A SEMANTIC NO-REFERENCE IMAGE SHARPNESS METRIC BASED ON TOP-DOWN AND BOTTOM-UP SALIENCY MAP MODELING

Sheng-hua Zhong, Yan Liu, Yang Liu and Fu-lai Chung

Department of Computing
The Hong Kong Polytechnic University

ABSTRACT

This work presents a semantic level no-reference image sharpness/blurriness metric under the guidance of top-down & bottom-up saliency map, which is learned based on eye-tracking data by SVM. Unlike existing metrics focused on measuring the blurriness in vision level, our metric more concerns about the image content and human's intention. We integrate visual features, center priority, and semantic meaning from tag information to learn a top-down & bottom-up saliency model based on the eye-tracking data. Empirical validations on standard dataset demonstrate the effectiveness of the proposed model and metric.

Index Terms—Image quality assessment, Top-down & bottom-up saliency map, No-reference

1. INTRODUCTION

Assessing the quality of images automatically in agree with human's judgment, is desirable in various applications such as image compression and enhancement. Objective image assessment can be divided into three categories: full-reference, no-reference (blind quality assessment), and reduced-reference. Full-reference quality assessment assumes that a complete reference image is known. In many practical applications, however, the reference image cannot be available, and a no-reference approach is desirable. Most existing no-reference quality assessment techniques focus on measuring the sharpness of the images [1-6]. Rony Ferzli and Lina J. Karam divided the images into 64*64 blocks, which are corresponded to the foeval region of Human Visual System (HVS). They judge the blurriness of the image using classical Just Noticeable Blur (JNB) model based on the spread of the edges in these local areas, which is calculated according to contrast measures [4]. In [5], more low-level visual features, such as intensity, color and orientation information are considered in the saliency weighted JNB to accentuate the blur distortions. In [1], a locally-adaptive iterative edge refinement algorithm is proposed based on the classical JNB model to more reliably detect edges in highly blurred images.

However, all these metrics only assess the quality of the image in the visual level although it is widely known that cognitive understanding influences the perceived quality of the images. For example, in Fig.1, (a) is the image without distortion; (b) and (c) include blurriness in different areas. If the user is more concerned about the girl, (c) is considered as the image with better quality. But if the user is more concerned about the apple, (b) will be the better one. Obviously, prior information regarding the image content, or human's attention, may also affect the evaluation of the image quality [7]. But most of image quality metrics do not consider these effects, as they are difficult to quantify and not well understood.

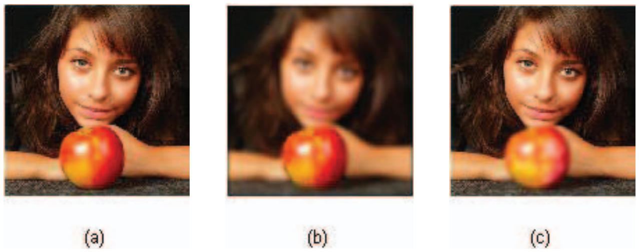

Fig. 1. Example of images quality influenced by the tag. (a) The image without distortion. (b) Blurriness mainly on the girl. (c) Blurriness mainly on the apple.

To address this problem, we propose a semantic image sharpness metric with the aid of rich tag information from Internet. Now, many web applications, such as Flickr, allow users to upload photos with their own annotated tags, which generally indicate the objects users concerned or the targets they took photos. To integrate tag information, this paper utilizes a new saliency map model based on both bottom-up & top-down schemes. Currently, most saliency map, which calculates the likelihood of a location to attract attention, is built based on bottom-up computational model, i.e., from low-level visual features to high-level judgment of saliency. However, this bottom-up scheme only measures the conspicuity of the image in visual level. In this paper, we also consider the image saliency in semantic level, which at least can be partially indicated by the tags of the image annotated by humans. Hence, top-down scheme is proposed to map human's intention to the saliency of each pixel. Support vector machine is used to learn the contributions to human's attention from visual aspect and semantic aspect based on eye-tracking information

2. PROPOSED METRIC

Fig. 2 shows a block diagram of the proposed sharpness metric process. Firstly, WordNet [8] is used to get the target information according to the tag. Then a top-down & bottom-up saliency model is learnt by eye-tracking data. Third, the saliency regions are calculated based on the target information, visual information, and the saliency map. Guided by the saliency information, we use computer edge block distortion to assess the image quality.

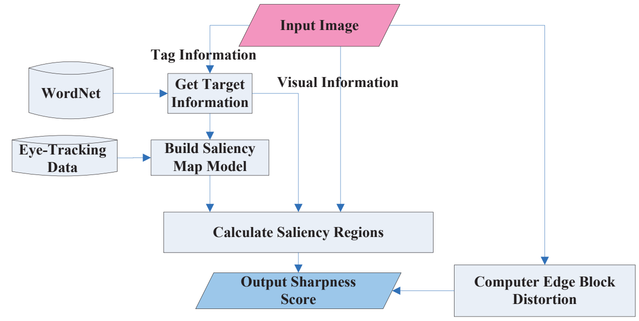

Fig. 2. Flowchart illustrating of the proposed image sharpness assessment metric.

2.1. Target information acquisition

Firstly, we acquire target information of the image with the help of corresponding tags. To avoid the interference of irrelevant or trivial tags, we should use a lexicon to remove all tags that do not belong to the 'physical entity' group. Due to the function of finding hypernym of words, WordNet is chosen as the lexicon. So the tag information automatically be transformed to the target information and then used to calculate the saliency regions.

2.2. Top-down & bottom-up saliency map model

After determining the target of the image in 2.1, we utilize a top-down & bottom-up saliency model to detect the saliency regions which affect the quality. Saliency map first appeared in [9]. Typically, multiple low-level visual features such as intensity, color, orientation, texture and motion are extracted at multiple scales. After a feature map is computed for each of the features, they are normalized and combined into a master saliency map that represents the saliency of each pixel. However, this kind of bottom-up saliency map is limited to represent semantic information. Based on the eye tracking data collecting from 15 viewers on 1003 images in [10], bottom-up saliency model does not match actual eye movements. In Fig. 3, (b) shows the eye movements of humans to watch the image in (a). And (c) shows the eye fixation points covered by the saliency regions based on bottom-up saliency model while (d) shows the result from our top-down and bottom-up model. Obviously, our model covers most eye fixation points.

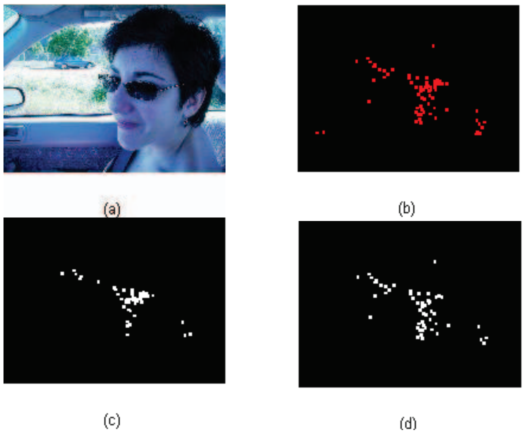

Fig. 3. Examples of bottom-up vs. our top-down & bottom-up saliency model judge results. (a) Original image. (b) Eye-tracking locations. (c) Eye fixation points covered by bottom-up saliency model [5]. (d) Eye fixation points covered by our saliency model.

The flowchart of top-down and bottom-up saliency map modeling is shown in Fig. 4. Eye-tracking data includes three kinds of information: tag, visual information, and eye fixation points. Target information is acquired from image tag by the method mentioned in section 2.1. Then we search the targets in the image according to the visual information. According to the research in [10], we know that the center priority is an important feature to represent semantics of the image because human photographers tending to place objects of interest in the center of photographs. We also extract low-level visual features using Itti bottom-up saliency model [11].

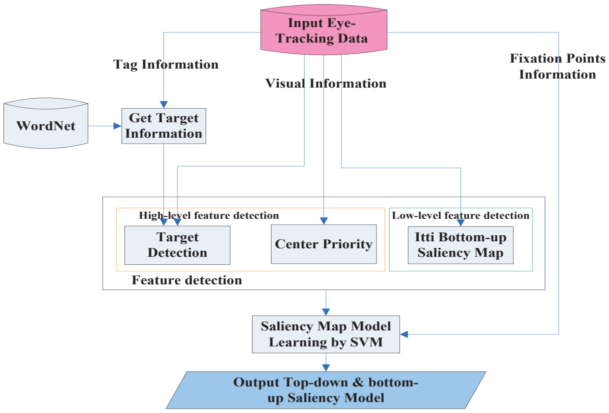
Input Eye-Tracking Data
Tag Information
Visual Information
Fixation Points Information
WordNet
Get Target Information
High-level feature detection
Target Detection
Center Priority
Low-level feature detection
Itti Bottom-up Saliency Map
Feature detection
Saliency Map Model Learning by SVM
Output Top-down & bottom-up Saliency Model

Fig. 4. Flowchart illustrating of the proposed top-down & bottom-up model algorithm.

To determine the contributions to the human's attention from target, center priority, and low-level visual features, SVM is used based on the eye fixation points. To choose the positively and negatively labeled pixels for saliency model learning, we build a ground truth map according with the contrast sensitivity research [12]. The function of contrast sensitivity as a function of pixel position (x,y) is given by

$$S_c = \frac{e_2 \ln(\frac{1}{CT_0})}{\alpha[e_2 + \tan^{-1}(\frac{d(x,y)}{Lv})]} \quad (1)$$

Where α is spatial frequency decay constant; e_2 is half resolution eccentricity constant (degrees); L is the image width (measured in pixels), v is the viewing distance from viewer to computer screen (measured in image width). $d(x,y)$ is the distance from (x,y) to the fixation point. After removing the constant which doesn't affect the saliency sequence of locations, the ground truth map I is created by convolving the function of contrast sensitivity over the top N fixation locations for all M users.

$$g(x,y) = \sum_{i=1}^{M}\sum_{j=1}^{N}\delta_{i,j}(u - f_x(i,j), v - f_y(i,j)) \otimes \frac{1}{e_2 + \tan^{-1}(\frac{d(x-u,y-v)}{Lv})} \quad (2)$$

$$I(x,y) = g(x,y) / \max_{x,y}(g(x,y)) \quad (3)$$

$$d(x,y) = \sqrt{x^2 + y^2} \quad (4)$$

Fig. 5 (a) demonstrates one example of contrast sensitivity model, where brightness indicates the normalized strength of contrast sensitivity. Fig. 5 (b) shows the ground truth map, which measures the salient degree of every pixel of image shown in Fig. 3 (a) according to the eye fixation points shown in Fig. 3 (b).

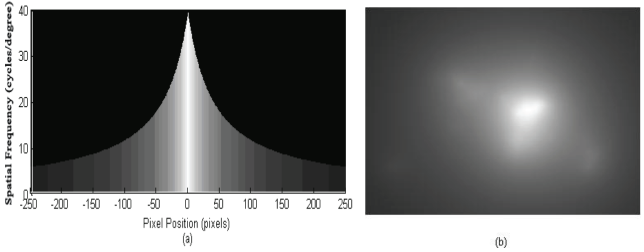

Fig. 5. (a) Contrast sensitivity (brightness indicates the strength of contrast sensitivity) for L=1024 and viewing distance $v = 1$. (b) Ground truth map is found by convolving a cutoff frequency function over the fixation locations of Fig. 3 (b) where M =15, N = 6, $e_2 = 2.3$, $v = 2$ based on [10][12].

After generating the ground truth map, we choose the saliency locations as positively labeled pixels and the non-saliency locations as negatively labeled ones to learn the top-down & bottom-up saliency model by SVM. Then the model is utilized to find saliency regions in blurred images.

2.3. Saliency guidance combined with JNB

In this part, we combine the saliency guidance to measure the blurriness of image. Firstly, the input image is divided into 64x64 blocks. Then the JNB metric [2] is used to calculate the blurriness in every local edge block. The perceived blur distortion within an edge R_b is given by

$$D_{R_b} = (\sum_{e_i \in R_b} \left| \frac{W(e_i)}{W_{JNB}(e_i)} \right|^{\beta})^{\frac{1}{\beta}} \quad (5)$$

where β, which sets as 3.6 in [2], is chosen to increase the correspondence of (5) with the experimentally determined psychometric function. $W(e_i)$ is the measured width of the edge and $W_{JNB}(e_i)$ is the JNB width which depends on the local contrast C which is defined as the magnitude of the difference between the maximum and minimum intensities. W_{JNB} is modeled as follows:

$$W_{JNB} = \begin{cases} 5 & C \le 50 \\ 3 & C > 50 \end{cases} \quad (6)$$

Then the saliency guidance combined with blurriness in each block is utilized to assess the image quality. Blur distortion in saliency regions D_s and non-saliency regions D_{ns} are defined as (7), where R_{bs} is the salient part in R_b, and $Num(\cdot)$ is to calculate the number of pixels in corresponding region.

$$D = \begin{cases} D_s = (\sum_{R_b} |D_{R_b}|^{\beta})^{\frac{1}{\beta}} & Num(R_{bs}) / Num(R_b) > \alpha_{thresh} \\ D_{ns} = (\sum_{R_b} |D_{R_b}|^{\beta})^{\frac{1}{\beta}} & Num(R_{bs}) / Num(R_b) \le \alpha_{thresh} \end{cases} \quad (7)$$

The proposed objective sharpness metric is given by

$$S = \alpha_s \cdot (\frac{L_s}{D_s}) + (1 - \alpha_s) \cdot (\frac{L_{ns}}{D_{ns}}) \quad (8)$$

where L_s, L_{ns} are the numbers of saliency blocks and non-saliency blocks in the image. α_s is the weight of saliency part to the sharpness metric.

3. SIMULATION RESULTS

In this section, we conduct two experiments. The first experiment is used to demonstrate the effectiveness of the proposed top-down & bottom-up saliency map modeling technique. We choose 200 training images and 64 testing images which contain person from standard eye-tracking dataset [10]. In every training image, we randomly choose 30 saliency pixels as positively labeled data from the 10% most salient locations and 30 non-saliency pixels as negatively labeled data from the 10% least salient locations. The experiment results show that 84.156% saliency points detected by our model are inside and 76.344% non-saliency points are outside the saliency regions judged by human. This performance is better than the results from existing saliency map modeling methods [10]. Fig. 6 compares the different models to determine the salient region. Our model covers most informative area of the image.

The second experiment demonstrates that the proposed image assessment algorithm based on top-down & bottom-up saliency map outperforms the representative blurriness metrics of classical JNB, saliency weighted JNB, and JNB with edge refinement. For blur image quality evaluating, we use the real online dataset Flickr. Due to the page limitation, we only demo the images with the tag "person", because more than 4 million images in Flickr use this tag. We randomly select 160 images with the tag or the hypernym of tag of 'person'. We averagely partition the images into eight groups blurred with eight different 7x7 Gaussian masks of σ values 0.8, 1.6, 2.0, 2.4, 3.2, 4.0, 4.8 and 5.6. For each displayed image, fourteen subjects are asked to rate the quality of the images in terms of perceived blurriness using a scale from 1 to 5 corresponding to "Very annoying", "Annoying"', "Slightly Annoying", "Perceptible but not annoying", and "Imperceptible", respectively. We provide the correlation analysis between the objective measures and the mean opinion scores (MOS). In our model, we set the parameters $\alpha_{thresh} = 1/3$ and $\alpha_s = 0.8, \alpha_{ns} = 0.2$. Table 1 shows the comparison of different sharpness metrics in terms of the Pearson (indicates the prediction accuracy), Spearman (indicates the prediction monotonicity), MAE (mean absolute prediction error) and RMS (root mean squared error) coefficients after nonlinear regression. Obviously, our technique has better results under most cases. Gerneally speaking, intergrating the saliency model will improve the proformance. However, if the saliency model doesn't consist with the real fixation points, it may hurt the proformance. It is why the performance of Saliency Weighted JNB [5] and JNB with Edge Refinement [1] is even lower than classical JNB [2] in Table 1.

4. CONCLUSION

In this paper, a semantic level no-reference image quality assessment metric is proposed. It utilizes the top-down & bottom-up saliency map to detect the possible saliency regions in blurred images, which is learned based on eye-tracking data by SVM. This metric exhibits increased correlation with perceived quality. In future, we will apply the proposed techniques to different kinds of images and extend current algorithm to the images with multiple tags.

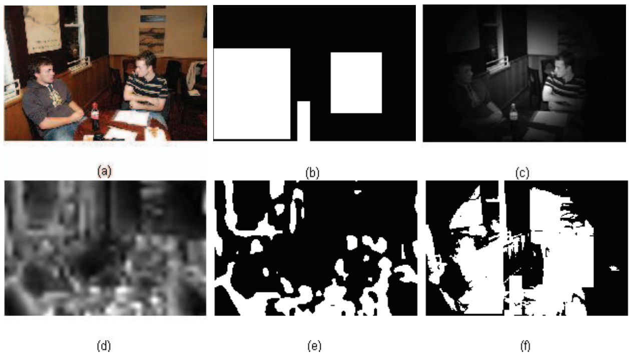

Fig. 6. A sample image with tag "person". (a) Original image. (b) Person detection [13]. (c) Center prior detection. (d) Bottom-up saliency map detected by [11]. (e) Bottom-up saliency regions which value is over the mean value. (f) Top-down & bottom-up saliency map.

Table 1. Evaluation of the proposed metric performance.

	Nonlinear Pearson	Spearman	MAE	RMS
Proposed Metric	**0.914**	**0.86**	**0.173**	0.25
Classical JNB [2]	0.885	0.815	0.219	0.292
Saliency Weighted JNB [5]	0.863	0.801	0.317	**0.232**
JNB with Edge Refinement [1]	0.618	0.466	0.387	0.494

5. REFERENCES

[1] Srenivas Varadarajan and Lina J. Karam, "An Improved Perception-Based No-Reference Objective Image Sharpness Metric using Iterative Edge Refinement," IEEE International Conference on Image Processing, pp. 401-404, Oct. 2008.

[2] Rony Ferzli and Lina J. Karam, "A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)," IEEE Transactions on Image Processing, Vol. 18, No. 4, pp. 717-728, Apr.2009.

[3] Rony Ferzli and Lina J. Karam, "A No-Reference Objective Image Sharpness Metric Based on Just Noticeable Blur and Probability Summation," IEEE International Conference on Image Processing, Vol. 3, pp. 445-448, Sept. 2007.

[4] Rony Ferzli and Lina J. Karam, "A Human Visual System Based No-Reference Objective Image Sharpness Metric," IEEE International Conference on Image Processing, pp. 2949-2952, Oct. 2006.

[5] Nabil G. Sadaka, Lina J. Karam, Rony Ferzli, and Glen P. Abousleman, "A No-Reference Perceptual Image Sharpness Metric Based on Saliency-Weighted Foveal Pooling," IEEE International Conference on Image Processing, pp. 369-372, Oct. 2008.

[6] Luhong Liang, D. Jianhua Chen, Siwei Man, Debin Zhao and Wen Gao, "A no-reference perceptual blur metric using histogram of gradient profile sharpness, pp.4369-4372. Apr.2009.

[7] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600-612, Apr. 2004.

[8] G. Miller. WordNet: A Lexical Database for English. Communications of the ACM, 1995.

[9] Koch, C. & Ullman, S., "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," Human Neurobiology. Vol. 4, No. 4, pp. 219-227, 1985.

[10] Tilke Judd, Krista Ehinger, Fr´edo Durand and Antonio Torralba, "Learning to Predict Where Humans Look," IEEE International Conference on Computer Vision, Sep. 2009.

[11] L. Itti, C. Koch, & E. Niebur, "A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention," Vision Research, Vol. 40, pp. 1489-1506, Apr. 2000.

[12] Zhou Wang and Alan Conrad Bovik, "Embedded Foveation Image Coding,"IEEE Transactions of Image Processing, Vol. 10, No.10, Oct. 2001.

[13] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A Discriminatively Trained, Multiscale, Deformable Part Model," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-8. June 2008.