IMAGE ENTROPY OF PRIMITIVE AND VISUAL QUALITY ASSESSMENT

Wuzhen Shi, Feng Jiang, Debin Zhao

School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

ABSTRACT

Recently, the concept of Entropy of Primitive (EoP) has been proposed to measure the image visual information. Some successful EoP based application also be developed. In this paper, we further explore the concept of EoP and propose an improved version: the L1 norm based EoP. Our EoP takes full account of the properties of a dictionary's layered structure and the characteristic of a basis pursuit method. Experimental results show that the L1 norm based EoP is superior to the L0 norm based one in measuring the image visual information. The curve of L1 norm based EoP holds a more consistent monotonicity with SSIM, its values is not trapped in the local convergence and the convergence value is less than that of the L0 norm based one. With the convergence characteristics of EoP, we further explore its application in stereoscopic image quality assessment (SIQA). With EoP as monocular cue and mutual information of primitive (MIP) as binocular cue, the relative entropy between the original stereoscopic image and the distorted one is used to compute the quality score by a prediction function which is trained using support vector regression (SVR). Extensive experimental results show that our new EoP based SIQA outperforms many state-of-the-art on the LIVE phase II databases.

Index Terms— Entropy of primitive, mutual information of primitive, visual quality assessment, stereoscopic image quality assessment, visual information

1. INTRODUCTION

To date, information theory has made a profound impact on many fields including electrical engineering, computer science, mathematics, physics, philosophy, and economics [1]. In this paper, we interpret information theory principles in the context of visual information representation. For data analysis, one may naturally wonder whether information theory can be applied to improve our understanding of the data and furthermore, to assist us to extract hidden salient data features. Based on the established sparse coding theory, Zhang et al. [2] proposed the concept of Entropy of Primitive and used it to measure the image visual information.

After the concept of EoP been proposed, some works have been done based on it. In [3], Ma et al. proposed a different version of EoP with respect to that the visual signal can be decomposed into structural and non-structural layers according to the visual importance of sparse primitives. In [4], Zhang et al. proposed an EoP based perceptual lossless profile to efficiently measure the minimum noticeable visual information distortion. In [5], Wang et al. proposed a reduced reference image quality assessment algorithm based on the EoP based distortion metric. In [6], Qi et al. proposed a reduced reference stereoscopic image quality assessment (RR-SIQA) metric by using binocular perceptual information (BPI). BPI is represented by the distribution statistics of visual primitives in left and right views' images, which are extracted by sparse coding and representation. In this paper, we interpret explicitly the BPI as the set of EoP and mutual information of primitive (MIP).

In recent years, 2D image quality assessment has made great progress, such as the structural similarity (SSIM) [7] and visual signal-to-noise ratio [8], but the stereoscopic image assessment is at the very beginning. According to the accessibility of reference image information, existing stereoscopic image assessment method may fall into three classes, i.e. full reference SIQA (FR-SIQA), reduced reference SIQA (RF-SIQA) and no-reference SIQA (NR-SIQA). In many applications the reference images are not available, and thus creating IQA algorithms that depend on much less information from the reference image is another focus of research. However, NR-SQIA are less efficient in providing a high correlation with the subjective quality evaluations because of the absence of the reference image information. To achieve a good tradeoff between the FR and NR algorithms, RR-SIQA algorithms become more popular.

In [3], authors apply k-means cluster algorithm to adaptively divide primitives into three categories, namely the primary, sketch and texture, respectively. In other words, they think each primitive has different importance. However, in spite they have fond this property of dictionary, it cannot be reflected in their new version of EoP. To take a full consideration on this property of dictionary, we propose a L1 norm based EoP and give experimental results to demonstrate that our L1 norm based EoP is superior to the L0 norm based one to measure the image visual information. With the convergence characteristics of EoP, we further explore its application in visual quality assessment, and propose a new EoP based RR-SIQA metric. Experimental results show that proposed method outperforms many state-of-the-art methods.

978-1-4673-9961-6/16/$31.00 ©2016 IEEE

2. ENTROPY OF PRIMITIVE

In this section, two types of entropy of primitive are introduced. Then, we give theory analysis and experimental results to prove that L1 norm based EoP is superior to the L0 norm based one.

2.1 Image primitive coding

The scheme of image primitive coding is established on the Sparseland model, which assumes that natural signals, such as images, admit a sparse decomposition over a redundant dictionary. It contains two stages: dictionary learning stage and sparse decomposition stage. In the dictionary learning stage, for an input image X, the dictionary learning process starts by partitioning the image into many overlapped patches, which are denoted by $x_1, x_2, \cdots, x_i,\ i = 1, 2, \cdots, N$. These patches are then collected as training samples. Assuming a local Sparse-Land model on image patches, the K-SVD dictionary training algorithm [7] is applied to the set of patches $\{x_i\}$, generating a content adaptive dictionary D:

$$D, \{a_i\} = \arg\min_{D,\{a_i\}} \sum_k \left\| x_i - Da_i \right\|_2^2 \quad s.t.\ \left\| a_i \right\|_0 < L \ \ \forall i, \tag{1}$$

where $\{a_i\}$ are the sparse representation vectors for $\{x_i\}$. In the sparse decomposition stage, for a patch x_i , the process of finding its sparse representation vector a_i with respect to a known over-complete dictionary D is called sparse coding. To obtain the sparse representation, sparse coding can be formulated as

$$a_i = \arg\min_{a_i} \left\| x_i - Da_i \right\|_2^2 \quad s.t.\ \left\| a_i \right\|_0 < L. \tag{2}$$

The orthogonal matching pursuit (OMP) [9] algorithm is used to solve this problem for its simplicity and efficiency.

2.2 Image Entropy of Primitive

The second law of thermodynamics states that the entropy of an isolated system never decreases and isolated systems always evolve toward thermodynamic equilibrium, a state with maximum entropy. For the biological agents such as HVS, this maximum entropy state could never be reached as their internal states are limited to a relative low entropy level for keeping themselves within some physiological bounds. This bound is determined by the level of "surprise" in a particular visual scene which is known as the "free energy principle" [10]. Analogous to this law, the entropy of primitive (EoP) measures the state in the image representation system when responding to a "surprise" environment.

Before introducing the EoP, useful mathematical notations should be firstly defined. Let's define n_j^0 as the number of every image primitives used for representing the patches. It means that the sparse representation vector a_i has one donation to the n_j^0 when the j^{th} coefficient of a_i is nonzero. Let n_j^1 indicate the sum of coefficient of every image primitives used for representing the patches.

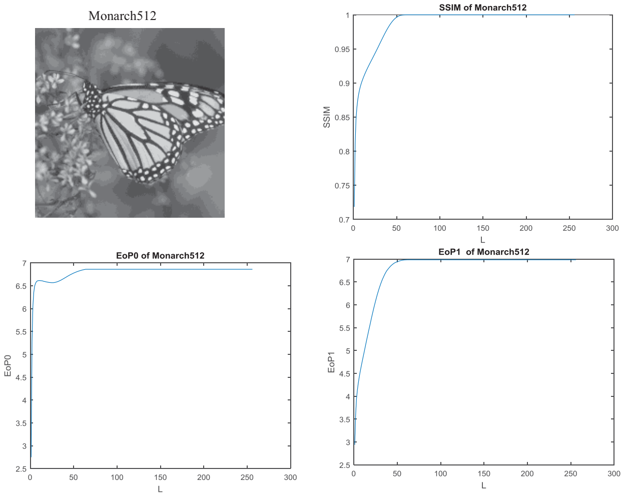

Fig. 1 SSIM, EoP0 and EoP1 of Monarch512 image in terms of L

Formally, given the sparse representation matrix $A = \{a_i\}$, n_j^0 and n_j^1 can be calculated as follows, respectively,

$$n_j^0 = \left\| A_j \right\|_0 , \tag{3}$$

$$n_j^1 = \left\| A_j \right\|_1 , \tag{4}$$

where $\|\cdot\|_0$ and $\|\cdot\|_1$ are L0 and L1 norm, respectively. Then two probability density functions (PDF) can be given by:

$$p_j^0 = \frac{n_j^0}{\sum_{t=1}^k n_t^0} = \frac{\left\| A_j \right\|_0}{\sum_{t=1}^k \left\| A_t \right\|_0} , \tag{5}$$

$$p_j^1 = \frac{n_j^1}{\sum_{t=1}^k n_t^1} = \frac{\left\| A_j \right\|_1}{\sum_{t=1}^k \left\| A_t \right\|_1} . \tag{6}$$

According to the relationship between the receptive fields of simple cells and primitives, we evaluate how surprise the receptive field reflects on the encountered input scene by estimating the information of primitives. Specifically, based on the Shannon theory, two types of entropy of primitive (EoP) are defined as,

$$EoP_0 = -\sum_{j=1}^k p_j^0 \log p_j^0 , \tag{7}$$

$$EoP_1 = -\sum_{j=1}^k p_j^1 \log p_j^1 . \tag{8}$$

It needs to note that EoP_0 is the original entropy of primitive proposed by zhang et al [2]. We just give other form of expression and point out it is based on L0 norm. As a result, this two kind of EoPs have similar form for easy comparison. They all have some interesting statistical properties based on a large number of statistics. Their values increase with the number of primitives L, and tends to converge to stable values, as illustrated in Fig.1. Moreover, they also have similar monotonicity with the SSIM curves. It demonstrates that both EoP_0 and EoP_1 have high correlation perceptual visual quality. The question is which one is better.

Let's recall the sparse coding process. It can be formulated as equation (2). The OMP method works in a greedy fashion that chooses the most similar primitive with the residual at each iteration. Note that the residual at the first iteration is the original patch itself. Then the original signal is subtracted by the chosen primitive to update the residual. The residual become smaller in each iteration until less than a predefined threshold. In other words, each primitive has different importance to the image patch. The primitive which is most similar with the image patch is picked first, followed by some details primitives for shaping the visual contents. It implies that computing the EoP should set different weight to each primitive. In this paper, we propose the sparse coefficients as an alternative. Experiment results show it work well.

Fig. 1 shows that both EoP_1 and SSIM tend to converge when L equal to about 60. In contrast, EoP_0 converge at about L equal to 10. Furthermore, EoP_0 is trapped in local convergence. In order to further quantitatively compare EoP_0 and EoP_1 , four popular evaluation criteria are chosen to compare the predicted quality score with SSIM, e.g. PLCC, SRCC, KRCC and RMSE. The SSIM score can also be treat as the ground true opinion score, then EoP_0 and EoP_1 are two corresponding predicted scores. Table 1 shows that EoP_1 gets better performance than EoP_0 .

Table 1 Performance comparison of EoP0 and EoP1

	PLCC	SRCC	KRCC	RMSE
EoP0-SSIM	0.8131	0.9966	0.9632	5.8070
EoP1-SSIM	**0.9846**	**1**	**1**	**5.8042**

Above all, we demonstrate that L1 norm based EoP is a better choice than L0 norm based EoP to measure image visual information. This is our main contribution. Next, we further discuss its application in visual quality assessment, and the EoP in the next section is EoP_1 .

3. EOP BASED SIQA

The stereoscopic image contains left view image I_L and right view image I_R .Given a dictionary D, we can get the sparse representation matrix A_L and A_R for I_L and I_R , respectively. Then the probability density of visual primitive d_k for I_L is calculated by

$$p_L^k = \frac{\left\|A_L^k\right\|_1}{\sum_{t=1}^{M}\left\|A_L^t\right\|_1}, \tag{9}$$

According to equation (8), the EoP of I_L can be calculated by

$$EoP\left(I_L\right) = -\sum_{k=1}^{M} p_L^k \log\left(p_L^k\right). \tag{10}$$

Similar to the left view image, the probability density and EoP of visual primitive d_k for I_R can be calculated by,

$$p_R^k = \frac{\left\|A_R^k\right\|_1}{\sum_{t=1}^{M}\left\|A_R^t\right\|_1}, \tag{11}$$

$$EoP\left(I_R\right) = -\sum_{k=1}^{M} p_R^k \log\left(p_R^k\right). \tag{12}$$

Human eyes are the front-end binocular system. It has been discovered that cells in the retina of each eye individually encode their received visual signal, and then the coded information, later merged in lateral geniculate nucleus, formulate the ultimate stereoscopic image in the brain [13]. Human visual perception depends on both monocular and binocular cues. Image artifacts may make the unnatural perception of HVS. When the artifacts are asymmetric, the balance of monocular perception is broken and the binocular perception is also changed. Therefore, both the monocular and binocular cues should be taken into account in SIQA metric. EoP_L and EoP_R can just be treat as the monocular cues. To get a more robust EoP based metric, some EoP based binocular cues need to be considered. In this paper, we use the mutual information of primitive (MIP) as an alternative.

The sum of coefficients that d_k is used to reconstruct both the i^{th} patch in left view and the j^{th} patch in the right view can be expressed as

$$a_{d_k}\left(i,j\right) = \begin{cases} \left|A_L^i\left[k\right] + A_R^j\left[k\right]\right|, & if\ A_L^i\left[k\right] \neq 0\ and\ A_R^j\left[k\right] \neq 0 \\ 0 & otherwise \end{cases}, \tag{13}$$

where $A_L^i\left[k\right], A_R^j\left[k\right], \left(i,j=1,2,\cdots,n\right)$ are the coefficients of d_k in the i^{th} patch of left view's image and the j^{th} patch of right view's image, respectively. Then, the sum of coefficients that d_k is used to reconstruct the patches in both the left view and the right view is calculated by

$$a^k = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{d_k}\left(i,j\right). \tag{14}$$

Then, the joint probability density of visual primitive d_k for I_L and I_R is calculated by

$$p^k = \frac{a^k}{\sum_{k=1}^{M} a^k}. \tag{15}$$

With the probability density distribution p_L , p_R and p, the mutual information of primitive (MIP) can be defined by

$$MIP\left(I_L;I_R\right) = \sum_{k\in\Omega}^{M} p^k \log\left(\frac{p^k}{p_L^k p_R^k}\right), \tag{16}$$

where $\Omega = \left\{k \mid p_L^k \times p_R^k \neq 0\right\}$.

The EoP and MIP are two information representation methods for stereoscopic images. However, they are global methods that lack structural information. To alleviate this problem, we divide the stereoscopic image into multiple sub-image. When seeing an image, ones' eye will focus on a few different points, and the accumulation of their responses is the final opinion score. To imitate this physiological characteristic, we divide the stereoscopic image into several sub-images in which each sub-image can be regard as a fixation point. For each pair sub-images, the BPI can be computed. If the stereoscopic image I is divided into N sub-images (in this paper N=4), then the BPI can be calculated by

$$BPI = \left\{ BPI_1, BPI_2, \cdots, BPI_N, BPI_C \right\}, \qquad (17)$$

where $BPI_i = \left\{ EoP\left(I_L\right), MIP\left(I_L; I_R\right), EoP\left(I_R\right) \right\}$. BPI_C is the BPI computed using the complete image. It needs to be noted that BPI_C equals to the BPI of Qi's method [6]. This makes our method has higher versatility since the BPI_C can be regarded as a special case of the proposed method.

As aforementioned, when observer watches an asymmetric distorted stereoscopic image, both monocular and binocular visual perceptual information have some loss. The value of visual perceptual information loss reflects the distortion level between the original and distorted stereoscopic images. The loss of binocular perceptual information can be computed by the relative entropy:

$$\begin{aligned} E_r &= BPI^o - BPI^d \\ &= \left\{ \begin{aligned} & EoP^o\left(I_L^o\right) - EoP^d\left(I_L^d\right) \\ & MIP^o\left(I_L^o; I_R\right) - MIP^d\left(I_L^d; I_L^d\right) \\ & EoP^o\left(I_R^d\right) - EoP^d\left(I_L^d\right) \end{aligned} \right\}, \end{aligned} \qquad (18)$$

where o and d indicate the originate image and the distorted one respectively.

The quality score of a stereoscopic image is computed using the perceptual loss vector by a prediction function $f\left(\cdot\right)$. That is, the final quality score is given by

$$Q = f\left(E_r\right), \qquad (19)$$

where $f\left(\cdot\right)$ is trained in advance using support vector regression ($\varepsilon - SVR$). Here, $f\left(\cdot\right): R^2 \to R$ takes E_r as input and produces output as a corresponding quality score. In the $\varepsilon - SVR$, the unknown function $f\left(\cdot\right)$ is constructed by linearly combining the results of a nonlinear transformation of the input samples

$$f\left(x\right) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) K\left(x_i, x\right) + b, \qquad (20)$$

where α_i and α_i^* are the Lagrange multipliers, and $K\left(x_i, x\right)$ is the kernel function to perform nonlinear transformation. This paper uses the radial basis function (RBF) kernel:

$$K\left(x_i, x\right) = e^{-\gamma \left\| x_i - x_j \right\|^2}, \qquad (21)$$

where γ is a positive number, which represents the variance of the kernel function. In the experiment, the regression of the prediction function is performed using the LIBSVM. Three parameters are finally fixed by a grid search, e.g. $\varepsilon = 0.5, C = 32, \gamma = 1$.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, not any more other experimental results are given to demonstrate the superiority of L1 norm based EoP. We just verify our new EoP based SIQA metric. To make this a fair comparison, we use the popular LIVE PhaseII 3D IQA database [11] to verify the performance of proposed improved SIQA metric, because most of the compare algorithms use it. It was created at the University of Texas at Austin, consists of 8 reference stereoscopic images and 360 distorted stereoscopic images. The database provides 5 distorted types, including White Noise (WN), JP2K, JPEG, Gaussian Blur (GB), and Fast Fading (FF). Each distorted image has difference mean opinion scores (DMOS).

Two popular evaluation criteria (PLCC and SROCC) are chosen to compare the predicted quality score after nonlinear regression with DMOS. A good objective method should have high PLCC and SROCC values. These two metrics can efficiently evaluate the prediction accuracy, monotonicity, and consistency of the IQA algorithms.

To evaluate the efficiency of the proposed SIQA metric, we choose two FR-SIQA metrics, i.e. Benoit et al.'s metric [12], You et al.'s metric [13], and three RR-SIQA metrics, i.e. Hewage et al.'s metric [14], Xu et al.'s metric [15], and Qi et al. [6] for comparison. The performance comparisons of PLCC and SROCC values for each distortion type on the LIVE phaseII database are listed in Table 2, where the indicator that gives the best performance is highlighted in bold. Table 2 shows that our EoP based SIQA metric performs better than the other five state-of-the-art ones.

Table 2 performance comparison of the metrics on LIVE

	Criteria	Benoit	You	Hewage	Xu	Qi	Proposed
WN	PLCC	**0.926**	0.912	0.891	0.918	0.891	0.892
	SROCC	0.923	0.909	0.880	**0.940**	0.904	0.929
JP2K	PLCC	0.784	**0.905**	0.664	0.752	0.858	0.888
	SROCC	0.751	0.894	0.598	0.751	0.776	**0.922**
JPEG	PLCC	0.853	0.830	0.734	0.788	0.871	**0.887**
	SROCC	**0.867**	0.795	0.736	0.768	0.736	0.753
GB	PLCC	0.535	0.784	0.450	0.938	**0.981**	0.957
	SROCC	0.455	0.813	0.028	**0.900**	0.871	0.826
FF	PLCC	0.807	0.915	0.746	0.914	0.925	**0.937**
	SROCC	0.773	0.891	0.684	**0.920**	0.854	0.790
ALL	PLCC	0.784	0.800	0.558	0.824	**0.915**	**0.915**
	SROCC	0.728	0.786	0.501	0.820	0.867	**0.903**

5. CONCLUSIONS

In this paper, we bridge the sparse representation and stereoscopic image quality assessment with the concept of entropy of primitive and mutual information of primitive. Firstly, the L1 norm based EoP is introduced based on sparse coding theory. It is demonstrated that EoP is highly relevant with visual information. Secondly, experimental results show that the proposed L1 norm based EoP is superior to the L0 norm based one in measuring the image visual information. Finally, with EoP as monocular cue and MIP as binocular cue, a new SIQA metric is proposed. Experimental results show that our EoP based SIQA metric performs better than many state-of-the-art SIQA methods.

6. ACKNOWLEDGEMENTS

This work has been supported in part by the Major State Basic Research Development Program of China (973 Program 2015CB351804), the National Science Foundation of China under Grant No. 61572155.

7. REFERENCES

[1] Cover, T.M.; Thomas, J.A. Elements of Information Theory, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.

[2] J. Zhang, S. Ma, R. Xiong, D. Zhao and W. Gao, "Image Primitive Coding and Visual Quality Assessment," PCM 2012, vol. 7674, pp. 674-685, 2012.

[3] Ma S, Zhang X, Wang S, Zhang J, Sun H, Gao W. Entropy of Primitive: From Sparse Representation to Visual Information Evaluation. IEEE Transactions on Circuits and Systems for Video Technology. 2015:1-1.

[4] Zhang X, Wang S, Ma S, Liu S, Gao W. Entropy of primitive: A top-down methodology for evaluating the perceptual visual information. 2013 Visual Communications and Image Processing (VCIP). IEEE; 2013:1-6.

[5] Wang S, Zhang X, Ma S, Gao W. Reduced reference image quality assessment using entropy of primitives. 2013 Picture Coding Symposium (PCS). IEEE; 2013:193-196.

[6] Qi F, Zhao D, Gao W. Reduced Reference Stereoscopic Image Quality Assessment Based on Binocular Perceptual Information. IEEE Transactions on Multimedia. 2015;17:2338-2344.

[7] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing. 2004; 13:600-612.

[8] Chandler DM, Hemami SS. VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. IEEE Transactions on Image Processing. 2007; 16:2284-2298.

[9] Tropp, J.A., Gilber, A.A.: Signal Recovery from Random Measurements via Orthogonal Matching Pursuit. IEEE Trans. on Information Theory 53(12), 4655–4666 (2007)

[10] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," IEEE Trans. Image Process., vol. 21, no. 1, pp. 41–52, 2012.

[11] M. J. Chen, L. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," IEEE Trans. Image Process., vol. 22, no. 9, pp. 3379–3391, Sep. 2013.

[12] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," EURASIP J. Image Video Process, vol. 659024, pp. 1–13, 2008.

[13] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in Proc. Int.Workshop Video Process. QualityMetrics Consum. Electron., 2010, vol. 9, pp. 1–6.

[14] C. Hewage and M. G. Martini, "Reduced-reference quality assessment for 3D video compression and transmission," IEEE Trans. Consum. Electron., vol. 57, no. 3, pp. 1185–1193, Aug. 2011.

[15] Q. Xu, G. Zhai, M. Liu, and K. Gu, "Using structural degradation and parallax for reduced-reference quality assessment of 3D images," in Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast., Jun. 2014, pp. 1–6.