Reduced-reference Video Quality Metric Using Spatial Information in Salient Regions

Farah Diyana Abdul Rahman*[1], Dimitris Agrafiotis[2], Othman O. Khalifa[3], Fan Zhang[4]
[1,3] Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia
[2,4] Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK
*Corresponding author, e-mail: farahdy@iium.edu.my

Abstract
In multimedia transmission, it is important to rely on an objective quality metric which accurately represents the subjective quality of processed images and video sequences. Maintaining acceptable Quality of Experience in video transmission requires the ability to measure the quality of the video seen at the receiver end. Reduced-reference metrics make use of side-information that is transmitted to the receiver for estimating the quality of the received sequence with low complexity. This attribute enables real-time assessment and visual degradation detection caused by transmission and compression errors. A novel reduced-reference video quality known as the Spatial Information in Salient Regions Reduced Reference Metric is proposed. The approach proposed makes use of spatial activity to estimate the received sequence distortion after concealment. The statistical elements analysed in this work are based on extracted edges and their luminance distributions. Results highlight that the proposed edge dissimilarity measure has a good correlation with DMOS scores from the LIVE Video Database.

Keywords: Video quality, quality of service, video signal processing, reduced-reference metric, objective video quality metric

1. Introduction

In the recent emerging multimedia systems and applications, user requirements are going beyond requirements on connectivity, and users now expect the services to meet their requirements on quality. So, it is important to evaluate the quality of the received video sequence with minimal reference to the transmitted one. In this paper, a novel reduced-reference video quality metric using edge-based feature on salient regions is developed and analysed. The metric relies on saliency as well as spatial information (SI) differences within frames of transmitted videos and received videos. A novel video quality measure is developed and known as Spatial Information in Salience Region Reduced Reference Metric (SISRR). The aim of this work is to discover an improved reduced-reference (RR) video quality metric (VQM) that makes use of the spatial information values incorporated with saliency maps of a transmitted and received sequences to estimate the quality of the received video in the presence of errors.

Regions in a video frame are considered salient if they attract the visual attention of the viewer. Visual attention has been investigated in numerous research fields such as cognitive psychology, neuroscience and computer vision [1]. Saliency models can generally be divided into two categories based on the applications. Some saliency detection models are developed to predict human fixations [2]-[9] and some models are developed to identify salient regions or objects [10]-[14]. Fixation prediction models generally generate sparse separated salient regions whereas salient object detection produces smooth linked saliency regions in video quality assessment the salient object detections method is mostly employed. The aim of this paper is to design a reduced reference quality metric that makes use of salient region detection. Each pixel in a saliency map represents the importance of the object that contains the pixel in the scene. A saliency map allows high resolution analysis on the most relevant parts of the visual field. Therefore, image or video processing would be very efficient in processing complex scenes and its fixated region by the fovea [15].

2. Previous Work

The visual quality assessment task may seem simple, but it involves a set of complex mechanisms that are not completely understood. Koch and Ullman [16] proposed a model generating saliency maps and it was one of the significant and influential models in the field. The model has been since extended and advanced maps have been proposed over time [3,17]. These models fundamentally, find the dissimilarity of intensity, colour, orientation and other factors to determine the saliency of the image or frame. Based on each property, the dissimilarities values are summed over several scales to form conspicuity maps. The conspicuity maps are then normalised and merged to construct the saliency map. From the saliency map, maxima values are identified based on the thresholds designated for the models. Other approaches define saliency as local complexity where scale localised features with high entropy are measured [18]. Another approach to measure saliency is by using several variations of local symmetry operator as proposed in [19]. A better outcome with relation to human perception was observed compared to contrast saliency work of [3].

The work in [20] considered prior high-level knowledge instead of typical bottom-up process when computing saliency maps. The high-level knowledge consists of combination of trained decision trees, where different sized windows are grouped into different object classifications. Pixels grouped into non-background are then considered to be in the saliency map.

There are two general approaches in determining low-level saliency, using biological models or computational models [21]. These approaches can be further divided into the number of scales involved in the model algorithm, either a single scale or a number of scales. The salient region detection is based on the contrast determination filter over a few scales to produce several saliency maps. These maps are then pooled to create the final saliency map. This method has previously been used in [21] and has produced a good result in detecting saliency regions. The segmentation would be based on the hill-climbing algorithm. The novelty of this work is to find saliency maps that incorporated with HVS that is the edges as well spatial energy of the whole frame. In general, all saliency detection methods are based on finding the local contrast of the image or frame by comparing regions using different features. These features consist of the colour, intensity and orientation. Typically, each feature produced its own saliency map, and the combinations of all the features' maps generate the final saliency map [22]-[25]. An approach to find the saliency map using center-surround differences using several feature maps of colour, intensity and orientation is proposed in [23]. This method reduced the computational time by using integral images in Visual Object Detection with a Computational Attention System (VOCUS). However, due to reducing the feature saliency maps size to a lower scale, the final saliency map has lower resolution and losing data. However, the work done in this paper managed to maintain the resolution by resizing the feature saliency maps at   each scale.

In [24], Hu et al. presented a thresholding analysis approach instead of a scale-space one. The approach uses histogram entropy thresholding of colour, intensity and orientation. In this work, the measures used are based on the spatial compactness measure as well as saliency density. The spatial compactness measure is performed by rounding up the exterior body of the salient regions and the saliency density is a function of weighing each individual magnitude of saliency features before combining them. A spatial attention saliency-based biological-driven computational model is proposed in [25]. The features that are taken into account to generate the saliency maps are luminance, colour and orientation at different scales. Then these magnitudes are aggregated and combined for each location in the image, which in turn using a bottom-up approach, are combined to generate the concluding saliency map. Work in [22] proposed a local contrast-based method to generate saliency map by using single scale operation and does not considered any biological model. The input to the operation consists of image that has been resized, and then colour quantized using CIELab space and also then sub-divided into pixel blocks. The proposed operation consists of summing up differences between image pixels and their surrounding pixels within a small neighbourhood, which then produces the output of the saliency map.

3. Proposed Method

The proposed RR metric generates saliency maps using low level features as well as using spatial information of the reference as well as distorted sequences' frames. The algorithm of finding salient regions consists of computing the saliency map at different scales of the input image or frame, adding each of the saliency maps based on the different scales at each pixels and then the added values are averaged and normalised. This process generates the comprehensive visual saliency map. Figure 1 shows the flow diagram of the proposed SISRR. For the purpose of this work, no segmentation is performed due to the objective of this work. The aim is to produce a RR video quality metric using the information from the saliency regions combined with the SI. Therefore, there is no need for any segmentation to be performed.
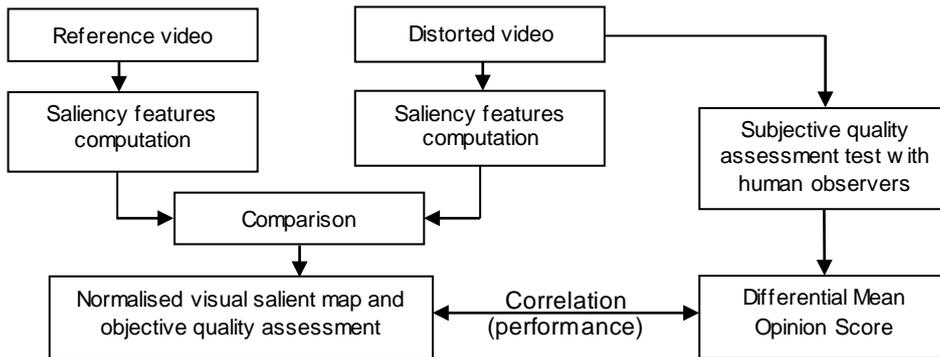

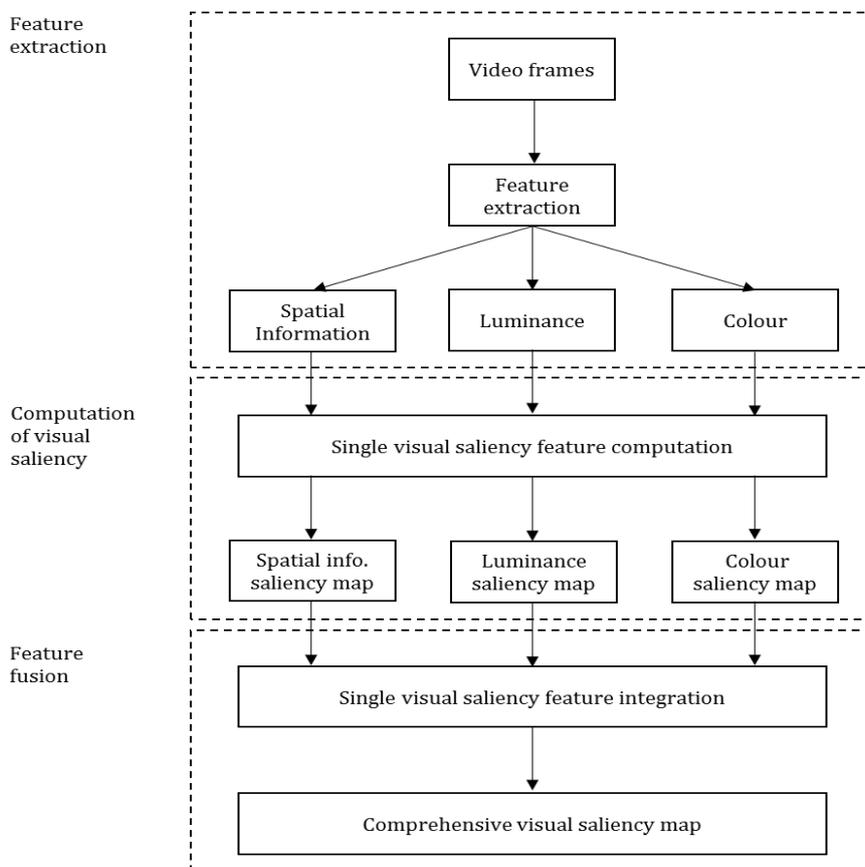Figure 1. Block diagram of the proposed quality assessment, SISRR

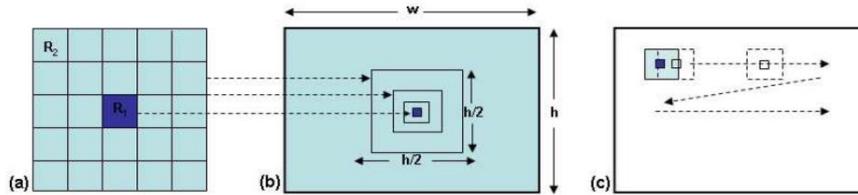Figure 2. Visual saliency computation (adapted from [27])

Figure 3. (a)The sub-region (R₁) and its surrounding neighbour region (R₂), (b) The width of R₁ remains the same, whereas the width of R₂ varies by reducing by a factor of two each time for a smaller scale, (c) The frame is filter by scanning in raster method. Figure reproduced from [21]

In this work, the local contrast of a region within a video frame with respect to its neighbourhood is used to determine the saliency. The local contrast is found by calculating the distance between average vectors of the pixels within the region to its respective average vectors pixels in the neighbourhood. as suggested by Achanta et al. [21] which used two low level features, colour and luminance. The method is easy to implement, noise tolerant and fast to compute compared to other complex saliency models. The saliency detection managed to capture the salient regions successfully and the region scope is not too limited and also not too wide [26]. The flow diagram of the visual computation is shown in Figure 2.

For each scale, the contrast value, $c(i,j)$ is determined as the distance D between the average vectors of the region R_1 and region R_2. The coordinate (i,j) is the pixel position within the frame as shown in Figure 3 and can be calculated as follows:

$$c(i,j) = D\left[\left(\frac{1}{N_1}\sum_{p=1}^{N_1} \boldsymbol{v}_p\right),\left(\frac{1}{N_2}\sum_{q=1}^{N_2} \boldsymbol{v}_q\right)\right] \tag{1}$$

where N_1 and N_2 are the number of pixels in regions R_1 and R_2 respectively, and $\mathbf{v}$ is the vector of feature elements corresponding to a pixel. In order to generate feature vectors for colour and luminance, the CIELab colour space [28] is used. The average feature vector values of R_1 and R_2 are computed by using the integral image approach as applied in [29]. In this method, filter region scaling is performed instead of image scaling. This allows generating saliency maps of the same size and resolution as the input. Filtering is performed at three different scales for each frame as seen in Figure 3 and the final saliency map is determined as the summation of saliency values across all three scales S:

$$m(i,j) = \sum_S c(i,j) \tag{2}$$

where $m(i,j)$ is an element of the summation saliency map M. The saliency map itself is obtained by pixel-wise summation of saliency values across the scales.

Another feature to be extracted is the spatial information (SI) itself. SI measurement evaluates the spatial information details. It is closely related to the perception of the human viewer, where the human viewer notices that changes or distortions occur spatially. SI measurement is standardised in ITU-T Recommendation P.910. The measurement has low computational complexity, as it is easy to calculate using well-known technique that is the Sobel filter. The Sobel filter is a simple high-pass, edge enhancement digital filter which is widely used in image processing. In short, SI is an indicator of edge energy. In order to calculate the value of SI for one video frame, a Sobel filter is first applied on the luminance values. The SI value of frame Fn at time n is then equal to the standard deviation of the image resulting from convolving frame Fn with the Sobel kernel:

$$\text{SI} = \left\{\text{std}_{\text{space}}[\text{Sobel}(\text{F}_\text{n})]\right\} \tag{3}$$

4. Simulation Setup
The test sequences were obtained from the LIVE Video Quality database [30]. The proposed metric is tested against the subjective quality score, DMOS, provided from the

subjective study carried out by LIVE [31-][32]. The DMOS values range from 0 to 100, where the smaller value expresses the greater quality and the larger value states the worse quality, are collected using the subjective test model specified in ITU-R BT 500.11. The subjective study was conducted using a single stimulus procedure with hidden reference removal and the subjects indicated the quality of the video on a continuous scale. Subjects also viewed each of the reference videos to facilitate computation of difference scores using hidden reference removal. From the database, 80 distorted video sequences were obtained from 10 different high-quality videos with a wide variety of content as reference videos.

A set of 80 distorted videos are tested using two different distortion types: H.264 compression and simulated transmission of H.264 compressed bitstreams through error-prone wireless networks, as these types of distortions relate the most to the work performed in this chapter. The diversity of distortion types is to test the ability of the proposed objective model to predict visual quality consistently across distortions. The H.264 compression system produces fairly uniform spatial and temporal distortions in the video. Network losses, however, cause transient distortions in the video, both spatially and temporally. The H.264 compressed videos exhibit a visual appearance of typical compression artifacts such as blur, blocking, ringing and motion compensation mismatches around the edges of the main body in the frame. Videos obtained from the wireless transmission error exhibit errors that are restricted to small regions of a frame. Errors sustained by an H.264 compressed video stream in a wireless environment are also spatio-temporally localised distortions, due to the small packet sizes or temporally transient and appear as glitches in the video. A packet transmitted over a wireless channel is susceptible to transmission errors due to various factors such as shadowing, attenuation, fading and multi-user interference in wireless channels.

All of the ten uncompressed high-quality YUV sequences used have the resolution of 768 x 432 pixels. Each sequence was assessed by 29 valid human subjects in a single stimulus study where the scores are based on a continuous quality scale. The DMOS from the subjective evaluations are used to compare with the differences between transmitted and received visual comprehensive saliency maps. The work scope deals mainly with streaming video over a multicast network that required the transported bitstream to be able to be decoded and displayed in real-time. In this work, all of the frames are used to determine the most suitable similarity measure. However, only the last reference frames in each GOP from both the reference and received sequences perform as inputs in the proposed quality assessment system. This is due to the fact that it is crucial in keeping the overhead bit rate as low as possible, as well its practicality and realistic in keeping with real-time wireless transmission over a multicast network scenario. The system outputs a value to quantify the quality of the distorted sequence. The Live Video Database has been evaluated by many researchers and has been verified with various objective performance metrics [32]-[41].

5. Result Analysis

The first analysis performed is the correlation between the DMOS and the quality measure resulting from the differences between the reference and the distorted. The correlation coefficients are acquired between the two parameters in order to compare and justify the performances relatively, in terms of prediction accuracy, monotonicity and consistency performances. Table 1 shows the performance of all objective models using LCC for each sequence and its average for every objective quality metric. The averaged performance across all sequences shows that SISRR, with LCC equals to 0.851, outperforms SSIM and VIFP, whereas PSNR has the highest correlation at 0.901. This shows that with additional information, that is the saliency features, the quality measure has increased its correlation with the subjective quality scores if compared to only using edge feature. Table 2 shows the performance of all objective models using LCC for each distortion types. All sequences' LCCs are averaged for each distortion types and the overall average across both distortions shows that SISRR, with LCC equals to 0.941, outperforms PSNR and VIFP, whereas SSIM has the highest correlation at 0.943. This is again another improvement for SISRR when compared to EDIRR, where the LCC average across both distortions is 0.938. Table 3 compares the performance of all objective models using LCC, SROCC and KRCC for the entire LIVE Video Quality Database. These correlation coefficients are computed on each sequence and then

averaged together to get the results. The results once again show that PSNR has the highest correlation, followed by the proposed metric, SISRR, which outperforms SSIM and VIFP.

Table 1. The LCC between DMOS and SISRR, PSNR, SSIM and VIFP for each test sequence

Video Dataset	Sequence	SISRR	PSNR	SSIM	VIFP
LIVE	BlueSky	0.962	0.957	0.967	0.953
(Both distortions)	Mobile&Calendar	0.791	0.936	0.670	0.709
	PedestrianArea	0.758	0.915	0.651	0.503
	ParkRun	0.815	0.975	0.827	0.851
	RiverBed	0.766	0.926	0.730	0.728
	RushHour	0.849	0.877	0.820	0.656
	SunFlower	0.721	0.656	0.670	0.638
	Shields	0.981	0.952	0.965	0.989
	Station	0.893	0.878	0.927	0.919
	Tractor	0.970	0.933	0.947	0.936
	Average	0.851	0.901	0.817	0.788

Table 2. The LCC between DMOS and EDIRR, PSNR, SSIM and VIFP for different distortion sources

Video Dataset	Correlation Coefficient	SISRR	PSNR	SSIM	VIFP
LIVE (Wireless distortions)		0.892	0.877	0.896	0.862
LIVE (H.264 Compressions)	LCC	0.990	0.988	0.990	0.989
Average		0.941	0.933	0.943	0.926

Table 3. Comparing the quality scores between SISRR and DMOS using LCC, SROCC and KRCC

Video Dataset	Sequence	SISRR	PSNR	SSIM	VIFP
LIVE	LCC	0.851	0.900	0.817	0.788
(Both distortions)	SROCC	0.801	0.867	0.767	0.714
	KRCC	0.689	0.779	0.636	0.571
	Average	0.780	0.849	0.740	0.691

The results reported for the different distortion types for all of the databases presented in Table 2 also shows that SISRR outperforms PSNR and VIFP on wireless distortion induced images and it also performs on a par with PSNR, SSIM and VIFP for H.264 compressions distortions. It can be observed from Table 3 that these results are comparable with the verified results in [33] where extensive work has been done to compare subjective scores. The performance of all objective quality assessments used in this paper were validated using metrics relating to prediction accuracy, monotonicity and consistency as recommended in [42]. In addition, the proposed method exhibits very little complexity relative to all other methods (except PSNR) as shown in Table 4. Complexity was measured as the average execution time on an Intel i7-2600 CPU @ 3.40GHz PC and was normalised relative to the execution time of PSNR. All test metrics were realised in Matlab except MOVIE, which is realised in C.

Table 4. Comparison of the performance of video quality assessment (VQA) algorithms for wireless distortion (LIVE DATABASE)

Prediction Model	VQA	Complexity
PSNR	FR	1
SSIM	FR	13
VIFP	FR	49
VQM	FR	681
MOVIE	FR	2206
STRRED	RR	97
STIS-SSIM	RR	9
STIRR	RR	3
SISRR	RR	2

The results in Figure 4 shows the screenshots, salient pixels, the SI pixels as well as the comprehensive visual saliency maps using the proposed method on *Bluesky* sequence from LIVE database. From the results, it can be observed that the proposed method highlights the edges within the salient regions only. This observation reflects on the correlations achieved with the DMOS as edge differences have a higher perceptual significance to quality, in terms of structural distortion.
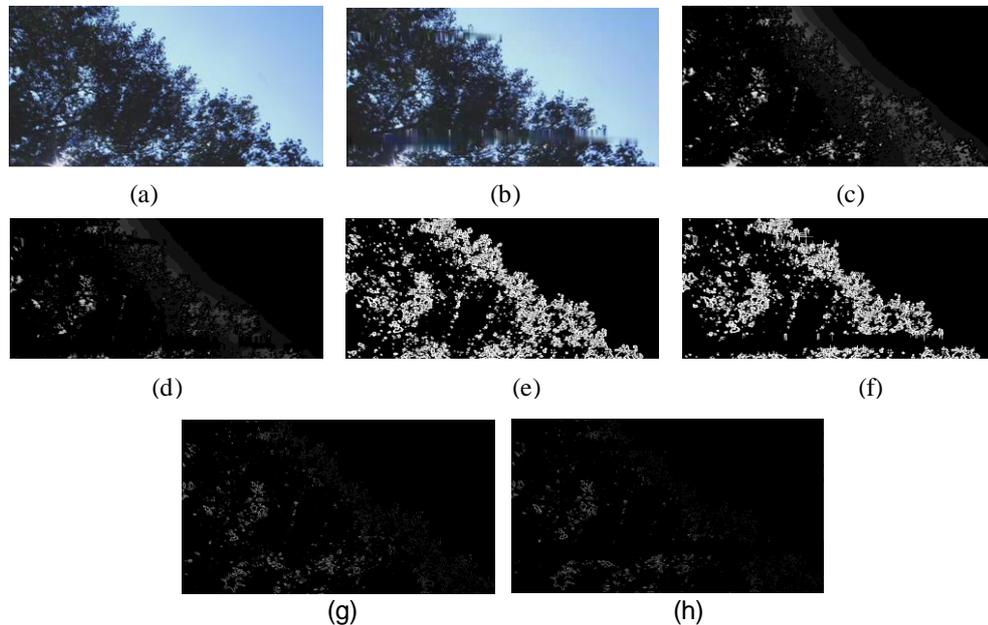

Figure 4. *Bluesky* screenshot of the (a) reference (ref.) frame (b) tested frame. Saliency map of the (c) ref. frame (d) tested frame. SI filtered of the (e) ref. frame (f) tested frame. The comprehensive visual saliency map of the (g) ref. frame and (h) tested frame.

6. Conclusion

In this paper, the possibility of extracting video quality information using the RR video quality metric by analysing the SI and salient regions using low level features of luminance and colour is examined. A novel method named Spatial Information in Salience Region Reduced Reference Metric (SISRR) is proposed in this paper. The metric is performed by comparing the combination of spatial and salient information of the original and distorted sequences based on the idea that edge differences have a higher perceptual significance to quality, in terms of structural distortion. The method is easy to implement, low complexity and fast enough to be used in real-time applications. The saliency maps have high resolutions which reflect the same resolutions as the input frames.

Even though the results show that PSNR has the highest performance in correlation with DMOS, the rank is followed by the proposed metric, SISRR, which outperforms SSIM and VIFP.The results obtained from SISRR shows some moderate correlations with quality values estimated by a number of full reference objective quality metrics which shows its suitability for simple albeit less accurate video quality assessment. It was also shown to outperform some full reference metrics when tested on the wireless distortion part of the LIVE video database.

References

[1] X Gao, T Wang, J Li. A content-based image quality metric. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. 2005; 231-240.

[2] GH Chen, CL Yang, SL Xie. *Gradient-based structural similarity for image quality assessment*. in Image Processing, 2006 IEEE International Conference. 2006: 2929-2932.

[3] JF Zhao, HJ Feng, ZH Xu, Q Li. An image quality assessment based on regional weight. *Optik-International Journal for Light and Electron Optics*. 2012; 123(6): 494-497.

[4] M Carnec, PL Callet, D Barba. *Visual features for image quality assessment with reduced reference*. in Image Processing, 2005. ICIP 2005. IEEE International Conference on. 2005: I-421-4.

[5] S AlZahir and F Kashanchi. *A new image quality measure,"* in Electrical and Computer Engineering (CCECE). 2013 26th Annual IEEE Canadian Conference on. 2013; 1-5.

[6] MS Prieto and AR Allen. A similarity metric for edge images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2003; 25(10): 1265-1273.

[7] MAR Díaz and HS Cruz. Refined fixed double pass binary object classification for document image compression. *Digital Signal Processing.* 2014; 30: 114-130.

[8] Y Yaegashi, K Tateoka, K Fujimoto, T Nakazawa, A Nakata, Y Saito, T Abe, M Yano, K Sakata. Assessment of Similarity Measures for Accurate Deformable Image Registration. *Journal of Nuclear Medicine & Radiation Therapy*. 2012.

[9] P Willett, JM Barnard, GM Downs. Chemical similarity searching. *Journal of chemical information and computer sciences*. 1998; 38(6): 983-996.

[10] KB Ho. A Review on Image and Video Processing. *International Journal of Multimedia and Ubiquitous Engineering*. 2007; 2(2): 49-64.

[11] R Jain, R Kasturi, BG Schunck. Machine vision. McGraw-Hill New York. 1995.

[12] JF Canny. Finding edges and lines in images. *Massachusetts Inst. of Tech. Report*. 1983; 1.

[13] J Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 1986; 6: 679-698.

[14] MSYL Bin. Comparison for Image Edge Detection Algorithms. *IOSR Journal of Computer Engineering (IOSRJCE).* 2012; 2(6).

[15] AH Lipkus. A proof of the triangle inequality for the Tanimoto distance. *Journal of Mathematical Chemistry*. 1999; 26(1-3): 263-265.

[16] S Theodoridis and K Koutroumbas. Pattern recognition. *IEEE Transactions on Neural Networks*. 2008; 19(2): 376.

[17] T Tanimoto. An elementary mathematical theory of classification and prediction, IBM Report (November, 1958). Cited in G. Salton. Automatic Information Organization and Retrieval. McGraw-Hill. 1968.

[18] JC Gower. Measures of similarity, dissimilarity, and distance. *Encyclopedia of statistical sciences*. 1985; 5(397-405): 3.

[19] JB Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika.* 1964; 29(1): 1-27.

[20] RP Woods, SR Cherry, JC Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *Journal of computer assisted tomography*. 1992; 16(4): 620-633.

[21] K Seshadrinathan, R Soundararajan, A Bovik, L Cormack. LIVE video quality database. *Laboratory for Image and Video Engineering*. 2009.

[22] TN Pappas, RJ Safranek, J Chen. Perceptual criteria for image quality evaluation. *Handbook of image and video processing*. 2000: 669-684.

[23] K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack. A subjective study to evaluate video quality assessment algorithms. *in IS&T/SPIE Electronic Imaging*. 2010: 75270H-75270H-10.

[24] I Recommendation, 500-11. Methodology for the Subjective Assessment of the Quality of Television Pictures. Recommendation ITU-R BT. 500-11. *ITU Telecom. Standardization Sector of ITU*. 2002.

[25] K Okarma. Image and video quality assessment with the use of various verification databases. *New Electrical and Electronic Technologies and their Industrial Implementatio*n. 2013: 142.

[26] S Li, L Ma, KN Ngan. Full-reference video quality assessment by decoupling detail losses and additive impairments. *Circuits and Systems for Video Technology, IEEE Transactions on*. 2012; 22(7): 1100-1112.

[27] AM Eskicioglu and PS Fisher. Image quality measures and their performance. *Communications, IEEE Transactions on*. 1995; 43(12): 2959-2965.

[28] L Ma, S Li, KN Ngan. Reduced-reference video quality assessment of compressed video sequences. *Circuits and Systems for Video Technology, IEEE Transactions on*. 2012; 22(10): 1441-1456.

[29] D Vranjes, D Zagar, O Nemcic. *Comparison of objective quality assessment methods for scalable video coding.* in ELMAR, 2012 Proceedings. 2012: 19-22.

[30] M Fang, J Xiuhua, S Hui, Y Shuang. Objective Perceptual Video Quality Measurement using a Foveation-Based Reduced Reference Algorithm. 308-311.

[31] KC Yang, CC Guest, KE Maleh, PK Das. Perceptual temporal quality metric for compressed video. *IEEE Transactions on Multimedia.* 2007; 9(7): 1528-1535.

[32] T Yamada, Y Miyamoto, M Serizawa. *End-user video-quality estimation based on a Reduced-Reference model employing activity-difference for IPTV services.* Consumer Electronics, 2009. ICCE '09. Digest of Technical Papers International Conference on. 2009: 1-2.

[33] T Yamada and T Nishitani. *No-reference quality estimation for compressed videos based on inter-frame activity difference*. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. 2012: 2325-2328.

[34] DS Turaga, C Yingwei, J Caviedes. No reference PSNR estimation for compressed pictures. 3: III-61-III-64.

[35] A Eden. No-Reference Estimation of the Coding PSNR for H.264-Coded Sequences. *Consumer Electronics, IEEE Transactions on*. 2007; 53(2): 667-674.

[36] SY Shim, JH Moon, JK Han. PSNR estimation scheme using coefficient distribution of frequency domain in H.264 decoder. *Electronics Letters*. 2008; 44(2): 108-109.

[37] C Keimel, M Klimpke, J Habigt, K Diepold. No-reference video quality metric for HDTV based on H.264/AVC bitstream features. 3325-3328.

[38] IP Gunawan and M Ghanbari. Efficient reduced-reference video quality meter. *IEEE Transactions on Broadcasting*. 2008; 54(3): 669-679.

[39] A Divakaran, K Peker, H Sun. A region based descriptor for spatial distribution of motion activity for compressed video. 287-290.

[40] Y Zhong, I Richardson, A Sahraie, P McGeorge. Influence of task and scene content on subjective video quality. 295-301.

[41] HR Wu and KR Rao. Digital video image quality and perceptual coding. *CRC press*. 2005.

[42] GE Legge and JM Foley. Contrast masking in human vision. *JOSA*; 70(12): 1458-1471.