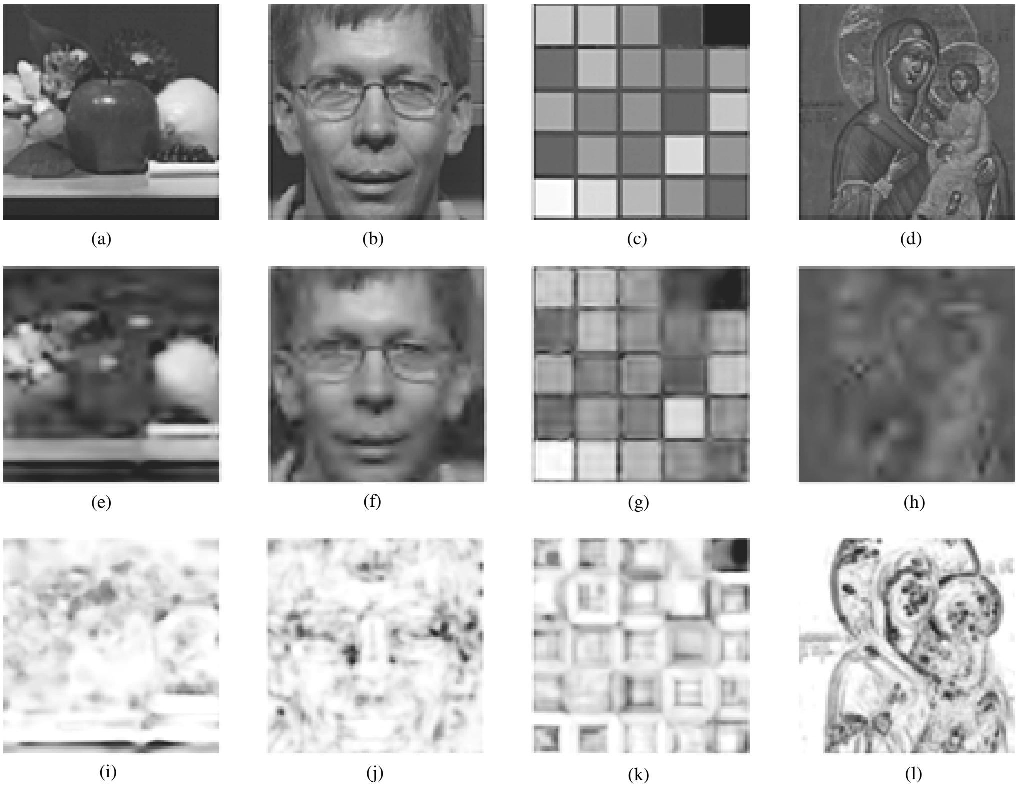

Figure.1. *Sample PCA + SPIHT images compressed to different quality levels. Original images (a) "Fruits and Flowers", (b) "Jussi", (c) "Colorchecker", (d) "Icon". Compressed to (e) 0.0005 bits/pixel, (f) 0.0025 bits/pixel, (g) 0.001 bits/pixel, (h) 0.0001 bits/pixel. (i), (j), (k) and (l) show error maps of the compressed images compared to the originals applying two dimensional SSIM index to every band of the image and meaning the result.*

$$MSSIM(\mathbf{X},\mathbf{Y}) = \frac{1}{M}\sum_{j=1}^{M} SSIM(\mathbf{x}_j, \mathbf{y}_j) \tag{14}$$

where **X** and **Y** are the reference and the distorted images, x_j and y_j are the image contents of the jth local window, M is the number of local windows of the image.

Experimental Result

We apply our image quality assessment algorithm to distorted images created from the same original image, using the same type of distortion, based on a certain compression method.

It is known from publications [1,12] that one of the best results in spectral image compression gives the usage of PCA compression in spectral domain and wavelet compression in spatial domain. From all wavelet based encoders we used the so called SPIHT, because it is easy to use and achieves good results at high compression ratios [1].

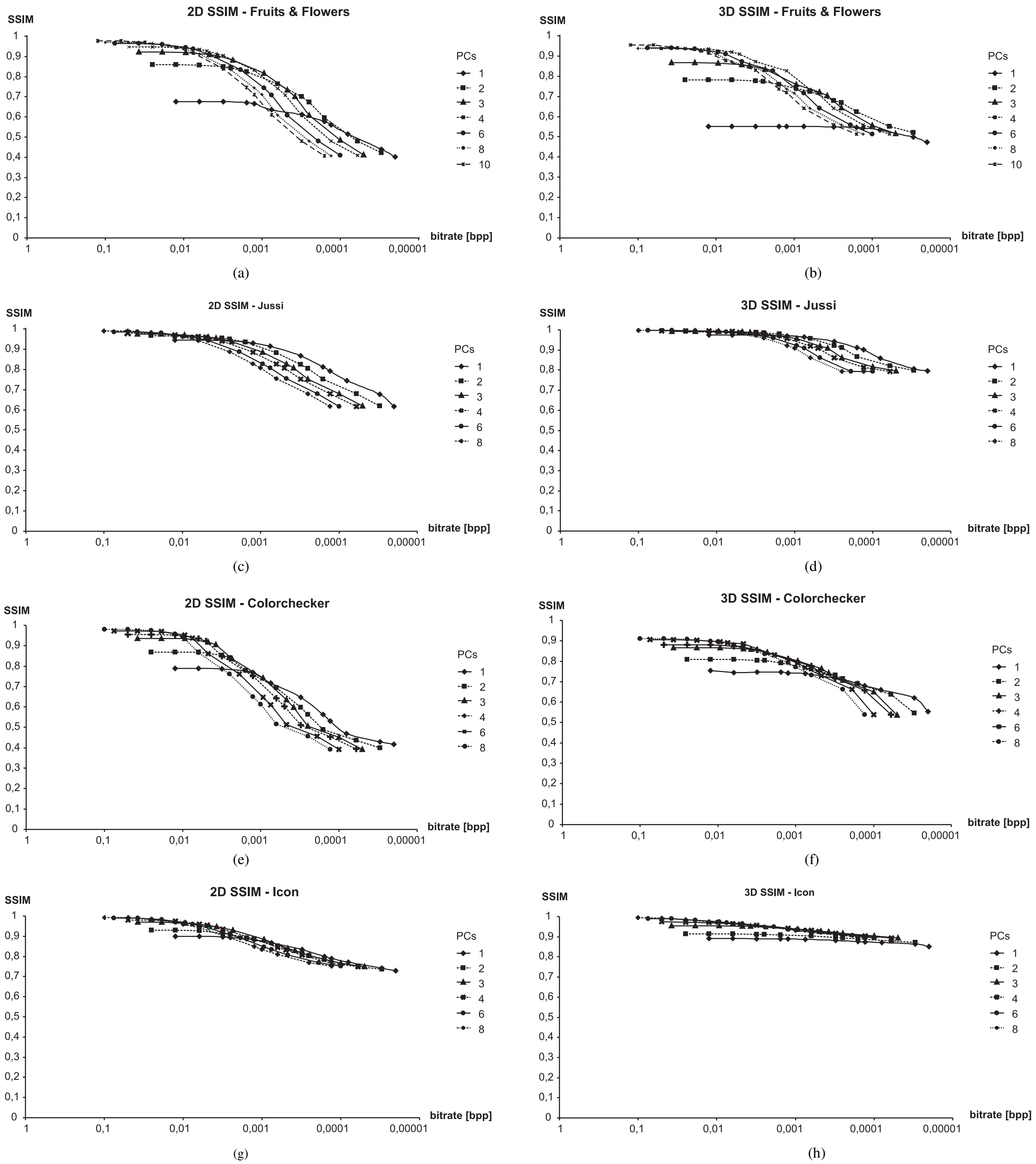

Figure.2. *Plot of the 2D and 3D SSIM index results versus the bitrate. Each sample point represents one test image. (a), (c), (e) and (g) – using two dimensional SSIM index; (b), (d), (f) and (h) – using three dimensional SSIM index. (a) and (b) – image "Fruits and Flowers", (c) and (d) – image "Jussi", (e) and (f) – image "Colorchecker", (g) and (h) – image "Icon"*

For the tests were used 4 spectral images consisting of 81 components - from 380nm to 780nm through 5nm. The spatial size of the images was 120x200 pixels. Each image was compressed with PCA in spectral domain and SPIHT in spatial domain to different ratios and after that restored and compared with the original one using SSIM index. The PCA compression ratios were from 8 (10 principle components kept after PCA) to 80 (1 principle component kept after PCA) and SPIHT compression ratios from 1 to 1000. Thus the common bitrate after compression was between 10^{-1} – 10^{-5} bpp leading to a total set of 336 pictures.

It is important to remember, that a multispectral image can not be observed directly. Every component can be shown as a grayscale image after scaling and rounding or the whole image must be converted to another color space that could be observed by humans (RGB). Another important feature comes from the cuboid structure of the spectral images. Although the two spatial directions represent the structure of the image, and the third (spectral) direction – the color, these are not independent. It is proved in [4] that changes in structure lead to changes in color and vice-versa. All original and compressed images and error maps were converted to RGB color space for the purpose of observation and afterwards visual assessment tests.

The four images we used during testing have different content and behavior according to SSIM. This comes both from their colorfulness and their structure. The “Colorchecker” is an artificial image consisting of many sharp edged, equal colored fields. The “Jussi” picture contains a not much detailed human face, which colors and shades are one of the most difficult things for interpretation. “Fruits & Flowers” is a natural image, which is not high detailed, but is very colorful, covering big diapason of spectra. “Icon” image consists of just few colors, but a lot of small details, which structure could be lost during compression. Analyzing the plots in Fig. 2 we can observe the main features of 2D and 3D SSIM in spectral imaging.

To check the correctness of the analytical results received we decided to perform visual assessment tests (Fig.3). For the tests 108 images, grouped in 4 sets, were chosen out from these 336 in such a way, that they cover the whole diapason of reasonable color and structure compression levels used in SSIM image testing. Some of the images with similar characteristics and obviously no quality difference were skipped for the ease of the observers. Subjects viewed trying to keep the recommendations of ITU-R BT.500 [11], although some new proposals about visual assessment appeared lately [8,9]. The observation of images was made from comfortable viewing distances that allow the data to reflect natural viewing conditions. People were asked to give their opinion about every image on continuous linear scale that was divided into ten equal regions marked with adjectives “Awful”, “Very Bad”, “Bad”, “Not So Bad”, “Poor”, “Reasonable”, “Reasonably Good”, “Good”, “Very Good” and “Excellent”. The 4 sets of pictures were assessed twice, in normal and reverse order of the pictures, as recommended in [9]. The first picture in a set was the original one, but it was also included 2 more times in each set, so that subjects didn’t have information about it. This way we can obtain the mean opinion score about the originals also. All the results for every image were averaged by the number of subjects and number of observations.

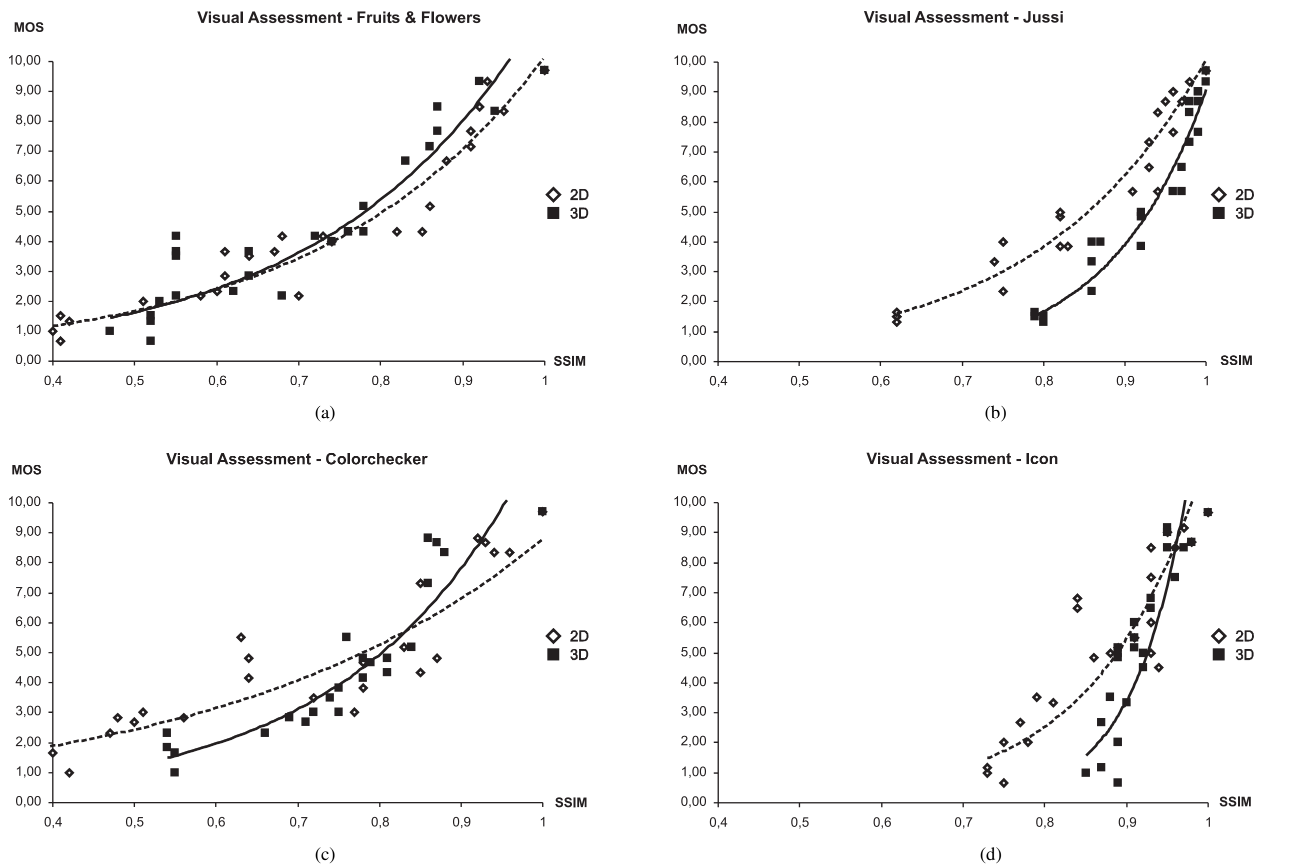

Figure.3. *Scatter plots of subjective mean opinion score (MOS) versus 3D SSIM index for different images*

Discussion

a) 2D SSIM accesses the image band by band and obtains the consecutive error maps, which are then averaged to obtain an overall error map. Although this way is implementing only the structure similarity test, it calculates indirectly color similarity. As mentioned above, compressing only spectral domain leads to change in color, but also to change in structure, that is observed with 2D SSIM (the first points in every principal component (PC) with no spatial compression). On the other hand 3D SSIM calculates error maps in both spectral and spatial domains and takes into account color and structure similarity. This is the reason why the differences between first dots of every principal component (no spatial compression) are bigger in 3D case than 2D case (Fig.2).

b) The plots of 2D SSIM go down faster with increasing spatial compression because of the same reasons. 3D SSIM index observes the absence or presence of color in the region during image error acquisition in addition to similarity changes. Comparing each line in plots between the 2D and 3D variants it is obvious that if structure is changed but color is somehow preserved in restored image, then the quality in 3D case is decreasing slower with increasing the level of spatial compression. Although spatial compression changes only structure of images, it also leads to color errors on most places, which can be understood from the error maps in Fig.1

c) According to the results we have a color-oriented similarity metric that tolerates the preservation of color in restored images after that type of used compression methods. This can be observed most clearly from plots of color images in Fig.2 in the case when only one PC is kept after PCA compression. In this situation, according to 3D SSIM, pictures are so much color-distorted, that the additional spatial compression almost cannot worsen them. In addition it can be seen from the pairs of pictures on Fig.2, where SSIM gives equal indices to image with less spatial compression (preserved structure) and bigger color loss on the one hand, and on the other image with distorted structure but somewhat preserved color range.

References

[1] A.Kaarna, "Multispectral Image Compression Using The Wavelet Transform", *Thesis for the degree Doctor of Science*, Lappeenranta University of Technology, 2000.

[2] Z.Wang and and A.C.Bovik, "A Universal Image Quality Index", *IEEE Signal Processing Letters*, vol.9, pp.81–84, Mar.2002.

[3] A.Kaarna and J.Parkkinen, "Blockwise Distortion Measure For Lossy Compression Of Multispectral Images", *Proc. EUSIPCO,* Sept.2000, pp. 2197-2200.

[4] Z.Wang, A.C.Bovik, H.R.Sheikh and E.P.Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. Image Processing, vol. 13, pp. 600-612, Apr.2004

[5] R.Janssen, "Computational Image Quality", *SPIE Press*, Bellingham, Washington, USA, 2001.

[6] A.M.Eskicioglu and P.S.Fisher, "Image Quality Measures and Their Performance", *IEEE Trans. on Communications*, vol. 43, pp. 2959-2965, Dec.1995.

[7] M.Miyahira, K.Kotani and V.Algazi, "*Objective Picture Quality Scale (PQS) for image coding*", IEEE Trans. on Communication, vol. 46, pp. 1215-1226, Sept.1998

[8] B.Keelan and H.Urabe, "*ISO 20462, A psyhophysical image quality measurement standart*", Proc. of SPIE—IS&T Electronic Imaging, vol. 5294, pp. 181-189, 2004

[9] W.Zou and Ph.Corriveau, "*Methods for Evaluating Of Digital Television Picture Quality*",

[10] D.Shen, J.Sung, "*Application of JND Visual Model to SPIHT Image Coding and Performance Evaluation*",

[11] Recommendation ITU-R BT.500-10, "*Metodology For The Subjective Assessment Of The Quality Of Television Pictures*", 2000

[12] R. Gonzalez, R. Woods: Digital Image Processing. New York. Addison-Wesley, 1992.

Author Biography

***Dobromir Dochev** received his MSc degree in communication technologies from Rousse University, Rousse, Bulgaria (1999). Since then he has worked in the Technology Division at Dunav press JSC and Rousse University. His research work has focused on image and video compression, image quality and color science. Now he is pursuing his Ph.D. degree at Rousse University.*