Quality Assessment Method For Synthesized Virtual Viewpoint Image

You Yang[1,2], Gangyi Jiang[1,2], Mei Yu[1,2], Yo-Sung Ho[3]
[1] Faculty of Information Science and Engineering, Ningbo University, Ningbo, 315211, China
[2] Institute of Computing Technology, Chinese Academy of Science, Beijing, 100080, China
[3] Dept. of Inform. & Comm., Kwangju Institute of Science and Technology, Kwangju, 500-712, Korea

Abstract: Virtual viewpoint image (VVI) synthesis technology plays an important role in interactive three-dimensional audio-video (3DAV) system. Quality of VVI directly affects picture quality of service in 3DAV system. A new quality assessment method for VVI is proposed in this paper. The proposed method combines synthesis technology and fuzzy function to evaluate VVIs. Experimental results show the accordance to bio-characteristics of human vision system (HVS) and the quality of VVIs.

1. Introduction

Defining a suitable image quality assessment (IQA) method has been identified to be a common problem in the context of 3DAV[1], especially for the synthesized VVI, where there is no original signal for assessment methods. In many cases, only subjective criterias can be used for assessment purpose. In this evaluation phase, psychological scaling paradigms are constructed without objective metrics[2,3]. It is also difficult to use traditional error sensitivity based IQA methods to 3DAV system due to disparity information among viewpoints. A fuzzy IQA method is proposed, VVI is synthesized by its neighbor views and then assessed by these references, and shown by Fig.1.

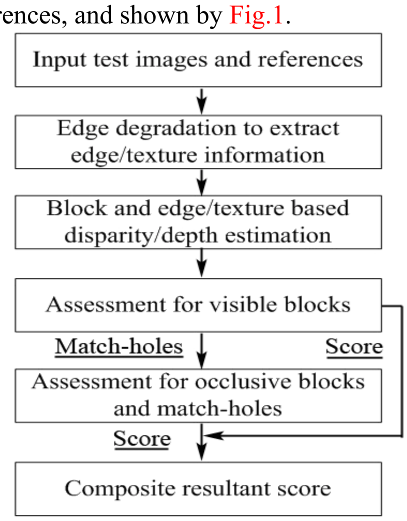


Fig.1 The proposed VVI quality assessment method.

2. Fuzzy Metric for VVI Quality

2. 1 Edge/texture extraction

It has been verified by many subjective tests on IQA that the evaluators tend to give low scores to an image whose edges/textures are noticeably blurred[4]. Therefore, it is important to extract edge/texture information in image. Let S_h, S_v, S_d and S_a denote horizontal, vertical, diagonal and anti-diagonal Sobel operator, respectively. Max-Sobel operator is defined here to preserve important direction information. An image with the size of $M \times N$ can be described by a fuzzy set. Let $X=\{(i,j)|0\le i\le M,\ 0\le j\le N\}$ be the image space, Max-Sobel is defined by

$$S_m(i,j)=\max\{S_h(i,j),S_v(i,j),S_d(i,j),S_a(i,j)\}. \qquad (1)$$

Some details in texture image have low energy that can not be observed clearly after being processed by Max-Sobel. Hence, a mask image based on $S_m(X)$ is needed. Let μ and σ denotes the mean and deviation of $S_m(X)$, respectively. μ can reflect the average energy of $S_m(X)$, while σ denotes energy fluctuate. Both of them can be used to express the statistical characteristics of $S_m(X)$. Then, the threshold α for $S_m(X)$ is determined by $\alpha=\mu+\sigma$. For α with ultra high value, the image must be a noisy image. On the other hand, the image is a planar one if α with low value and close to 0.

2.2 Dominant blocks for Comparison

VVI can be synthesized by its $2t$ (t>1) neighbor views, thus the blocks in VVI will contain disparity information with these views. Blocks in image are classified into visible or occlusive, determined by disparity estimation algorithm.

The source video sequences are captured by multi-cameras for 3DAV system. Illumination and chroma variation among different viewpoints will make disparity estimation far from precision. $S_m(X)$ is a binary image without these variations, and it is appropriated for disparity estimation. A structural MSE (*SMSE*) is defined for block based disparity estimation based on $S_m(X)$.

$$SMSE=\frac{\sum_{(i,j)\in block}\varepsilon^2}{block\_size}=\frac{\sum_{(i,j)\in block}(S_{m\_r}(i,j)-S_{m\_l}(i,j))^2}{block\_size}, \qquad (2)$$

where $S_{m\text{-}r}$ and $S_{m\text{-}l}$ are mask images for the right and left references, respectively. The disparity d is determined by minimum *SMSE* for blocks in image. Visible block is then determined by $d_{l\to r}=d_{r\to l}$.

2.3 Fuzzy metric for VVI quality

Let $\mathrm{P}(X)$ be a norm function on image $X(i,j)$. Fuzzy metric for multi-reference based VVI quality assessment is a fuzzy binary relation on $\mathrm{P}(X)$, i.e. a reflexive, symmetric mapping with the form of $\mathrm{F}(X)=\mathrm{P}(X)\otimes\mathrm{P}(X)\otimes\ldots\ldots\otimes\mathrm{P}(X)\to[0,1]$. Based on dominant blocks obtained above, the fuzzy metric for assess the quality of VVI is defined by

$$\begin{aligned}\mathrm{F}(X)&=\underbrace{\mathrm{P}(X)\otimes\mathrm{P}(X)\otimes\ldots\ldots\otimes\mathrm{P}(X)}_{2t+1}\\&=\sum_{i=1}^{t}f(L_i,V,R_i)\frac{|Mat|}{|V|}\to[0,1]\end{aligned} \qquad (3)$$

where t is the number of neighbor references on each side, *Mat* be the set of matched blocks. L_i, R_i and V are the ith left, right reference and VVI. $f(L_i,V,R_i)$ is a factor corresponding to the ith left or right references defined by

$$f(L_i,V,R_i)=\begin{cases}\dfrac{2i2\alpha_{S_{VIV}}/(\alpha_{S_{ith-l}}+\alpha_{S_{ith-r}})}{t(t+1)} & (4a)\\[2ex] \dfrac{2i(\alpha_{S_{ith-l}}+\alpha_{S_{ith-r}})/2\alpha_{S_{VIV}}}{t(t+1)} & (4b)\end{cases}.$$

Eq.(4b) is for noisy test images, while Eq.(4a) for ordinary images. It is noted that the coefficient $2i/(t(t+1))$ in Eq.(4a) and (4b) satisfies

$$\sum_{i=1}^{t}\frac{2i}{t(t+1)}=1 \cdot \qquad (5)$$

Thus, the coefficient is set to describe the fact that reference which far away from VVI will have less effectiveness than the near one. The parameter *Mat* in Eq.(3) is calculated by *SMSE* comparison between VVI and references

```
if (SMSE_VIV→left=SMSE_VIV→right) Mat=Mat+1
```

3. Experimental Results

"Barn 2" and "Sawtooth" image sequences provided by the test bed[5] are selected for experiments. The algorithm in [6] is applied for creating VVIs. For the sake of convenient, the original reference for VVI is preserved for PSNR. All experiments are performed with block size from 4×4 to 64×64 in the block matching. Small and large blocks are both approximate in evaluation. Therefore, the results based on 16×16 are listed.

Experiments are divided into "Adaptation Test" and "Real Test". Adaptation Test is performed to testify the effectiveness of the proposed metric. For Adaptation Test, the test images are obtained by the original image in the middle of left and right references. There are 33 JPEG compressed test images with different quantization parameters and 21 images with different degree of salt&pepper noise for Barn2 and Sawtooth (i.e. 108 images). Experimental results of Adaptation Test are shown in Fig. 2. As noted from these charts, scores change sharply at normal image quality, but laggardly at high and poor quality level. It is a bio-characteristic of HVS, which is much less sensitive in the case the image has overly high or poor quality, but maintains normal to well balanced quality. Hence, the resultant score can simulate HVS to predict the perceptual quality of image.

Real Test on VVIs is performed next. Results are listed in Talbe1. Clearly, the images with low distortions will have high objective score, and results of the proposed metric can well reflect this situation.

Table 1. Assessment scores for VVIs.

Name	Number of references	PSNR	Score
Barn 2	2	39.55	0.9470
Sawtooth	2	36.25	0.9082

Acknowledgements

This work was supported by the Natural Science Foundation of China (grant 60472100, 60672073), NSFC/KOSEF Joint Research Project, Natural Science Foundation of Zhejiang Province (grant Y105577), and the Key Project of Chinese Ministry of Education (grant 206059).

References

[1] A. Smolic and P. Kauff, "Interactive 3D Video Representation and Coding Technologies", *Proceedings of the IEEE*, Vol.93, No.1 pp.98-110, 2005.

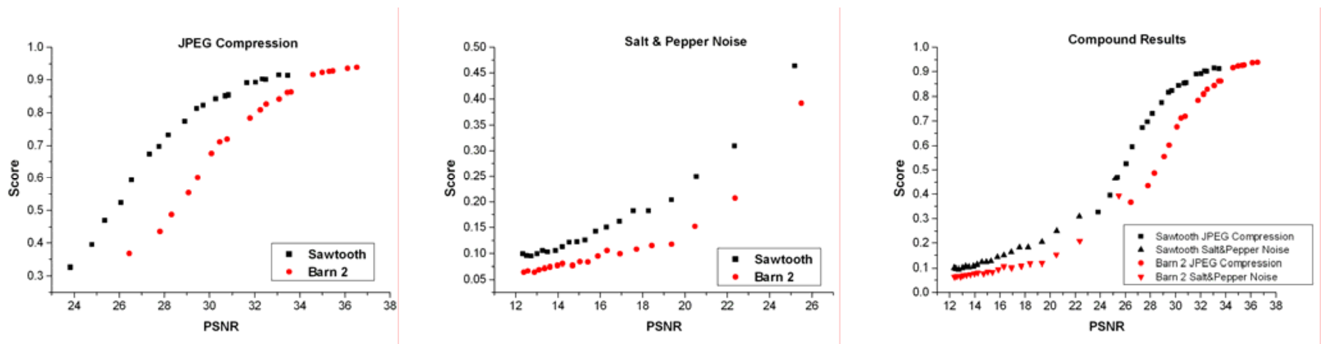
[2] L. M. J. Meessters, W.A. IJsselsteijn, and P. J. H. Seuntiëns, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV", *IEEE Trans. on CSVT*, Vol.14, No.3, pp. 381-391, 2004.

[3] K. Masaoka, A. Hanazato, M. Emoto, H. Yamanoue, Y. Nojiri, and F. Okano, "Spatial distortion prediction system for stereoscopic images", *Journal of Electronic Imaging*, Vol. 15(1), pp. (13002)1-12, 2006.

[4] C. Lee, S. Cho, J. Choe, T. Jeong, W. Ahn, and E. Lee, "Objective Video Quality Assessment", *Optical Engineering*, 45 (1), pp.1-11, 2006.

[5] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, Vol.47(1-3), pp. 7-42, 2002.

[6] G. Jiang, L. Fan, M. Yu, et al, "Fast Ray-Space Interpolation Based on Occlusion Analysis and Feature Points Detection", *Part II. LNCS* 3802 Springer, pp. 935-940, 2005.



(a)Test results for JPEG compression (b)Test results for Salt&Pepper noise (c)Compound results

Fig. 2 Resultant scores derived by test image set based on the proposed method. The outcome scores show the accordance with bio-characteristics of HVS.