Improved Entropy of Primitive for Visual Information Estimation

Shurun Wang #1, Zhenghui Zhao #1, Xiang Zhang #1, Jian Zhang #1, Shiqi Wang **2, Siwei Ma #*1 , Wen Gao #1

Institute of Digital Media, Peking University, Beijing 100871, China

* *Peking University Shenzhen Graduate School, Shenzhen, China*

1 $\{always1898pku, zhzhao, x\_zhang, jian.zhang, swma, wgao\}@pku.edu.cn$

** *Nanyang Technological University, Singapore 639798*

2 $wangshiqi@ntu.edu.sg$

***Abstract*—Sparse representation has been observed to be highly efficient in dealing with rich, varied and directional information in natural scenes. Based on the statistical analysis of primitives in sparse coding, the entropy of primitive (EoP) was proposed for measuring visual information of images, and its changing tendency has been shown to be highly relevant with the human visual system (HVS). But the sparse coefficient energy was ignored when calculating EoP, which may be critical in accounting for the primitive characteristics. To tackle this, an improved EoP is developed in this work via ℓ_2 norm calculation. We further give mathematical derivations for its convergence verification. Experimental evaluations have also demonstrated that the improved EoP can achieve more stable convergence tendencies, which is consistent with the perceptual experiences.**

***Index Terms*—Entropy of primitive, sparse representation, orthogonal matching pursuit, visual information estimation.**

I. INTRODUCTION

Sparse representation develops swift and becomes a powerful method in describing signals owing to its efficient in dealing with rich, varied and directional information obtained from natural scenes [1]. Extensive successful applications have been observed based on sparse models, e.g. image denoising [2], image restoration [3], [4], [5], [6] and image quality assessment [7], [8], [9]. Sparse theory claims that signals can be well represented by a few bases from an over-complete dictionary, where the dictionary is assumed to be highly adaptive to a set of signals within a limited subspace. There are two basic problems in sparse models, which are dictionary training and sparse decomposition, respectively. A typical algorithm for dictionary learning was the K-SVD [10], while the orthogonal matching pursuit (OMP) was proposed to solve the second problem.

What attracts us is how to measure the visual information in the natural scene with the form of sparse representation. To solve this, a novel concept was developed recently, i.e. entropy of primitive (EoP) [11], [12], [13], where the sparse coefficients distribution and its entropy value were used for measuring the visual information. It was also found that the EoP changing tendency was relevant to the visual cognitive process. Successful applications have been achieved from image quality assessment to just-noticeable difference (JND) model.

However, we observed that EoP tendency was not always consistent with the HVS, especially for complex scenes with unnaturalness. This may be caused by the definition of EoP, where only the count of nonzero coefficient values was taken into account while the coefficient amplitude was completely ignored during the calculations. The sparse value may be, however, very significant in accounting for the image representation. Therefore, an improved approach in calculating EoP is presented in this work, and its convergency performance has shown to be both theoretically and experimentally superior to the original EoP.

The rest of the paper is organized as follows. In Section II, the concept of EoP is introduced briefly. The improved EoP and its mathematical convergence proof are discussed in Section III. We further give experimental verifications in Section IV. Section V concludes this paper.

II. ENTROPY OF PRIMITIVE

In this section, the concept of entropy of primitive (EoP) is briefly reviewed.

A. Sparse Representation via Learned Dictionary

The theory foundation of EoP is the sparse-land model [14] assuming that natural images can be approximately represented by a linear combination from an over-complete dictionary. Formally, we can write $\forall Y \in \mathbb{R}^{s\times n}$, $Y \approx DX$, where $D \in \mathbb{R}^{s\times m}$ is the over-complete dictionary, and $X \in \mathbb{R}^{m\times n}$ is the sparse coefficient matrix. Each column of D is also called a *primitive*, and its corresponding coefficient vector x_i performs *sparse*, i.e. $||x_i||_0 \ll m$.

In order to train the over-complete dictionary, the K-SVD algorithm [10] is employed in this work, which is a kind of generalization of K-means. It consists of two processes, i.e.

VCIP 2016, Nov. 27 – 30, 2016, Chengdu, China

sparse coding and dictionary training. The problem can be formalized as the following formula:

$$min_{D,X} \|Y - DX\|_F, \ s.t. \|x_i\|_0 \leq L, i = 1, 2, ..., n, \quad (1)$$

where $\|\cdot\|_F$ is the matrix norm and is a measure of the error between Y and DX, and $\|\cdot\|_0$ represents the number of nonzero components. L is the sparsity level parameter that governs the ℓ_0 norm of sparse coefficients.

The orthogonal matching pursuit (OMP) [15] is a popular extension of matching pursuit family, which is a type of sparse decomposition algorithm targeting at finding the "best matching" projections of multidimensional data onto the over-complete dictionary D. The OMP method works in a greedy fashion that the most similar primitive with the residual is chosen after each iteration. It should be noted that all coefficients extracted so far are updated after each iteration, by computing the orthogonal projection of the residual onto the selected atoms, which can lead to more efficient representation.

B. Entropy of Primitive

Some mathematical notations should be introduced at first. Let n_j^i denote the incremental number of the j^{th} primitive selected in the i^{th} iteration during the OMP algorithm. N_j^i represents the total number of the j^{th} primitive selected in the previous i iterations, it can be written as follows,

$$N_j^i = \sum_{t=1}^{i} n_j^t. \quad (2)$$

Accordingly, the probability density function (PDF) is defined by,

$$P^i(j) \triangleq \frac{N_j^i}{\sum_t N_t^i}. \quad (3)$$

It represents the distribution of primitives in the previous i iterations which can also be comprehended as the cumulative distribution of the primitives during the OMP algorithm. Next, based on the $Shannon$ theory, the entropy of primitive (EoP) value can be defined as follows [13],

$$EoP_i \triangleq -\sum\nolimits_j^k P^i(j) \log P^i(j), \quad (4)$$

where k is the number of the primitives.

III. Improved EoP via ℓ_2 norm

A. Improved EoP

Despite its success, EoP was not always consistent with the HVS, especially for complex scenes with unnaturalness. This may be caused by the reason that only the count of nonzero values was taken into account in the definition of EoP. The amplitude of coefficient values may be however significant in image representations. EoP is an individual measurement of the information of image for sparse representation. In order to overcome this inconsistency, a new version of EoP based on the ℓ_2 norm is proposed in this work.

Put more formally, let $X_t = (x_{i,j}^t)_{m \times n}$ be the sparse coefficient matrix at the t^{th} iteration, then a PDF considering the coefficient energy can be defined as follows,

$$p_{t,i} = \frac{N_{t,i}}{N_t}, \quad (5)$$

where

$$N_t = \sum_{i=1}^{m} N_{t,i}, \quad (6)$$

$$N_{t,i} = \sqrt{\sum_{j=1}^{n} X_t^2(i,j)}. \quad (7)$$

Consequently, the improved EoP can be defined as,

$$I\text{-}EoP = -\sum_{i=1}^{m} p_{t,i} \log_2 p_{t,i}. \quad (8)$$

B. Convergency Verification

To verify the convergence of the proposed EoP, we try to prove its robustness and stability with mathematical derivations in this subsection. The basic idea is that the difference between two PDFs of successive iterations approaches to zero as the iteration increases, such that $I\text{-}EoP$ can be stable.

Firstly, we define two EoPs of successive iterations as follows,

$$I\text{-}EoP_t = -\sum_{i=1}^{m} p_{t,i} \log_2 p_{t,i}, \quad (9)$$

$$I\text{-}EoP_{t+1} = -\sum_{i=1}^{m} p_{t+1,i} \log_2 p_{t+1,i}, \quad (10)$$

where $p_{t+1,i} = \frac{N_{t+1,i}}{N_{t+1}}$, $X_{t+1} = (x_{i,j}^{t+1})_{m \times n} = (x_{i,j}^t + \Delta_{i,j}^t)_{m \times n}$ and $\Delta_{i,j}^t$ represents the coefficient difference between successive iterations in the sparse representation. Then the PDF of the $t+1$ iteration can be rewritten as follows,

$$p_{t+1,i} = \frac{\sqrt{\sum\limits_{j=1}^{n} X_t^2(i,j) + 2\sum\limits_{j=1}^{n} X_t(i,j)\Delta_{i,j}^t + \sum\limits_{j=1}^{n} {\Delta_{i,j}^t}^2}}{\sum\limits_{k=1}^{m} \sqrt{\sum\limits_{j=1}^{n} X_t^2(k,j) + 2\sum\limits_{j=1}^{n} X_t(k,j)\Delta_{k,j}^t + \sum\limits_{j=1}^{n} {\Delta_{k,j}^t}^2}}. \quad (11)$$

The ultimate goal is to verify that $\lim_{t\to\infty} |p_{t+1,i} - p_{t,i}| = 0$, which can be split to the following two parts,

$$\lim_{t\to\infty} |p_{t+1,i_{\min}} - p_{t,i}| = 0, \quad (12)$$

$$\lim_{t\to\infty} |p_{t+1,i_{\max}} - p_{t,i}| = 0, \quad (13)$$

where $p_{t+1,i_{min}}$ and $p_{t+1,i_{max}}$ are the lower bound and upper bound of $p_{t+1,i}$, respectively. In the following derivations, only the first part in (12) will be addressed, while the second part in (13) can be derived in a similar way.

In order to derive (12), we also need to split the discussion in whether the formula $\sum_{j=1}^{n} X_t \Delta_{i,j}^t$ is negative or positive, where only the case when $\sum_{j=1}^{n} X_t \Delta_{i,j}^t < 0$ will be given in the following.

According to (11), we have,

$$
\begin{aligned}
&p_{t+1,i} \\
&\geq \frac{\sqrt{\sum\limits_{j=1}^{n} X_t^2(i,j) + 2\sum\limits_{j=1}^{n} X_t\Delta_{i,j}^t}}{\sum\limits_{k=1}^{m}\sqrt{\sum\limits_{j=1}^{n} X_t^2(k,j) + \sum\limits_{j=1}^{n} {\Delta_{k,j}^t}^2} + \sum\limits_{k\in A_t}\sqrt{2\sum\limits_{j=1}^{n} X_t(k,j)\Delta_{k,j}^t}} \\
&\geq \frac{\sqrt{\sum\limits_{j=1}^{n} X_t^2(i,j)} - \sqrt{2|\sum\limits_{j=1}^{n} X_t\Delta_{i,j}^t|}}{\sum\limits_{k=1}^{m}\sqrt{\sum\limits_{j=1}^{n} X_t^2(k,j)} + B_t + \sum\limits_{k=1}^{m}\sum\limits_{j=1}^{n}|\Delta_{k,j}^t|} \\
&= \frac{N_{t,i} - \sqrt{2|\sum\limits_{j=1}^{n} X_t(i,j)\Delta_{i,j}^t|}}{N_t + B_t + ||\Delta^t||_1},
\end{aligned}
\tag{14}
$$

where

$$A_t = \{k|\sum_{j=1}^{n} X_t(k,j)\Delta_{k,j}^t > 0\}, \tag{15}$$

$$B_t = \sum_{k\in A_t}\sqrt{2\sum_{j=1}^{n} X_t(k,j)\Delta_{k,j}^t}, \tag{16}$$

$$||\Delta^t||_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}|\Delta_{i,j}^t|. \tag{17}$$

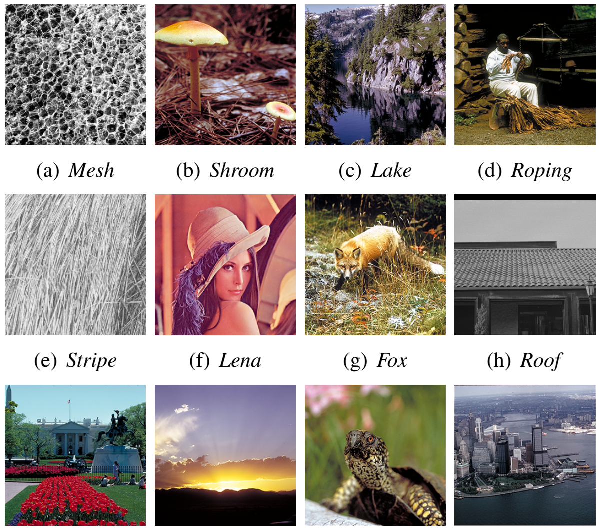
Subsequently, (12) can be further derived by incorporating (14) as follows,

$$
\begin{aligned}
&|p_{t+1,i_{min}} - p_{t,i}| = \left| \frac{N_{t,i} - \sqrt{2|\sum\limits_{j=1}^{n} X_t(i,j)\Delta_{i,j}^t|}}{N_t + B_t + ||\Delta^t||_1} - \frac{N_{t,i}}{N_t} \right| \\
&\leq \frac{\sqrt{2\max\limits_{1\leq j\leq n}\{|\Delta_{i,j}^t|\}\sum\limits_{j=1}^{n}|X_t(i,j)|}}{N_t} + \frac{||\Delta^t||_1}{N_t} + \frac{B_t}{N_t} \\
&\leq \frac{\sqrt{2||\Delta^t||_1\sum_{j=1}^{n}|X_t(i,j)|}}{N_t} + \frac{||\Delta^t||_1}{N_t} \\
&\quad + \frac{\sum\limits_{k\in A_t}\sqrt{2\sum_{j=1}^{n}|X_t(k,j)|||\Delta^t||_1}}{N_t}.
\end{aligned}
\tag{18}
$$

Assuming that $\lim_{t\to\infty}||\Delta^t||_1 = 0$, the conclusion in (12) can be finally obtained. This assumption seems reasonable because the r_t and r_{t+1} are closer to zero when the iteration number increases, and the experimental validation for supporting this assumption will be given in Section IV.

IV. Experimental Results

To evaluate and verify the improved EoP, common natural images in most image processing applications are used in this work, and some examples are shown in Fig. 1.



(a) *Mesh* (b) *Shroom* (c) *Lake* (d) *Roping*
(e) *Stripe* (f) *Lena* (g) *Fox* (h) *Roof*
(i) *Flower* (j) *Sunset* (k) *Turtle* (l) *City*

Fig. 1. Examples of test images.

Firstly, we verify the assumption of $\lim_{t\to\infty}||\Delta^t||_1 = 0$ when deriving (18) by measuring the residual energy of each iteration, i.e. $\frac{||r_t||_F^2}{N_t}$, where r_t indicate the residual matrix in t^{th} iteration. The corresponding results are shown in Fig 2, from which one can observe that the residual energy approaches to zero after several times of iteration, and it can provide strong support for this assumption.

Put it more formally, we have the following equation,

$$DX_t + r_t = DX_{t+1} + r_{t+1}. \tag{19}$$

Then we can derive,

$$D(X_{t+1} - X_t) = r_t - r_{t+1}, \tag{20}$$

$$\Delta^t = X_{t+1} - X_t = D^{\dagger}(r_t - r_{t+1}). \tag{21}$$

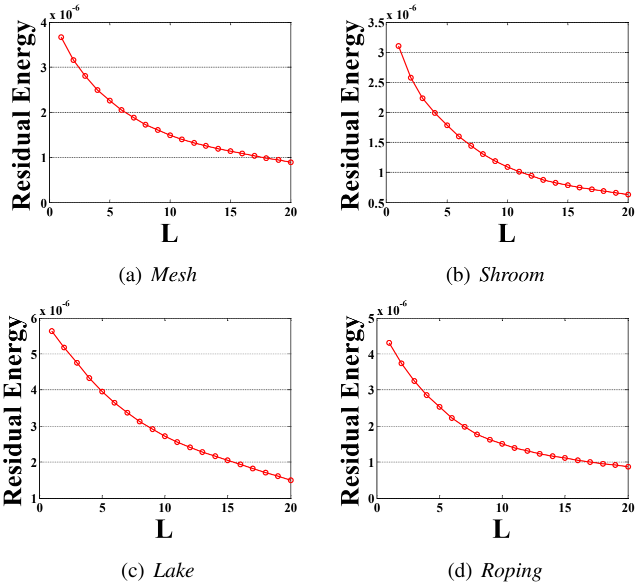
Therefore, it can be concluded that $\lim_{t\to\infty}||\Delta^t||_1 = 0$ since the residuals tend to zero.

Secondly, to further demonstrate the robustness and convergency of *I-EoP*, we employ the KL divergence to measure the difference between two PDFs of successive iterations as defined in (5). The results are visualized in Fig. 3, where the KL divergence reaches close to zero after several iterations indicating the final convergence of *I-EoP*.

Finally, the comparisons between original EoP and improved EoP are given in Fig. 4, from which one can observe that the improved one achieves more stable and robust convergence tendencies against original EoP, which is more consistent with human eyes.

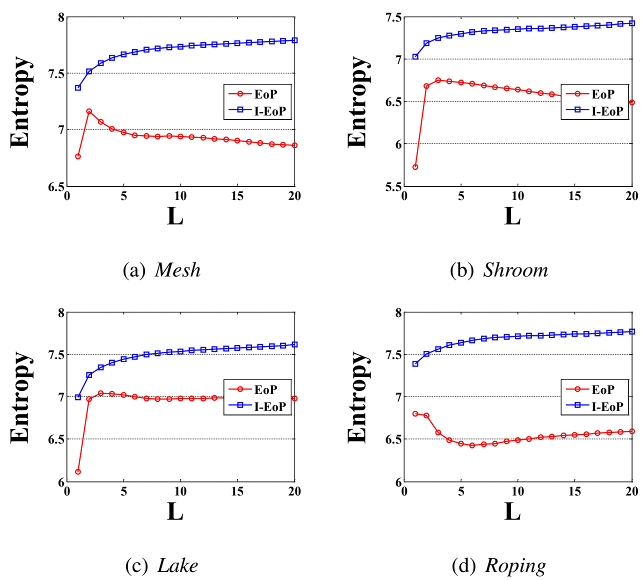
V. Conclusion

In this work, an improved version of EoP is proposed for visual information estimation. The motivation is that the original EoP definition takes no consideration of the amplitude of sparse coefficient values and its convergence tendency sometimes gets worse for complex natural scenes with unnaturalness. Therefore, in the improved EoP, we propose to



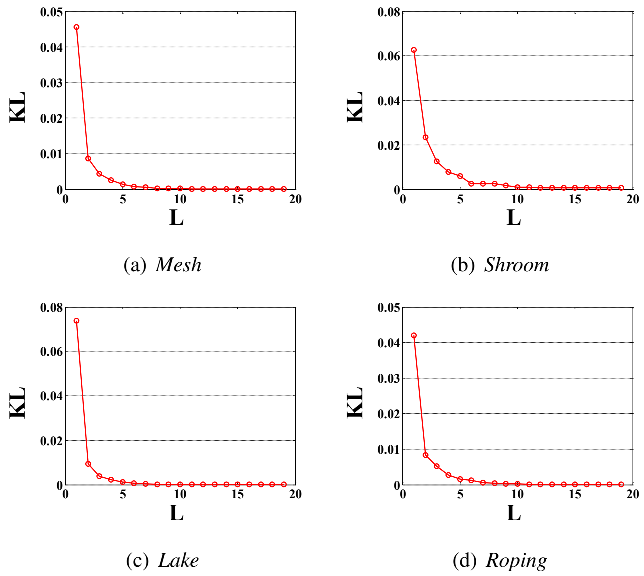
(a) *Mesh* (b) *Shroom* (c) *Lake* (d) *Roping*

Fig. 2. Residual energy in terms of iteration times.



(a) *Mesh* (b) *Shroom* (c) *Lake* (d) *Roping*

Fig. 4. Comparisons between original EoP and improved EoP.



(a) *Mesh* (b) *Shroom* (c) *Lake* (d) *Roping*

Fig. 3. KL divergence between PDFs of two successive iterations.

use the ℓ_2 norm instead of ℓ_0 norm, and its convergence performance and robustness are guaranteed by both mathematical derivations and experimental verifications.

Acknowledgment

This work was supported in part by the National Basic Research Program of China (973 Program, 2015CB351800), and National Natural Science Foundation of China (61322106, 61421062), and Shenzhen Peacock Plan, which are gratefully acknowledged.

References

[1] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[3] J. Zhang, C. Zhao, R. Xiong, S. Ma, and D. Zhao, "Image super-resolution via dual-dictionary learning and sparse representation," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, 2012, pp. 1688–1691.

[4] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, 2014.

[5] J. Zhang, C. Zhao, D. Zhao, and W. Gao, "Image compressive sensing recovery using adaptively learned sparsifying basis via l0 minimization," *Signal Processing*, vol. 103, pp. 114–126, 2014.

[6] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, 2014.

[7] C. Hua-Wen, Y. Hua, G. Yong, and W. Ming-Hui, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 4007–4018, 2013.

[8] X. Zhang, S. Wang, K. Gu, T. Jiang, S. Ma, and W. Gao, "Sparse structural similarity for objective image quality assessment," in *IEEE Int. Conf. Systems, Man, and Cybernetics*, 2015.

[9] T. Guha, E. Nezhadarya, and R. Ward, "Learning sparse models for image quality assessment," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 151–155.

[10] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

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

[12] X. Zhang, S. Wang, S. Ma, and W. Gao, "Towards accurate visual information estimation with entropy of primitive," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, 2015.

[13] S. Ma, X. Zhang, S. Wang, H. Sun, and W. Gao, "Entropy of primitive: From sparse representation to visual information evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, 2016.

[14] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[15] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.