Consumer Content Framework for Blind Photo Quality Evaluation

Michele A. Saad, Philip Corriveau, and Ramesh Jaladi

Abstract—We present a framework for blind quality assessment that is targeted at consumer content evaluation. We show that popular state-of-the-art blind image quality assessment algorithms that are successful on the popular publicly available databases for quality assessment do not deliver sufficiently accurate prediction in real consumer usage models (such as when comparing the quality of images captured by two consumer cameras). Towards solving this problem, we propose a model of quality assessment targeted at consumer content evaluation.

I. INTRODUCTION

O-Reference quality assessment methods have steadily achieved higher correlations with subjective judgments of visual quality on publicly available databases, which have been designed for the purpose of quality assessment research. Two of the more popular databases for image quality assessment (IQA) research are the LIVE-IQA database [1] and the TID2013 database [2]. While significant progress in blind image quality assessment algorithm design has been made [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], these methods have been developed and tested on databases that support either single distortions (JPEG, JPEG2000, Blur, Noise), or simulated distortions that are not necessarily representative of what one encounters in a real consumer device system. The simulated distortions on which these methods have been designed and tested are also more focused towards compression and transmission errors and are not representative of typical distortions that arise in photographs captured by the multitude of consumer devices. The type of image distortions encountered within the plethora of captured, shared, and transmitted photos are largely attributed to the optical system of the camera as well as (and to a great extent) to the post processing imaging chain that follows the optical system employed by the device. Simulating this type of complex image distortion is extremely challenging. Most current blind IQA approaches fail to predict image quality in a manner that agrees well with human perception on these kinds of complex consumer images. In addition to the complexity of how distortions manifest during image capture and post processing, the range of qualities produced by consumer devices is narrower than what is typically represented in the popular IQA databases. Fig. 1 shows two examples of unrealistically low quality images (not a typical representation of low quality photographs from consumer devices) from IQA databases typically used for IQA algorithm design and evaluation.

While an objective quality assessment approach may perform well (as measured by correlation between predicted and subjective ratings) on a wide range of qualities, the correlation may exhibit a significant drop in a narrower quality range. Fig. 2 illustrates what is meant by the range effect visually.

Fig. 1. Two images with unrealistically low quality from commonly used IQA databases for quality assessment algorithm design and evaluation.

Fig. 2. The range effect illustrated. The correlation between x and y variables are quite high as exhibited in this scatter plot. They are however much lower in the narrower range of the quantities depicted inside the red box.

A wide range of qualities is essential to ensure a proper coverage of the space, but this spectrum needs to be a representative one of images encountered in practical user scenarios. In this work, we propose a new approach to quality assessment that targets images from consumer devices (mobile phones, tablets, compact point-and-shoot cameras as well as higher end DSLRs). The proposed approach tackles structural properties and includes some color reproduction aspects of consumer images.

To test our approach, we conducted a subjective study for the purpose of understanding and modeling blind perception.
of consumer content. The result of the study is a database of consumer content (images) captured from 27 different devices (mobile phones, tablets, and higher end cameras) of various qualities, and associated subjective MOS scores. Unlike previous databases, this image corpus does not contain simulated image distortions, but the quality of the images in it is dictated by the consumer devices used to capture the images (different devices result in different quality images). This makes the image distortions as well as the ranges they cover more representative of real consumer content. We show that our approach outperforms state of the art blind IQA methods on this type of consumer content.

Consequently, the contributions of this work are twofold: 1) A number of quality aware interpretable features that have a direct impact on overall perceived quality (measures of sharpness, noise, dynamic range, illumination, color warmth, and color saturation) are defined. 2) A mapping to quality scores that we show outperforms existing quality assessment methods on consumer data is derived.

II. FRAMEWORK FOR BLIND CONSUMER CONTENT IMAGE QUALITY ASSESSMENT

When deciding on vectors of investigation for the development of this approach, it was decided to take a consumer centric view. When consumers talk about quality in their words, terms like noise and sharpness are prevalent in their feedback. Since the methodology adopted here is meant to define a new space of investigation, it was defined to be within these key parameters referenced by consumers. We define a list of quality related features that have a direct relevance on quality; these are measures of sharpness and detail, noise, dynamic range, goodness of the scene illumination, color warmth, and saturation.

A. Image Sharpness and Detail

Image sharpness is computed along the edges of an image at various scales ranging from coarse to fine, where finer scales represent finer image detail and the optical system’s ability to capture them. The proposed sharpness measure consists of decomposing the image along major axes of orientation. The major axes are the vertical, horizontal, and the two main diagonals (0°, 90°, 45°, 135° respectively). To do so, we define a number of Gaussian oriented filters (similar in nature to the real part of a Gabor filter). The difference between these and Gabor filters is that in this work these filters are real and do not entail a complete Gabor decomposition. An illustration of the real Gabor-like filters at 4 scales and 4 orientations is depicted in Fig. 3. A sharpness map at a particular scale is then obtained according to Fig. 4. The edge map ensures that the flatter areas in an image, which could potentially be noisy, do not contribute to the sharpness map. This is based on our observation that many images that exhibit lower image sharpness tend to be highly noisy as well (due to a poorer optical system and post capture processing). Computing sharpness along the edges at multiple scales and orientations ensures that noise contributes less to the final sharpness measure. A sharpness map at 9 scales is obtained and the mean sharpness for each map is obtained. This decomposition along 9 image scales (each of which results in a sharpness map at a particular scale) is essential in quantifying image detail. It is desirable to have a sharp image (strong edge responses), and it is also desirable that fine image detail be captured by the optical system. The finer Gabor-like decompositions capture whether a system is able to represent fine detail well or not. An image that is both sharp and exhibits fine details has a strong response for each scale. On the other hand, an image could exhibit a strong response in the coarser scale decompositions (due to boosting of the edge response in post processing), but still fail to exhibit as strong a response in the finer scale decompositions (poorer ability to represent detail).

B. Image Smoothness

The way in which noise manifests in real consumer-like images is highly complex in nature and is not well approximated...
as Additive White Gaussian noise variations. Realistic noise is signal dependent and its properties in the final image are a function of the camera processing chain which transforms the raw camera sensor data into a final image for display. In order to characterize noisy image regions we have developed a two stage approach to noise quantification. The first stage identifies regions in an image that possess noise signatures (random variations that are not structurally meaningful). These noise signatures are distinguishable from image regions that are textured and hence characterized by a relatively high degree of structural regularity. Fig. 5 shows examples of image patches representing noisy and textured image regions. These image patches are obtained from real consumer images as opposed to texture and noise simulations. It is observed that intensity variation is not enough to distinguish between texture and noisy flat image regions. Fig. 6 shows a high degree of overlap in pixel intensity standard deviation between flat (possibly noisy) and textured image regions.

\[
\tilde{G}[i,j] = \frac{G[i,j]^2}{\sum_{i,j} G[i,j]^2},
\]

The Renyi entropy for an image patch is then given by

\[
R = -\frac{1}{2} \log \left( \sum_{i,j} \tilde{G}[i,j]^3 \right)
\]

In addition to the Renyi entropy of the patch filter response, the ratio of the 25th percentile to the 75th percentile as well as the mean of the absolute value of each filter response is computed and used to characterize the image patch (texture versus noise). These features along with the Renyi entropy are used to determine candidate flat image regions where noise is highly visible. An unsupervised classifier (similar to k-means) is used to determine the candidate regions that exhibit noise signatures.

A noise measure is then computed on these candidate image regions. To quantify luminance noise the average pixel intensity variation (luminance variance) is computed locally from the candidate noise patch. By locally, it is meant that the image patch (100 × 100) is further subdivided into smaller blocks (5 × 5), where local intensity variation is computed. The intensity variation of the blocks is then averaged into a mean variation per image patch. A second order statistic (variance) of the local noise variation is computed as well and used later in the model for quality prediction. Color noise is then computed as the mean of the variation in the hue channel of the image patch after transformation of the image patch to HSV color space. The color noise is used in conjunction with the first and second order luminance variations later in the model that maps the computed features to overall image quality.

C. Dynamic Range and Goodness of Scene Illumination

Dynamic range and scene illumination are particularly important in their impact on overall quality in night time scenes where underexposure issues might arise or in bright light scenes where over-exposure issues are common.

1) Dynamic Range: Image dynamic range is computed as the discrete Kullback-Leibler divergence between a discretized

![Histograms of Standard Deviation of Noise Snips and Structure/Texture Snips](image)
uniform distribution and the empirical distribution of the image pixel intensity. A perfect representation of every possible grayscale level would result in a histogram that closely resembles a uniform distribution. The flatter the image histogram, the closer its empirical distribution to that of an ideal uniform distribution (the lower the KL divergence), and hence the higher the dynamic range. The discrete KL divergence is given as

\[
D_{KL}(P, U) = \sum_x P(x) \ln \frac{P(x)}{U(x)},
\]

where \( P \) and \( U \) are the empirical pixel intensity distribution and the discrete uniform distribution respectively. Fig. 7 shows two night images with varying dynamic ranges. The image on the left is a night image that is well illuminated with few under-exposed areas. Details are well captured and represented. The image on the right is highly under-exposed. Much of the image detail is lost. The image on the left hence has higher dynamic range that the image on the right.

![Fig. 7. Two night images with varying dynamic ranges. The image on the left is a night image that is well illuminated with few under-exposed areas. Details are well captured and represented. The image on the right is highly under-exposed. Much of the image detail is lost. The image on the left hence has higher dynamic range that the image on the right.](image)

2) Goodness of Scene Illumination: In addition to the dynamic range measure, a measure of illumination is computed as the average local image exposure modulated by the average local image contrast in the image. The idea behind the measure is that high illumination can be desirable so long as image detail is still captured (i.e. high contrast is maintained). Lower illumination on the other hand, or under-exposure is more tolerable if image contrast/detail is maintained than when an image regions is completely under-exposed (no detail or contrast maintained).

The local image contrast is computed at the Michelson contrast given by

\[
C = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}},
\]

and \( C \) is modulated by \( I_{\text{avg}} \), (i.e. \( C \times I_{\text{avg}} \) is computed for each block \( 17 \times 17 \) and then averaged over the entire image.) \( I_{\text{max}}, I_{\text{min}}, \) and \( I_{\text{avg}} \) are the minimum, maximum, and mean intensity of a local image block respectively.

D. Non-structural Image Quality Features: Color Information

The above indicators of quality are primarily structural measures that are impacted by structural image variations but are invariant to uniform image variations such as a homogeneous change in mean illumination, a shift in saturation, or image color temperature. However, these non-structural characteristics are tied to image quality perception albeit less so than structural image characteristics. To quantify some of these non-structural characteristics, we compute overall image saturation and color warmth. The image color saturation is given as the mean of the saturation channel of image post transformation to HSV color space.

Image color warmth is computed as the average distance of each image pixel from its current position in CIE XYZ color space to the closest point to it in the Planckian locus. The Planckian locus or black body locus is defined as the locus that an incandescent black body would take as its temperature changes. The locus goes from a deep red color at low temperatures to a blue color at high temperature (notice that the temperature of a black body and its color are inverse to what we associate as warm colors: red is a warm color and blue is a cool color, but by definition a red black body temperature is cool and a blue black body temperature is warm.) The color warmth of each pixel in the image are averaged into an overall image color warmth measure.

III. Experiments

The proposed approach is developed for the purpose of quality assessment of consumer-type photographs. These are photos taken by consumer devices such as mobile phones, tablets, point-and-shoots or higher end cameras. As explained in the introduction, these photos exhibit highly complex distortion characteristics that are difficult to simulate. To evaluate our consumer-centric quality assessment method, we conducted an overall quality photo rating study.

A subjective study was conducted with 42 participants (23 male, 19 female), all of whom fit in the following categories a) imaging engineers, b) professional and amateur photographers, c) graphic designers or d) frequent phone camera users. All participants were screened for normal or corrected to normal visual acuity and normal color vision. The database used was designed specifically to address characteristics of image capture typical of a consumer and of those not represented in other databases. This means that each group contained a night scene, a landscape scene, a close up scene and a wall scene. In other databases, this group contained a night scene, a landscape scene, a close up scene and a wall scene. In an efforts to have the captured content be repeatable, reliable, and representative portraits and in-scene motion were avoided. Another unique aspect of the database is that unlike similar databases available today, it does not contain the same image with a range of distortions for comparison. Instead, since 27 different devices were used for the photo capture, there are 27 distinct angles of view for each photo due to the fundamental nature of differing aspect ratios, fields of view and shot angles. It is felt that this is more representative of how consumers capture and compare images with each other. A total of 405 photos were evaluated (15 scenes from 27 capture devices). Each participant was asked to rate the overall image quality using a single stimulus evaluation method on a standard 5 point quality Likert scale where (1 Excellent, 4 Good, 3 Fair, 2 Poor and 1 Bad). All image evaluations were completed on a
secondary Dell UltraSharp 24 inch Monitor (U2412M) with a resolution of 1920*1080 so that the capture device display was removed as a variable. Some examples of the photos rated are in Fig. 8 and the database of images will be made available to the research community in the near future.

![Example photos](image)

**Fig. 8.** Example photos from the subjective study. The photos inherently have different resolutions and different aspect ratios (depending on the device by which they were captured.)

### IV. RESULTS

A linear support vector regressor is used to map the quality aware features described above to a mean opinion score. To evaluate the predictive performance of the proposed features we performed a scene-leave-one-out analysis, where 14 scenes (captured by 27 devices) are used for training and the remaining scene (27 photos from the left out scene) are used for testing. This is repeated until each scene set is used as a test set in the analysis. This ensures that the test set never contains a similar scene as one in the training set. For each test set we compute the Spearman rank order correlation coefficient (SROCC) between predicted and subjective MOS scores. We compare this SROCC with the SROCC between subjective and quality scores on the same test sets computed by the state-of-the-art quality assessment methods C-DIIVINE [9], BIQI [3], SSEQ [8], BRISQUE [6], and DESIQUE [10]. The results are shown in Table I. It is worth mentioning here that the algorithms we compared against did not perform well when applied to the images from the devices in their default original resolutions (2MP to 20MP), so the images were rescaled to a common resolution to allow for better comparison. In addition to the scene specific SROCC analysis, we combined the MOS scores of all the 15 scenes in one vector (from the leave-one-out analysis; i.e we still maintained the fact that there were no common scenes between train and test sets) and computed an overall SROCC. These results are shown in Table II. The results show that our approaches outperform the state-of-the-art approaches on consumer content data (photos from mobile and compact cameras).

### V. CONCLUSION

The approach presented in this paper serves to define a more consumer-relevant and tractable quality assessment space and develop and apply methods that quantify specific interpretable imaging characteristics. In this vein it is directed at consumer applications and gaps in current databases and research. Solving a specific problem and providing a methodology for repeatable reliable and representative captures allows for an evaluation method that is very stable for a subset of the imaging space.

### REFERENCES


### TABLE II

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall SROCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>0.711</td>
</tr>
<tr>
<td>C-DIVINE</td>
<td>0.522</td>
</tr>
<tr>
<td>BIQI</td>
<td>0.129</td>
</tr>
<tr>
<td>SSEQ</td>
<td>0.405</td>
</tr>
<tr>
<td>BRISQUE</td>
<td>0.523</td>
</tr>
<tr>
<td>DESIQUE</td>
<td>0.539</td>
</tr>
</tbody>
</table>
TABLE I
SROCC CORRELATIONS ON SUBJECTIVE STUDY 15 SCENES

<table>
<thead>
<tr>
<th>Content</th>
<th>C-DIVINE</th>
<th>BIQI</th>
<th>SSEQ</th>
<th>BRISQUE</th>
<th>DESIQUE</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scene 1</td>
<td>0.533</td>
<td>0.536</td>
<td>0.631</td>
<td>0.372</td>
<td>0.490</td>
<td>0.707</td>
</tr>
<tr>
<td>Scene 2</td>
<td>0.739</td>
<td>0.685</td>
<td>0.544</td>
<td>0.723</td>
<td>0.657</td>
<td>0.785</td>
</tr>
<tr>
<td>Scene 3</td>
<td>0.627</td>
<td>0.583</td>
<td>0.633</td>
<td>0.671</td>
<td>0.606</td>
<td>0.782</td>
</tr>
<tr>
<td>Scene 4</td>
<td>0.847</td>
<td>0.755</td>
<td>0.742</td>
<td>0.702</td>
<td>0.499</td>
<td>0.848</td>
</tr>
<tr>
<td>Scene 5</td>
<td>0.440</td>
<td>0.021</td>
<td>0.478</td>
<td>0.310</td>
<td>0.356</td>
<td>0.594</td>
</tr>
<tr>
<td>Scene 6</td>
<td>0.433</td>
<td>0.394</td>
<td>0.525</td>
<td>0.547</td>
<td>0.639</td>
<td>0.773</td>
</tr>
<tr>
<td>Scene 7</td>
<td>0.768</td>
<td>0.448</td>
<td>0.255</td>
<td>0.301</td>
<td>0.338</td>
<td>0.774</td>
</tr>
<tr>
<td>Scene 8</td>
<td>0.203</td>
<td>0.151</td>
<td>0.707</td>
<td>0.527</td>
<td>0.516</td>
<td>0.886</td>
</tr>
<tr>
<td>Scene 9</td>
<td>0.598</td>
<td>0.022</td>
<td>0.653</td>
<td>0.513</td>
<td>0.464</td>
<td>0.863</td>
</tr>
<tr>
<td>Scene 10</td>
<td>0.550</td>
<td>0.032</td>
<td>0.686</td>
<td>0.417</td>
<td>0.538</td>
<td>0.870</td>
</tr>
<tr>
<td>Scene 11</td>
<td>0.440</td>
<td>0.224</td>
<td>0.563</td>
<td>0.415</td>
<td>0.494</td>
<td>0.636</td>
</tr>
<tr>
<td>Scene 12</td>
<td>0.056</td>
<td>0.165</td>
<td>0.379</td>
<td>0.210</td>
<td>0.256</td>
<td>0.420</td>
</tr>
<tr>
<td>Scene 13</td>
<td>0.544</td>
<td>0.537</td>
<td>0.059</td>
<td>0.309</td>
<td>0.412</td>
<td>0.825</td>
</tr>
<tr>
<td>Scene 14</td>
<td>0.711</td>
<td>0.245</td>
<td>0.650</td>
<td>0.701</td>
<td>0.548</td>
<td>0.686</td>
</tr>
<tr>
<td>Scene 15</td>
<td>0.240</td>
<td>0.159</td>
<td>0.453</td>
<td>0.103</td>
<td>0.303</td>
<td>0.646</td>
</tr>
</tbody>
</table>