PPIQ: A PROBABILISTIC FRAMEWORK FOR IMAGE QUALITY ASSESSMENT

Koohyar Minoo and Truong Q. Nguyen
Electrical and Computer Engineering Department, UCSD, La Jolla, CA 92093
minoo@stanfordalumni.org, nugyent@ece.ucsd.edu

ABSTRACT

In this paper a framework for Image Quality Assessment (IQA) is introduced based on the properties of Receptive Fields (RFs) which are the primary mechanism for detection of visual patterns in the Human Visual System (HVS). The proposed framework offers a probabilistic approach to the perceptual IQA, based on the probability of detecting discrepancies (distortion) between the corresponding features of a test and a reference image. The proposed Probabilistic Perceptual Image Quality (PPIQ) framework facilitates defining specific perceptual metrics for specific applications. To give an example on how the PPIQ framework can be utilized to define an Image Quality Metric (IQM), a sample IQM is introduced based on the properties of simple RFs of the early vision in the HVS. The sample IQM, based on the PPIQ framework, exhibits comparable accuracy to that of the legacy methods in terms of predicting the outcome of subjective image quality experiments.

Index Terms— Full reference image quality metric, perceptual image distortion metric, feature discrepancy detection

1. INTRODUCTION

In this paper a framework for IQM is proposed. This framework is based on the probability of detecting discrepancies between two images in displaying the same visual feature at a given spatial coordinate. A visual feature can exist in one picture to reflect an impression from the natural world (therefore the lack of such a feature would be considered a distortion). Alternatively a feature may exist in one picture due to a modification (distortion) process (e.g. quantization), which is not expected in a natural scene (i.e., the detection of this feature is considered a distortion). In general, a feature can be any visual artifact which is relevant for measuring the quality of an image for a certain application.

As we elaborate in section 2, the proposed PPIQ framework captures the fundamental characteristic of the Human Visual System (HVS) which treats the detection of a feature in a visual stimulus as a random event. Although the concept of probabilistic detection of features has been considered in many literatures for the sub-threshold, or the JND conditions [1]-[3], in Section 3, we argue that certain aspects of conventional error pooling scheme (such as weighted Minkowski summation of errors) need to be revisited for supra-threshold conditions. The treatment of distortion pooling within the PPIQ framework, specifically, distortion pooling across the feature space and the spatial domain, facilitates the inclusion of foveae-weighted distortion [4] and feature importance due to specific requirements of a given application.

It is important to emphasize that the proposed PPIQ framework in this paper is not an IQM, but rather a model for measuring the distance between two images in terms of showing the same set of features. Once a set of relevant features (according to the application of interest) is defined then this framework can be used to define an IQM. To demonstrate the efficiency of the PPIQ framework in practice, Section 4 introduces an IQM based on the detection of a Non-Directional Contrast (NDC) feature. The proposed metric would be called PPIQ-NDC and it inherently factors in the viewing conditions such as the angular resolution of the image and the well established psychophysical properties of the HVS through the concept of RFs.

2. THE PROBABILISTIC METRIC MODEL

Many features and properties of the HVS for perceptual evaluation of an image can be accurately explained by the properties of RFs. In this section we introduce a framework based on the theory of RF which enables us to define a distortion (or similarity) metric for images, relative to a reference image. In general, an RF is a neural connection configuration, where a number of neurons send electrical impulses to the same node (ganglion cell). In our discussion, the RF model consists of a hierarchy of nodes. Each node performs a linear or non-linear transform on the input nodes (filtering) and then a non-linear operator sets the output (neural pulse rate activity) of that node based on the filter output. Our simple RF model is described by a feature extraction transform $T_f(\cdot)$ (not necessarily a linear transform). $T_f(a)$ maps the intensities from the image around location a, to a feature response value $r(f,a)$, which reflects the impulse rate (neural activity) at the RF which corresponds to a desired feature f.

In the PPIQ framework, when a specific feature (RF response) is assessed for the presence of a feature, a hard-limit threshold at $\Delta(f,a)$ is used to set the detection decision to either 0 (not detected) or 1 (detected). The detection of the feature f, at location a in an image is denoted by $Det(f,a)$. To formalize the notation, we assume that for a person with a threshold decision of $\Delta(f,a)$ the detection function can be defined as:

$$Det(f,a)=\begin{cases}1 & r(f,a)>\Delta(f,a)\\ 0 & r(f,a)\leq\Delta(f,a)\end{cases} \tag{1}$$

In (1) $r(f,a)$ is the activity of the RF (corresponding to the desired feature f at location a). Realizing that the value of $\Delta(f,a)$ is different from person to person, a probabilistic measure is chosen to predict the statistical nature for the detection of a given feature amongst all the test subjects (human observers). This probabilistic approach to feature detection is due to the different detection threshold and neural noise levels amongst human test-subjects. We assign a cumulative probability to the detection threshold value ($\Delta(f,a)$) amongst a large number of test subjects as $P_{\Delta(f,a)}(\Delta)=Prob(\Delta(f,a)\leq\Delta)$. Therefore the feature detection in a probabilistic manner can be interpreted as follows:

$$Det(f,a)=\begin{cases}1 & with\ prob= \ P_{\Delta(f,a)}(r(f,a))\\ 0 & with\ prob= \ 1-P_{\Delta(f,a)}(r(f,a))\end{cases} \tag{2}$$

In the PPIQ framework, the only assumption about $P_{\Delta(f,a)}(\Delta)$ is that it is a cumulative probability function, (non-decreasing function which goes from zero to one). This probability indicates what percentage of people would detect the feature f, at the location a. In Section 4 where the PPIQ-NDC metric is proposed, a specific parametric probability function will be presented to model $P_{\Delta(f,a)}(\Delta)$.

In the PPIQ framework, a discrepancy between two images (usually a reference image and a test image) happens when the same feature, f, at the same location a is detectable in one image and not detectable in the other image. A probability can be assigned to this discrepancy-detection in terms of the percentage of people who detect feature f, at the location a in one image, but do not detect the same feature at the same location in the other image. Fig. 1 depicts the probability of the three possible outcomes in detection of feature f at location a. Without loss of generality, in Fig. 1, it is assumed that the feature response at the reference image is smaller than the corresponding value at the test image. The S-shape curve in Fig. 1 represents the cumulative probability of the detection threshold amongst test subjects. It is easy to observe that the

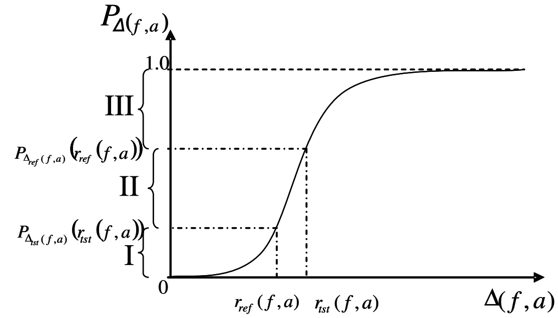

Fig. 1: Region I corresponds to the probability that the feature f would be detected in both images. Region III corresponds to the probability that the feature f would not be detected in either of images. Region II corresponds to the probability that the feature f would be detected in one image but not in the other image.

probability of detecting a discrepancy, according to (2) corresponds to the region II of Fig. 1 as follows:

$$P_{dis}(ref,tst,f,a)= \left|P_{\Delta_{ref}(f,a)}(r_{ref}(f,a))-P_{\Delta_{tst}(f,a)}(r_{tst}(f,a))\right| \tag{3}$$

In (3), $P_{dis}(ref,tst,f,a)$ is the probability of detecting a dissimilarity between the reference and the test images at location a for the feature f. Note that $r_{ref}(f,a)$ and $r_{tst}(f,a)$ are the feature extraction transform responses in the reference and the test image, respectively. Also $P_{\Delta_{ref}(f,a)}(\cdot)$ and $P_{\Delta_{tst}(f,a)}(\cdot)$ are the corresponding cumulative probabilities for detection threshold in the two images.

It is essential to recognize that the error visibility methods such as [5], subtract feature response values in one image from the corresponding values in the second image to measure the distance. This operation suggests that the HVS employs a *perfect photographic memory* which is capable of subtracting the exact values of feature responses from the test and the reference image, acquired at two different time instances. However the definition of distance in the PPIQ framework only assumes a *comparative memory* which relies on a more realistic notion that the only thing a test subject memorizes, is the detection of a given feature at a given location (not the exact value of the feature response in each of the two images).

3. DISTANCE (DISTORTION) POOLING

So far, the PPIQ framework offers a basic distance metric which is the probability of observing a discrepancy in the detection of a given feature at a given location between two images by equation (3). In this section we define a strategy by which, one can assign a single quality value to the entire image. This requires addressing the issues of distance (error) pooling across feature and spatial spaces. The concept of distance pooling has been extensively studied in the linear transform spaces such as DCT and Wavelet. In most of these studies a Minkowski summation (norm-p) or weighted Miknowski summation (weighting performed based on the masking effects or the contrast sensitivity [2] and [5]) is used to assign an overall distance value to a vector of errors across feature space and foveal-spatial space.

The justification for norm-p pooling comes from the concept of probability summation along with extra assumptions that 1- The probability density function for feature detection is exponential and 2- The detection at JND condition depends on the probability of detecting "any" error in the image (or within a foveal region [4]). Although this assumption is reasonable (and matches well with the results of psychophysical studies) for detection of "Just Noticeable Distortion (error)", it falls short of representing the subjective image quality at suprathreshold cases where the distortion is beyond the "just-visibility" and the probability of seeing "any" error is always close to one).

At suprathreshold conditions, where distortion is visible at many locations across the spatial span and feature space, it is a common practice to measure distortion in terms of the average time it takes to detect the first discrepancy between the test and reference images. Fortunately the probabilistic nature of PPIQ framework is facilitating to adopt this scheme of distance (error) pooling. Note that the distance metric in (3) indicates the probability that a given feature at certain location reveals a discrepancy between the two images. In this context, one can describe the act of eye fixation on one image in pursuit of finding a discrepancy or similarity, in terms of showing feature f_i (from a pool of M features) at location a, by a given probability [1], represented by $P_{fix}(f,a)$. It is intuitive to assume that the second fixation on the other image happens with probability 1 (as the test subject tries to find the feature of interest in the same location). Assuming that each fixation takes a constant time duration, it can be shown that the average time to detect the first discrepancy is proportional to the expected value of the probability of discrepancy detection in one fixation (note that expectation is according to the random fixation event). As result, (4) gives the overall distance $D(ref,tst)$, in the PPIQ framework . Note that (4) also affirms the intuitive notion of weighted (according to the probability of fixation) norm-1 distortion pooling which has already been practiced in some of the suprathreshold IQM such as SSIM [6].

$$D(ref,tst) = P_{dis}(ref,tst) = \sum_{a \in image} \sum_{i=1}^{M} P_{dis}(ref,tst,f_i,a) \cdot P_{fix}(f_i,a) \tag{4}$$

We conclude the description of PPIQ framework by giving a four step instruction on how an application can define an IQM based on the PPIQ framework: 1- Choose a set of desired feature extraction filters to produce the feature response $r(f,a)$. 2- Select a suitable cumulative probability function to produce the distortion detection probability $P_{dis}(ref,tst,f,a)$ from (3). 3- Adopting a fixation probability function $P_{fix}(f,a)$ to be used for distortion pooling across spatial and feature spaces as in (4). 4- If the preceding functions are not exactly known, a parameterized version of those functions can be used along with a labeled subjective image quality training set to learn what the unknown parameters should be.

Although the primary concern of this paper is to introduce the PPIQ framework and leave the task of instantiating specific IQMs to the adopters of the proposed PPIQ framework, in the next section we introduce an IQM just to exemplify how the PPIQ framework can be used to derive IQMs, comparable to the contemporary objective metrics, known to be performing superbly in predicting subjective test results.

4. PPIQ-NDC: A PERCEPTUAL IMAGE DISTORTION METRIC

In this section a sample PPIQ metric based on a specific feature extraction and detection threshold probability models is proposed to evaluate the principles based on which the PPIQ framework was proposed. To that end we utilize a feature that corresponds to an omnidirectional contrast filter (step 1 in Section 3) and an exponential detection threshold cumulative probability function [3] (step 2 in Section 3). As for step 3 in Section 3, we note that modeling of $P_{fix}(f,a)$ requires specific knowledge about several factors which can draw viewers' attention. With less knowledge about these elements, the probability distribution for the fixation becomes flatter (i.e. all locations are equally likely). For the sake of simplicity of the proposed example IQM (PPIQ-NDC) we resort to a flat fixation probability function.

The omnidirectional feature extraction filter for PPIQ-NDC is a parameterized, 2-D symmetric Laplacian of Gaussian (LoG) filter as in (5). Note that the LoG shape resembles simple RFs in the early vision [7] and as such it models the combined effect of point spread function of optical aperture and the bandpass property of the HVS.

$$T_{oc}(x,y) = \frac{1}{\pi \cdot \sigma^4}\left(\frac{x^2+y^2}{2\sigma^2} - 1\right)\exp\left[-\left(\frac{x^2+y^2}{2\sigma^2}\right)\right] \tag{5}$$

The parameter σ defines the band-pass characteristics of the RF filter in the spatial frequency domain. It should be noted that σ also defines the spatial span of the RF on the display. Since this number should reflect the actual size of the corresponding RFs on the retina, one can use the value of optimal σ for a given viewing distance and pixel pitch and calculate the optimal σ for a different viewing distance or pixel pitch. The adaptability to viewing conditions is one of the advantages of PPIQ-NDC as a distortion metric. Also the bandpass nature of the filter in (5) matches the results of psychophysical experiments and the concept of contrast sensitivity function (CSF).

Inspired by the probability function which is used for probability summation techniques in [1], a parameterized, exponential cumulative probability function is used for defining PPIQ-NDC as in (6). Δ_{norm} is the value of $r(f,a)$ at which 63% of test-subjects can detect the existence of feature f at location a and β influences the slope factor. Both Δ_{norm} and β parameters, in general, are functions of the feature f and the image content, especially around the location a to reflect the masking effects [5]. However in this paper to be able to have a simple parameterization of the detection probability function, we choose β and Δ_{norm} to be constants parameters.

$$P_{\Delta(f,a)}(r(f,a)) = 1 - \exp\left(-\left[\frac{|r(f,a)|}{\Delta_{norm}}\right]^{\beta}\right) \tag{6}$$

The parameterization of the proposed PPIQ-NDC metric offers an opportunity to examine the properties of the HVS's functionality which correspond to those parameters. To that end we used the subjective test results from the LIVE database [8] to optimize the PPIQ-NDC parameters. The LIVE database includes subjective test results for five categories of distorted images which include JPEG and JPEG 2000 compression distortion, additive white noise and Gaussian blur distortion and finally error concealed JPEG 2000 image distortion due to Fast Fading Rayleigh Channel Error (FFRCE) model. The details of the subjective test experiments can be found in [9]. The parameter optimization for PPIQ-NDC metric was done by minimizing the Root Mean Squared Error (RMSE) of the best logistic regressor to fit the metric value to the subjective test results per recommendations in [10]. Table I shows the RMSE value of PPIQ-NDC, VIF [9], SSIM [6] and MSE for prediction of the realigned DMOS values in the LIVE database [8]. This result shows the robustness of the PPIQ-NDC metric in predicting the subjective quality of images. The optimized parameters based on a training set of images in the LIVE database are as follows: $\sigma = 1.66$, $\beta = 0.4$, $\Delta_{norm} = 11.0$. Note that the optimized β value in our suprathreshold model is different than typical values of 2 to 4, quoted for JND experiments. (These optimized values have been used to generate the results for PPIQ-NDC in Table I.)

5. CONCLUSION

The proposed framework for perceptual assessment of image quality in this paper offers the following benefits: 1- A more realistic approach by suggesting the *comparative memory* model for perceptual quality assessment as opposed to the *perfect photographic memory* model used in conventional IQA methods. 2- Capturing the random nature of quality assessment amongst different observers and subjective test sessions. 3- Allowing applications to define quality metrics, by assessing specific feature(s) which matches their specific needs. 4- Meaningful interpretation of distance pooling, based on the expected probability of finding feature discrepancies at suprathreshold conditions.

TABLE I
RMSE FOR ALIGNED-DMOS REGRESSION
BASED ON DIFFERENT OBJECTIVE METRICS

Distortion Category	# of Images	PPIQ	VIF	SSIM	MSE
JPEG 2000	227	6.10	6.86	8.76	11.19
JPEG	233	7.10	6.57	10.38	14.58
White Noise	174	9.11	5.76	6.65	8.79
Gaussian Blur	174	5.16	4.55	9.34	11.67
JPEG 2000 with FFRCE	174	7.66	8.05	9.79	13.54
All categories	982	8.72	9.50	12.02	14.02

6. REFERENCES

[1] I. Hontsch, L. J. Karam, "Adaptive image coding with perceptual distortion control," *Image Processing, IEEE Transactions on,* vol. 11, pp. 213-222, 2002.

[2] Z. Liu, L.J. Karam, A.B. Watson, "JPEG2000 encoding with perceptual distortion control," *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on* , vol.1, pp. I-637-40, 14-17 Sept. 2003

[3] J.G. Robson, J.G.; Graham, N., "Probability summation and regional variation in contrast sensitivity across the visual field," *Vision Research* Volume 21, Issue 3, Pages 409-418, 1981.

[4] L. Sanghoon, M.S. Pattichis, A.C. Bovik, "Foveated video compression with optimal rate control" *IEEE Trans. on Image Processing*, pp. 977 – 992, July 2001.

[5] A. B. Watson, J. A. Solomon, "A model of visual contrast gain control and pattern masking," J. Opt. Soc. Amer., vol. 14, pp. 2397–2391, 1997.

[6] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol.13, no.4, pp. 600-612, April 2004.

[7] D. L. Ringach, "Mapping receptive fields in primary visual cortex," J. Physio. Lond., vol. 558, no. 3, pp. 717-728, Aug. 2004.

[8] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, " LIVE Image Quality Assessment Database Release 2", Available: http://live.ece.utexas.edu/research/quality.

[9] H.R. Sheikh, A.C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on* , vol.15, no.2, pp. 430-444, Feb. 2006.

[10] Ann M. Rohaly, et al, "Video Quality Experts Group: current results and future directions," *Proc. SPIE* 4067, 742, DOI:10.1117/12.386632, 2000.