TEMPORAL POOLING OF VIDEO QUALITY ESTIMATES USING PERCEPTUAL MOTION MODELS

Kwanghyun Lee†, Jincheol Park†, Sanghoon Lee†and Alan C. Bovik‡

†Wireless Network Lab., Center for IT of Yonsei University, Seoul, Korea, 120-749.
‡Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX

ABSTRACT

Emerging multimedia applications have increased the need for video quality measurement. Motion is critical to this task, but is complicated owing to a variety of object movements and movement of the camera. Here, we categorize the various motion situations and deploy appropriate perceptual models to each category. We use these models to create a new approach to objective video quality assessment. Performance evaluation on the Laboratory for Image and Video Engineering (LIVE) Video Quality Database shows competitive performance compared to the leading contemporary VQA algorithms.

1. INTRODUCTION

In parallel with the development in various multimedia applications, there has been considerable progress in developing objective image and video quality metrics designed to evaluate visual quality in agreement with subjective human judgments. Of course, for most applications, subjective quality assessment by humans is the most accurate method, but coordinating subjective studies requires considerable time, cost and human effort that is unsatisfactory except in rare instances other than for bench-marking quality assessment algorithms. Thus, a variety of fidelity measurement methods have been proposed to approximate the result of subjective quality measurement, including the hoary peak signal to noise ratio (PSNR), the structural similarity index (SSIM) and the Visual Information Fidelity index (VIF), video quality metric (VQM) etc [1].

The two terms, fidelity and quality, are often used interchangeably, but they are not always the same. Strictly, a fidelity measurement discriminates similarity of pixel values between two images, while a quality assessment evaluates the preference for one image sequence over another. Fidelity scores may fail to reflect accurately subjective observation since it may not reflect perceptual characteristics. Pooling fidelity scores to a representative quality score using perceptual motion models is the approach to the problem. We suggest such a framework of temporal pooling. We categorize each frame into stationary and moving scenes. According to the scene categorization, perceptual weights are defined and applied to the fidelity scores at both the frame level (FL) and sequence level (SL).

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the national HRD support program for convergence information technology supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2010-C6150-1001-0013)

2. COMPREHENSION OF MOTION PERCEPTION

Motion analysis is complicated by the fact that not only do objects in the environment move, but observers move as well [2]. For simplicity, we categorize moving images into three cases:

Stationary scene: Observer (camera) remains stationary and objects move.

Moving scene A: Objects move and camera follows a specific object.

Moving scene B: Objects are motionless and a camera moves in a stationary environment (i.e. panning).

We propose that for VQA, camera movements can be detected by statistics of the motion vector (MVs). When the camera is not moving, motions occur locally, which modifies the statistics.

Human eye is naturally attracted to moving objects. The authors of [4] illustrate a visual attention model attracted by objects in motion. It is assumed that moving objects are significant to the HVS relative to those of background with inducing more attention. Using the prior probability distribution in terms of the speed of motion, self-information, I, is formulated as the information entropy. This perceptual model can be applied to the stationary scene and the moving scene A. However, when the apparent velocity becomes too high in the stationary scene, it rather becomes harder to detect and distinguish object features such as contours and texture. In [5], the contrast sensitivity of a drifting bar linearly increases up to a threshold of velocity, then dramatically falls. In addition, as the width of the drifting bar objects increased (ranging from 1 to 80 degrees), it peak visibility occurred with increasing velocity. Likewise, the visual sensitivity is higher for smaller objects at lower velocities. In view of this, we propose the following contrast sensitivity model as a function of object

velocity and object width:

$$C(v,\, M_{deg}) \;=\; C_1 \left(M_{deg} \,+\, k_1\right)^{C_2} e^{-\left(v/k_2\right)^2}, \quad (1)$$

where v is the object velocity $(\deg / \sec)$ and M_{deg} is the object width $(\deg)$, where it is assumed that $v > 0$. M_{deg} is obtained by moving objects segmentation using refined MVs. We applied this model to the curves given in [5] and found that the following constants yield good fits to the behavior in [5]: $C_1 = 40{-}9.1 \ln\, M_{deg}, k_1 = 0.01, C_2 = \frac{1}{3+0.18 \ln\, M_{deg}}$ and $k_2 = 6.75 M_{deg}^2 + 48 M_{deg} - 4.75$. These fits are inexact, since the actual numerical data samples are no longer available, however, we believe that the fit is adequate to allow for a general accounting of visual motion importance.

In the moving scene, a global motion occurs according to the movement of a camera. It is difficult to perceive frames of the moving scene when the camera movement is too fast. In [4], this appearance is modelled as a noise of the distortion communication channel in the HVS or as a likelihood function of the noisy measurement, determining the perceptual uncertainty, U.

3. FRAMEWORK OF POOLING STRATEGY

3.1. Frame Level Pooling

In our proposed pooling strategy, we implement FL Pooling by applying FL motion weights to the lowest $p\%$ of the fidelity scores, and by applying a lower constant scaling factor r to the rest of the scores, similar to the percentile scoring in [3]. Fidelity scores, Q_m, are obtained from each m^{th} window in a frame. The fidelity scores are combined to yield the quality score for the f^{th} frame, s_f, as below:

$$s_f \;=\; \frac{\sum_{m\in\mathbf{P}^c} r \cdot Q_m + \sum_{m\in\mathbf{P}} w_m\left(\theta\right) \cdot Q_m}{\left|\mathbf{P}^c\right| \cdot r + \sum_{m\in\mathbf{P}} w_m\left(\theta\right)}, \quad (2)$$

where

$$w_m(\theta) \;=\; (1-\theta)\, w_m^S + \theta w_m^M, \quad (3)$$

and

$$\theta = \begin{cases} 0, & \frac{\mu_f}{\sigma_f} \le 1 \\ 1, & \frac{\mu_f}{\sigma_f} > 1 \end{cases}, \quad (4)$$

where $\mathbf{P}$ is the set of the lowest $p\%$ fidelity scores, $\mathbf{P}^c$ is the complement of $\mathbf{P}$ and $\left|\mathbf{P}^c\right|$ is the number of elements in $\mathbf{P}^c$. Here w_m is the FL motion weight of the m^{th} window and θ is a threshold used to identify the scene movement detection. If $\mu_f \le \sigma_f$, the current scene is assumed to be a stationary scene, and if $\mu_f > \sigma_f$, the current scene is a moving scene.

The FL motion weight of the moving scene is defined by normalizing the self-information of moving objects:

$$w_m^M = \kappa^{-1} \cdot I\left(v_m^r\right) \quad (5)$$

where v_m^r is the motion speed of the m^{th} sampling window calculated as relative motion [4] and $\kappa = \sum_n I\left(\bar{v}_n^r\right)$ is a normalization parameter.

By combining the motion contrast sensitivity with the self-information, the FL motion weight of the stationary scene is defined as below:

$$w_m^S = \kappa^{-1} \cdot I\left(v_m\right) \cdot C\left(v_m, M_{deg}^m\right) \quad (6)$$

where v_m is the motion speed of the m^{th} sampling window and M_{deg}^m is the width of an object including the m^{th} sampling window.

3.2. Sequence Level Pooling

When the video quality is time-varying, there is an 'negative-peak and duration-neglect effect' [6]. It implies that the important thing for overall subjective quality judgement of a video sequence is not the duration of a dip in quality, but rather the depth. Thus, we also calculate the final score, S, by means of the percentile scoring concept as below:

$$S = \frac{\sum_{f\in\widehat{\mathbf{P}}^c} \widehat{r} \cdot s_f + \sum_{f\in\widehat{\mathbf{P}}} \widehat{w}_f\left(\theta\right) \cdot s_f}{\left|\widehat{\mathbf{P}}^c\right| \cdot \widehat{r} + \sum_{f\in\widehat{\mathbf{P}}} \widehat{w}_f\left(\theta\right)}, \quad (7)$$

where

$$\widehat{w}_f(\theta) \;=\; (1-\theta)\, \widehat{w}_f^S + \theta \widehat{w}_f^M, \quad (8)$$

where $\widehat{\mathbf{P}}$ is the set of the lowest $p\%$ of s_f, $\widehat{\mathbf{P}}^c$ is the complement of $\widehat{\mathbf{P}}$, $|\widehat{\mathbf{P}}^c|$ is the cardinality of $\widehat{\mathbf{P}}^c$ and $\widehat{r}$ is a scaling factor.

The SL motion weight of the moving scene is defined as below:

$$\widehat{w}_f^M = 1 - \frac{U\left(\bar{v}_f\right) - U_{Min}}{U_{Max} - U_{Min}}, \quad (9)$$

where U_{Max} and U_{Min} are the maximum and minimum values of the uncertainty model, and $\bar{v}_f$ is the mean speed of the MVs in the f^{th} frame. In a stationary scene, there is little overall motion effecting human perception because there is no scene movement. Thus we set the visual weight of the SL motion weight of stationary scene be $\widehat{w}_f^S = 1$.

4. PERFORMANCE EVALUATION

We evaluated performance of the proposed method on the Laboratory for Image and Video Engineering (LIVE) Video Quality (VQ) database [7] and compared with several other video quality assessment (VQA) algorithms. Figure 1 shows the scatter plot of the proposed pooling method using SSIM to measure the fidelity of each window. The sampling windows 16×16 non-overlapped blocks.

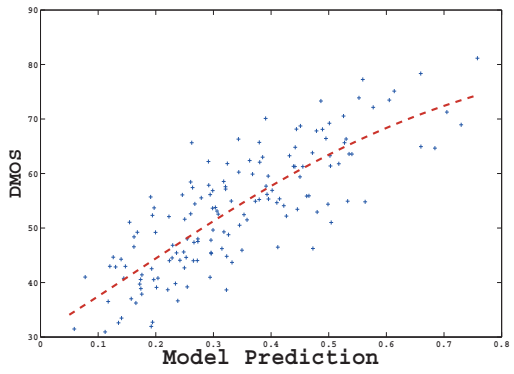

Fig. 1. Scatter plot of the proposed scheme versus DMOS for all videos in the LIVE database

The Spearman Rank Order Correlation coefficient (SROCC) and Pearson Linear Correlation Coefficient (LCC) are used as a metric for performance evaluation. In Fig. 2 and Fig. 3, the SROCC and the LCC of M-Pooling are compared with the several VQA algorithms in [7] for each distortion category. The results show that the performance of the proposed scheme is excellent throughout all the distortion types. In addition it can be seen that the performance of the M-Pooling is more stable than the other VQA algorithms in that the deviation of the SROCC and LCC values are relatively small.

5. CONCLUSION

A new framework of quality pooling strategy has been introduced for video quality assessment using motion perception characteristics. We designed the non-linear pooling strategy based on the motion situations, instead of a simple linear weighted mean method. Thus the motion perception models of psychophysics can be properly adopted on respective motion situations. In the results of the performance evaluation, the proposed scheme shows competitive performance throughout the most distortion categories. Although the MOV IE is better in several categories, the proposed scheme is still competitive comparing to the other VQA algorithms. Above all, the SROCC and the LCC values of the proposed scheme are evenly high over all distortion categories, while the performances of the others are uneven. Therefore, it can be concluded that the proposed scheme functions well for all of the distortion types, whereas the other VQA algorithms have distortion types whose quality is not well measured.

6. REFERENCES

[1] A.C. Bovik (Ed.), *The Handbook of Image and Video Processing*, Academic Press, 2005.

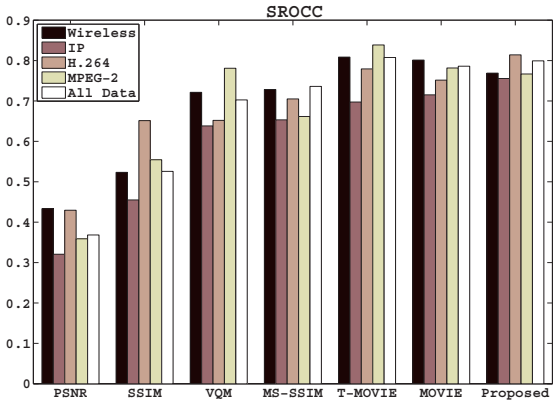

Fig. 2. SROCC comparison with the leading contemporary VQA algorithms on LIVE database

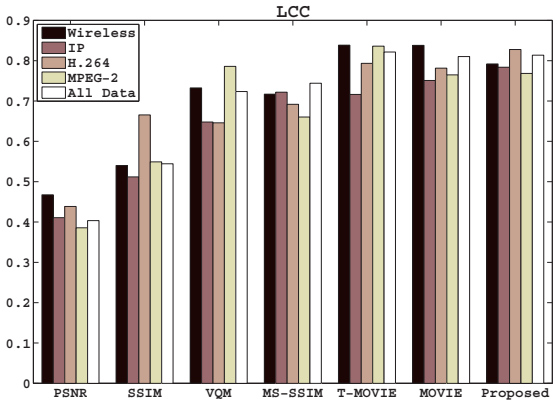

Fig. 3. SROCC comparison with the leading contemporary VQA algorithms on LIVE database

[2] E. B. Goldstein, *Sensation and perception*, Thomson Wadsworth, 2007, pp. 195-213.

[3] A. K. Moorthy and A. C. Bovik "Visual importance pooling for image quality assessment, *IEEE J. Special Topics in Signal* , vol.3, no.2, pp.193-201, Apr. 2009.

[4] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception *Journal of the Optical Society of America A*, vol. 24, no. 12, pp. B61-B69, Dec. 2007.

[5] D.C. Burr and J. Ross, "Contrast sensitivity at high velocities, *Vision Research,* no. 22, pp. 479-484, 1982.

[6] D.E. Pearson, "Viewer response to time-varying video quality. *SPIE Conf. Human vision and Electronic Imaging III,* San Jose, CA, 1998.

[7] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video, *IEEE Transactions on Image Processing to appear,* 2009.