A full-reference stereoscopic image quality metric based on binocular energy and regression analysis

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: GALKANDAGE, C. ...et al., 2015. A full-reference stereoscopic image quality metric based on binocular energy and regression analysis. IN: 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Lisbon, 8-10th July.

Additional Information:

- Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Metadata Record: [https://dspace.lboro.ac.uk/2134/19302](https://dspace.lboro.ac.uk/2134/19302)

Version: Accepted for publication

Publisher: © The Crown. Published by the IEEE

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
A FULL-REFERENCE STEREOSCOPIC IMAGE QUALITY METRIC BASED ON BINOCULAR ENERGY AND REGRESSION ANALYSIS

Galkandage C (1), Calic J (1), De Silva V (2), Dogan S (3)

(1) University of Surrey, UK; (2) Apical Ltd, UK; (3) Loughborough University in London, UK

ABSTRACT

The recent developments of 3D media technology have brought to life numerous applications of interactive entertainment such as 3D cinema, 3DTV and gaming. However, due to the data intensive nature of 3D visual content, a number of research challenges have emerged. In order to optimise the end-to-end content life-cycle, from capture to processing and delivery, Quality of Experience (QoE) has become a major driving factor. This paper presents a human-centric approach to quality estimation of 3D visual content. A full-reference quality assessment method for stereoscopic images is proposed. It is based on a Human Visual System (HVS) model to estimate subjective scores of registered stereoscopic images subjected to compression losses. The model has been trained with four publicly available registered stereoscopic image databases and a fixed relationship between subjective scores and the model has been determined. The high correlation of the relationship over a large number of stimuli has proven its consistency over the state-of-the-art.

Index Terms— Human Visual System, Binocular vision, Stereoscopic image quality metric

1. INTRODUCTION

Due to recent advances in multimedia and communication technology, 3D visual content has become a commonplace. However, the uptake and market impact has been limited, mainly due to user aspects of perceived 3D experience. In order to alleviate this, methods to analyse and utilise quantitative estimate of users Quality of Experience (QoE) is of a paramount importance to the area of 3D multimedia.

In order to be effective, estimate of perceived QoE should be utilised across the entire 3D media life cycle from capturing, processing and transmitting to displaying and interaction. Therefore, it is necessary to design the QoE estimation methodology so that the metric is as generic as possible, closely modelling the mechanisms of the ways users perceive the 3D visual content.

The estimation of QoE metric for 3D could ideally be done through subjective experiments, where human subjects would assess the quality of observed 3D content. However, this approach is expensive, time consuming and impossible to implement for real-time media. Therefore, the only feasible solution is to develop objective ways for quality estimation.

In this paper, a novel method for calculation of objective QoE metric for stereoscopic images is presented. The proposed approach is based on modelling of the mechanisms of human 3D perception, i.e. the binocular vision of the Human Visual System (HVS).

The rest of the paper is organised as follows. Section 2 presents state-of-the-art in stereoscopic quality metrics, while the proposed method is described with respect to analytical and statistical models in Section 3. Main outcomes of the proposed work are given in Section 4, while the discussion of achieved results are presented in Section 5, followed by the overall conclusions and our future research directions.

2. RELATED WORK

Development of quality metric for stereoscopic images and videos has been one of the major focuses of multimedia research community over the last decade. There are two different approaches to this problem in literature. First approach is to extend well established 2D quality metrics to 3D domain, while the second approach focuses specially on the 3D perception and defines models to mimic the 3D aspects of the HVS.

2.1. Extending 2D quality metrics to 3D domain

The approaches of extending 2D quality metrics to 3D domain are based on integration of separate metrics extracted from the left and right views. Well established metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [1] have been used to build algorithms for stereoscopic quality assessment. A study on performance of 2D quality metrics on stereoscopic images [2] has proved accuracy of correlation to subjective results was less significant with 3D images as compared to 2D images with 2D metrics.

Even with 3D information like disparity maps, depth maps or cyclopean images, it is still difficult to measure the stereoscopic quality by only extending the metrics developed for 2D content. Hence there was a research focus on interpretation and modelling of real 3D criteria for quality assessment both for 3D graphic objects [3] and stereoscopic images [4].
2.2. Model the HVS

Physiological studies of the HVS have introduced a signal processing model to mimic HVS behaviour. According to [5], the primary visual cortex (V1) is the main processing unit of HVS with simple cells and complex cells being main elements of V1. Simple cells are sensitive to size, orientation and phase of spatio-frequency information whereas a complex cell is connected to a pair of simple cells. Complex cells form relationships among themselves to generate a perceptual signal.

Research fields of physiology and signal processing have modelled the behaviours and functional roles of simple cells. There are monocular simple cells and binocular simple cells in the HVS. The founding definition of the receptive field of simple cells structure has been made in [6]. Appearance is primarily assumed to be the key for simple cells functional role, and this is understood over borders, contours and bars of a given stereoscopic view. Local spatial frequency analysis built on signal processing is used to model secondary roles of simple cells.

There are monocular complex cells and binocular complex cells just as the case with simple cells. The behaviour of complex cells is modelled in the proposal by [7]. The signal of two monocular simple cells feeds the receptive fields of monocular complex cells of the same retina. This signal in turn computes the monocular energy that corresponds to the occluded regions. The binocular energy is generated in the receptive fields of binocular complex cells from the input of two binocular simple cells which grouped in a quadrature phase. An illustration of monocular and binocular, simple and complex cells are shown in Fig. 1.

Based on the binocular fusion process characterising the 3D human perception, a full-reference metric is proposed for quality assessment of stereoscopic images in [5]. Binocular energy estimation is modelled in this model following a reproduction of the binocular signal generated by simple and complex cells. The relationship between the perceptual quality of stereoscopic images and visual information was explored in a model for binocular quality perception in [8]. Based on this model, a no-reference quality metric for stereoscopic images was proposed.

Level of development in the stereoscopic quality assessment research required a leap to match that of 2D quality assessment. To the best of our knowledge, there has been no perceptual 3D metric in the literature to match a 2D quality metric. Human 3D perception has to be explored deeply in order to understand and exploit phenomena in binocular vision process for the quality prediction.

3. PROPOSED METHOD

The proposed method is designed as a combination of analytical modelling and statistical data modelling. The analytical model is based on signal processing and comprises a sampling model, simple cell model and a complex cell model. The complex relationships among the complex cells are modelled statistically. These relationships terms are trained validated and tested using linear regression.

3.1. Analytical Model of the HVS

Based on the HVS based approach in the literature [5], a signal processing model is used in this research work. In addition to [5] complex cells response to binocular suppression in the presence of asymmetric stereoscopic contents are considered in the analytical model. Output of one complex cell is modulated by the output of another complex cell in the complex cell model in this research in order comply with physiological behaviour of recurrent excitation [9].

The output of complex cell model are formed into a number of objective scores based on luminance, chrominance and decomposed sub bands of spatial frequencies. In this research work decomposition has been done into three levels producing ten different spatial frequency bands representing different frequencies and orientations. Each such band is considered for luminance, chrominance, left-right fusion and left-right rivalry to calculate respective binocular energy values (10x4). A combination of all these 40 objective scores are mapped into one subjective score in statistical modelling [10].
3.2. Statistical Models

A number of statistical models are considered to find a relationship between HVS model outputs and subjective scores to overcome the limitations in [10]. This was done using a learning procedure of training, validation and testing on objective data of analytical models. As we have a number of objective scores for a single subjective score to map, the contribution from each objective score was required to determine through a learning procedure. This learning procedure is illustrated in Fig. 2.

A stepwise linear regression model is chosen to suppress unwanted objective scores from the relationship. Based on that a number of regression models were built with different combinations of training and validation sets while keeping the testing set constant with a composition shown in Table 1. All of the databases comprised stimuli of jpeg, jpeg2000, and blur artefact.

Pearson’s linear Correlation (PCC) and Spearman’s Rank Order Correlation (SROCC) were used to compare different regression models. The models with a better testing performance over PCC and SROCC are elected for the next stage of this research work. In order to select one model out of elected models, those models were applied directly without the learning process on any selected stereoscopic image to compare the correlation to the subjective score of that particular image. In this way a constant relationship between objective and subjective scores is available. The defined relationship in the regression model is considered as the quality metric formula in this research work. With the proposed quality metric, a state-of-the-art comparison is performed.

### Table 1. Composition of learning procedure

<table>
<thead>
<tr>
<th>Database</th>
<th>Trained images</th>
<th>Validated images</th>
<th>Tested images</th>
<th>Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live3D [11]</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>27x8</td>
</tr>
<tr>
<td>IRCYN [12]</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>15x5</td>
</tr>
<tr>
<td>NBU S [13]</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>21x12</td>
</tr>
<tr>
<td>NBU A [13]</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>27x10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>21</strong></td>
<td><strong>8</strong></td>
<td><strong>6</strong></td>
<td><strong>813</strong></td>
</tr>
</tbody>
</table>

IRCYN - Nantes University  
NBU - NingBo University; S-Symmetric, A-Asymmetric

Fig. 3. Results of learning procedure

models have been applied separately on arbitrary contents to predict subjective scores.

By assessing the linear regression models, a common candidate has been elected as the final regression model. Application of the final combination as a quality metric formula is performed and compared with state-of-the-art metrics. This formula is explained in (1) to match the recurrent excitation model in [10] and the importance of each term in the formula is represented in Table 2.

SSIMAvg [14], SSIMDdl [12], BEQM [5] and NoReferenceMetric [15] are used for above comparison. PCC indices of the proposed and state-of-the-art metrics are illustrated in Fig. 4. Moreover the total comparison to state-of-the-art metrics are presented in Table 3.

5. DISCUSSION

There are many aspects of the research work need to be discussed here. Hence a sequential explanation to the justification of selecting methods at different levels of this research is presented.

Using wavelet based modelling - Wavelets are sensitive to spatial impairments. Several physiological experiments
\[ Y = (a_1 \ a_2 \ \ldots \ a_n) \left( \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \right) + (x_1 \ x_2 \ \ldots \ x_n) \left( \begin{array}{cccc} b_{1,1} & b_{1,2} & \cdots & a_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n,1} & b_{n,2} & \cdots & b_{n,n} \end{array} \right) \]  

(1)

Table 2. Most significant terms of regression formula

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_{11}</td>
<td>-171.0585</td>
<td>16.3155</td>
</tr>
<tr>
<td>a_{39}</td>
<td>142.3221</td>
<td>16.3339</td>
</tr>
<tr>
<td>a_{31}</td>
<td>-2267.3</td>
<td>16.0514</td>
</tr>
<tr>
<td>a_{40}</td>
<td>-2267.3</td>
<td>280.9701</td>
</tr>
<tr>
<td>a_{16}</td>
<td>69.9404</td>
<td>8.7786</td>
</tr>
<tr>
<td>b_{16,38}</td>
<td>-406.9933</td>
<td>46.4231</td>
</tr>
<tr>
<td>b_{3,39}</td>
<td>928.6969276</td>
<td>128.5867041</td>
</tr>
</tbody>
</table>

Fig. 4. State-of-the-art comparison

Table 3. State-of-the-art comparison

<table>
<thead>
<tr>
<th>Metric</th>
<th>PCC</th>
<th>SROCC</th>
<th>AAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM Avg</td>
<td>0.7125</td>
<td>0.7760</td>
<td>26.9850</td>
<td>34.8257</td>
</tr>
<tr>
<td>SSIM Ddl</td>
<td>0.7353</td>
<td>0.7569</td>
<td>27.0595</td>
<td>34.8944</td>
</tr>
<tr>
<td>BEQM</td>
<td>0.6557</td>
<td>0.6739</td>
<td>25.5909</td>
<td>33.1214</td>
</tr>
<tr>
<td>NRIM</td>
<td>0.3556</td>
<td>0.3942</td>
<td>18.3200</td>
<td>22.5686</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.8392</td>
<td>0.8407</td>
<td>8.8708</td>
<td>12.1033</td>
</tr>
</tbody>
</table>

demonstrated that simple cells can be modelled using linear filters from their impulse response measured on the visual cortex. The impulse response has been approximated using a Gabor wavelet by [16]. Motivated by physiological studies of the visual perception, several sampling functions have been proposed, particularly directional wavelets.

Using Bandelet transform - Bandelet transform is very close to the behaviour of a simple cell [17]. It splits up the obtained sub bands in a quad tree of variable size following the image geometry and an orientation is computed and assigned to each block depending on the coefficients. The block is called a dyadic square and is characterised by its size, amplitude and orientation as a simple cell [18].

Using a stepwise linear regression - Stepwise suppression is used over linear regression with capability of filtering least required component from the analysis. In literature most of the quality metrics are formed of a weighted linear equation and those weights were determined using a learning procedure. Regression model also provides a linear equation of iteratively determined weights. This way a regression model behaves sufficiently close to the actual relationship, and hence could be considered as a metric. Selection of a linear regression model to be comparable with a quality metric was carefully performed to make sure that particular model is valid under most of the circumstances.

6. CONCLUSION AND FUTURE WORK

This paper proposed a novel method for calculation of stereoscopic image quality metric using HVS modelling and linear regression analysis. The analytical model of HVS is consistent with the physiology of human vision. We have utilised an existing simple cell model and worked towards refining a more precise complex cells model. The new definition of complex cells model is based on additions of complex cells behaviour to binocular suppression and implementation of recurrent excitation. A novel statistical model based on learning procedure is used as a technique to find a relationship between objective results of analytical model and subjective scores. Stepwise linear regression has used to find the mapping between objective scores of a stereoscopic image to its subjective score. Evaluation results show that the proposed human-centric approach to QoE estimation is feasible and achieves better results than the state-of-the-art methods. Registered stereoscopic images are used in this research work using four of publicly available databases. We have done better than the state-of-the-art methods with a consistent overall correlation to subjective results over a very large number of stimuli from 4 different databases.

Future directions of this research will be towards introduction of temporal properties to the existing HVS model. So far, temporal averaging has been tried, but the results were not consistently better when compared to the state-of-the-art methods. Moreover, application of this quality estimation method related to 3D visual media is also planned.
7. REFERENCES


