COMBINING THE BEST OF PERCEPTUAL QUALITY METRICS

Adriaan Barri, Ann Dooms, Peter Schelkens

Vrije Universiteit Brussel (VUB), Dept. of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, Belgium - Interdisciplinary Institute for Broadband Technology (IBBT), Dept. of Future Media and Imaging (FMI), Gaston Crommenlaan 8 (box 102), B-9050 Ghent, Belgium.

1. INTRODUCTION

With the rise of high resolution monitors and cameras, consumers have increasing demands concerning the visual quality of multimedia applications. The output, as received by the end-user, has undergone a series of processes, going from production and transmission to visualization. Each of these processes can cause quality loss of which the total impact is difficult to model. In order to meet the consumers' expectations, quality monitoring all along the E2E chain is of utmost importance.

The most accurate form of quality monitoring is obtained by performing subjective tests on a representative sample of distorted signals. However, subjective experiments are expensive and time-consuming, because they cannot be automated. These drawbacks triggered the research community to design *objective quality metrics*, which are models that can estimate the perceptual quality of degraded images or videos.

Today's most popular quality metrics assume the *full-reference* (FR) scenario, in which the original, non-degraded signal is available for comparison. In many real-world applications, however, there is no access to this information at the end user's side. This is for example the case for streaming applications and complex broadcast scenarios where multiple parties are involved. Hence, it is not possible to use FR quality metrics for automated quality monitoring.

The ultimate goal is to design *no-reference* (NR) metrics, that are completely independent of the original signal. However, due to our limited knowledge about the Human Visual System (HVS), it is not yet possible to construct reliable NR metrics.

A more realistic approach is based on *reduced-reference* (RR) techniques. Most applications allow to send some features about the original signal through a side-channel with a low bandwidth. This side-information can for example be sent through a synchronized data stream or using watermarking techniques. By employing a side-channel, it is possible to measure the similarity between the features of the original signal and the corresponding features of the degraded signal. In this way, RR quality metrics can be constructed.

In this paper, a novel data-driven *combination method* for RR quality assessment is proposed. The method's performance is validated by applying it on four simple distortion-specific quality metrics to produce a new RR quality metric for images.

2. METRIC COMBINATION METHOD

Consider a subjective database of test signals

$$\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N,$$

that covers the whole quality range. Every signal $\mathbf{x}_i$ is evaluated by a large number of test subjects in order to generate a subjective score, which is denoted $s(\mathbf{x}_i)$. Let

$$P_0, P_1, P_2, \ldots, P_{k+1}$$

be a partitioning of the database such that

$$s(\mathbf{x}) < s(\mathbf{x}') \text{ for every } \mathbf{x} \in P_i,\ \mathbf{x}' \in P_j \text{ with } i < j \quad (1)$$

An objective quality metric d should ideally satisfy $d(\mathbf{x}) < d(\mathbf{x}')$ for all $\mathbf{x} \in P_i$, $\mathbf{x}' \in P_j$ with $i < j$. In practice, however, most quality metrics fail to achieve this constraint, due to their limited correlation with the HVS. This suggests to introduce a weakened version of this property.

Definition. *Let $P_1, P_2, \ldots, P_k$ be a partitioning of a subjective database that satisfies (1). We say that P_i and P_j, $0 \leq i < j \leq k+1$, are λ-separated for a quality metric d if and only if*

$$\left(\mathbb{E}[d(P_j)] - \lambda\sqrt{\mathit{Var}[d(P_j)]}\right) - \left(\mathbb{E}[d(P_i)] + \lambda\sqrt{\mathit{Var}[d(P_i)]}\right) > 0. \quad (2)$$

Note that P_i and P_j are λ-separated for d if and only if

$$\lambda < \frac{\mathbb{E}[d(P_j)] - \mathbb{E}[d(P_i)]}{\sqrt{\mathrm{Var}[d(P_j)]} + \sqrt{\mathrm{Var}[d(P_i)]}}. \quad (3)$$

The upper-bound in (3) can be used as an optimization criterion to combine a series of quality metrics $d_1, d_2, \ldots, d_P$. Every linear combination of these metrics can be written in the form $\mathbf{v}^T\mathbf{d}$, with $\mathbf{v} \in \mathbb{R}^P$ and $\mathbf{d} = (d_1, d_2, \ldots, d_P)^T$. This leads to the following definition of $C_{i,j}$, $0 \leq i < j \leq k+1$:

$$C_{i,j}[\mathbf{v}] = \frac{\mathbb{E}[\mathbf{v}^T\mathbf{d}(P_j)] - \mathbb{E}[\mathbf{v}^T\mathbf{d}(P_i)]}{\sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_j)]} + \sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_i)]}}.$$

However, the corresponding optimization problem $\mathbf{v}_{i,j} = \arg\max_{\mathbf{v}} C_{i,j}[\mathbf{v}]$ is not trivial to solve. We therefore modify the definition of $C_{i,j}$ to

$$\tilde{C}_{i,j}[\mathbf{v}] = \frac{\mathbb{E}[\mathbf{v}^T\mathbf{d}(P_j)] - \mathbb{E}[\mathbf{v}^T\mathbf{d}(P_i)]}{\sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_j)] + \text{Var}[\mathbf{v}^T\mathbf{d}(P_i)]}}.$$

An easy calculation gives

$$\frac{1}{\tilde{C}_{i,j}[\mathbf{v}]^2} = \frac{\mathbf{v}^T A \mathbf{v}}{\left(\mathbf{v}^T\mathbf{b}\right)^2},$$

where $A = \text{Var}[d(P_i)] + \text{Var}[d(P_j)]$ and $\mathbf{b} = \mathbb{E}[d(P_i)] - \mathbb{E}[d(P_j)]$. Using basic techniques from linear algebra, one can efficiently find a vector $\tilde{\mathbf{v}}_{i,j}$ that minimizes the last fraction in the above equation. Hence, by assuming that

$$\mathbb{E}[\mathbf{d}(P_j)] > \mathbb{E}[\mathbf{d}(P_i)]$$

for every $1 \leq i < j \leq k$, we obtain that

$$\tilde{\mathbf{v}}_{i,j} = \arg\min_{\mathbf{v}} \frac{1}{\tilde{C}_{i,j}[\mathbf{v}]^2} = \arg\max_{\mathbf{v}} \tilde{C}_{i,j}[\mathbf{v}]. \qquad (4)$$

In order to describe the relationship between $C_{i,j}$ and $\tilde{C}_{i,j}$, we introduce the notation

$$R_{i,j}[\mathbf{v}] = \min\left(\frac{\sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_i)]}}{\sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_j)]}}, \frac{\sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_j)]}}{\sqrt{\text{Var}[\mathbf{v}^T\mathbf{d}(P_i)]}}\right)$$

for every $\mathbf{v} \in \mathbb{R}^P$.

Proposition 2.1. *Let $\mathbf{d} = (d_1, d_2, \ldots, d_P)^T$ be a series of quality metrics. The following equality holds for every $\mathbf{v} \in \mathbb{R}^P$:*

$$\tilde{C}_{i,j}[\mathbf{v}] = C_{i,j}[\mathbf{v}] \frac{1 + R_{i,j}[\mathbf{v}]}{\sqrt{1 + R_{i,j}[\mathbf{v}]^2}}.$$

In particular, we have

$$\min_{\mathbf{v}} C_{i,j}[\mathbf{v}] \leq \min_{\mathbf{v}} \tilde{C}_{i,j}[\mathbf{v}] \leq \sqrt{2}\min_{\mathbf{v}} C_{i,j}[\mathbf{v}].$$

Since the function that maps x to $(1+x)/\sqrt{1+x^2}$ is monotonically increasing on the interval $[0,1]$, we see that the modified criterion $\tilde{C}_{i,j}[\mathbf{v}]$ tends to be larger when the variances $\text{Var}[\mathbf{v}^T\mathbf{d}(P_i)]$ and $\text{Var}[\mathbf{v}^T\mathbf{d}(P_j)]$ are closer to each other and vice versa. This property ensures a better balance between the variances of $\tilde{\mathbf{v}}_{k,l}^T\mathbf{d}(P_i)$ and $\tilde{\mathbf{v}}_{k,l}^T\mathbf{d}(P_j)$, with $\tilde{\mathbf{v}}_{i,j}$ the optimal solution given in (4).

The proposed combination algorithm is based on the solutions $\tilde{\mathbf{v}}_{i,i+1}$, $i \in \{0, 1, 2, \ldots, k\}$. The following procedure describes how the quality score of a signal $\mathbf{x}$ is determined:

```
for i ← 0 to k do
    d^(i) ← ṽ^T_{i,i+1} d;
    λ^(i) ← C_{i,i+1}[ṽ_{i,i+1}];
    B_i ← E[d^(i)(P_i)] + λ^(i) sqrt(Var[d^(i)(P_i)]) ;
    if d^(i)(x) ≤ B_i then
        return d^(i)(x);
    end
end
```

3. A NEW RR QUALITY METRIC FOR IMAGES

We apply the proposed combination method on four RR image quality metrics, each of them focusing on a specific type of distortions to obtain a higher (local) correlation with the HVS.

The metrics are evaluated on the TID2008 database [1] and the resulting quality scores are plotted in Figure 1. To simplify the demonstration, we restrict the database to the distortions "Gaussian noise", "Gaussian blur", "JPEG compression" and "JPEG 2000 compression".

The metric d_{SI} compares edge information extracted from the original, undistorted image $\mathbf{x}$ and the observed, degraded image $\mathbf{y}$ to detect structural information loss. Only one feature is needed from the original image, namely

$$f_{\text{SI}}(\mathbf{x}) = \sigma\left(\sqrt{G_h^2(\mathbf{x}) + G_v^2(\mathbf{x})}\right),$$

where σ denotes standard deviation, $G_h(\mathbf{x}) = \left[\begin{smallmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{smallmatrix}\right] * \mathbf{x}$ and $G_v(\mathbf{x}) = \left[\begin{smallmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{smallmatrix}\right] * \mathbf{x}$. The structural information index is defined as

$$d_{\text{SI}}(\mathbf{x}, \mathbf{y}) = \min\left(\frac{f_{\text{SI}}(\mathbf{y}) + C_1}{f_{\text{SI}}(\mathbf{x}) + C_1}\ ; 1\right),$$

where a small constant C_1 is included to avoid instability when $f_{\text{SI}}(\mathbf{x})$ is close to zero (we take $C_1 = 1$ in this test). Note that a similar metric was previously proposed in [5].

The metric d_{HIST} is used to assess the quality of JPEG compressed images. The reference and distorted image are partitioned into 25 equal-sized blocks $\mathbf{x}_i$ and $\mathbf{y}_i$, with $i = 1 \ldots 25$. The corresponding block histograms are denoted by $H_{\mathbf{x}_i}$ and $H_{\mathbf{y}_i}$ respectively. Consider the local scores

$$\delta_{\text{HIST}}(\mathbf{x}_i, \mathbf{y}_i) = \|H_{\mathbf{y}_i}\|_2^2 / \|H_{\mathbf{x}_i}\|_2^2,$$

where $\|\cdot\|_2$ denotes the L^2 norm. Since JPEG compression algorithms attenuate detail at some places and introduce compression artifacts like ringing and blocking at others, the variance between the local scores $\delta_{\text{HIST}}(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \ldots 25$ will be high. This observation leads us to the definition

$$d_{\text{HIST}}(\mathbf{x}, \mathbf{y}) = \frac{1}{\sigma\big(\delta_{\text{HIST}}\big) + 1}.$$

The impact of spatial noise on the image quality is measured using the contrast comparison function d_{CON} of the structural similarity (SSIM) index [3]. More precisely,

$$d_{\text{CON}}(\mathbf{x}, \mathbf{y}) = 1 - \frac{(\sigma_{\mathbf{x}} - \sigma_{\mathbf{y}})^2 + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

where $C_2 = 58$. We evaluate this metric locally on the image blocks $\mathbf{x}_i$ and $\mathbf{y}_i$. The resulting scores are then combined by taking the mean.

The last metric d_{JPEG} is the NR quality measurement algorithm for JPEG compression developed in [2]. A MATAB implementation is publicly available from the author's website.

Before we can apply the combination algorithm on the above described quality metrics, we need to specify a partitioning P_j, $j = 0, 1, \ldots, k+1$ of the subjective database. We therefore divide the subjective scores of the database into a number of intervals $I_j =]s_j, s_{j+1}]$, $j = 0, 1, 2, \ldots, k$, for which

$$s_0 < s_1 < s_2 < \ldots < s_{k+1}.$$

and define P_j as the set of all images in the database with a subjective score between s_j and s_{j+1}. We will refer to this partitioning using the notation

$$[s_0 \ s_1 \ s_2 \ \ldots \ s_{k+1}].$$

The output of the combination algorithm is a piecewise linear combination of the input metrics with exactly k breakpoints. In this experiment, we achieved the best results for $k = 2$ using the partitioning $[0 \ 3.5 \ 5 \ 7]$.

We verified the robustness of the combination algorithm using a tenfold cross validation. We observed that the output was almost constant during the ten iterations. In Figure 2, the obtained quality metric is visualized, together with the output of the combination algorithm when the partitioning $[0 \ 4 \ 7]$ is used. This figure demonstrates the importance of the chosen partitioning.

In Table 1, we compare the proposed quality metric to the SSIM index [3] and to the RR quality metric (RRIQA) in [4].

Metric	Pearson	Spearman	Kendall
Proposed Metric	0.890	0.915	0.74
SSIM [3]	0.699	0.768	0.566
RRIQA [4]	0.639	0.655	0.487

Table 1. Performance comparison of the proposed metric, the SSIM index and the RRIQA index on a subset of the TID2008 database.

4. CONCLUSION

In this paper, we presented a new piecewise-linear combination algorithm for quality metrics. We demonstrated the performance of this algorithm by applying it on four RR distortion-selective quality metrics.

5. ACKNOWLEDGEMENTS

This research was supported by the Fund for Scientific Research Flanders (projects G.0206.08, G021311N and the Post-Doctoral Fellowship of Peter Schelkens) and by the Flemish Institute for the Promotion of Innovation by Science and Technology (IWT) (PhD bursary Adriaan Barri).

6. REFERENCES

[1] N.N. Ponomarenko, V.V. Lukin, K.O. Egiazarian, J. Astola, M. Carli, and F. Battisti, "Color image database for evaluation of image quality metrics," in Proc. MMSP, 2008.

[2] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-Reference Perceptual Quality Assessment of JPEG Compressed Images," in Proceedings of IEEE 2002 International Conference on Image Processing, 2002.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. 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.

[4] Z. Wang, E. Simoncelli, "Reduced-Reference Image Quality Assessment Using A Wavelet-Domain Natural Image Statistic Model," Human Vision and Electronic Imaging X, Proc. SPIE, vol. 5666, San Jose, CA, Jan. 2005.

[5] S. Wolf and M. Pinson, "Video quality measurement techniques," NTIA Report 02-392, 2002.

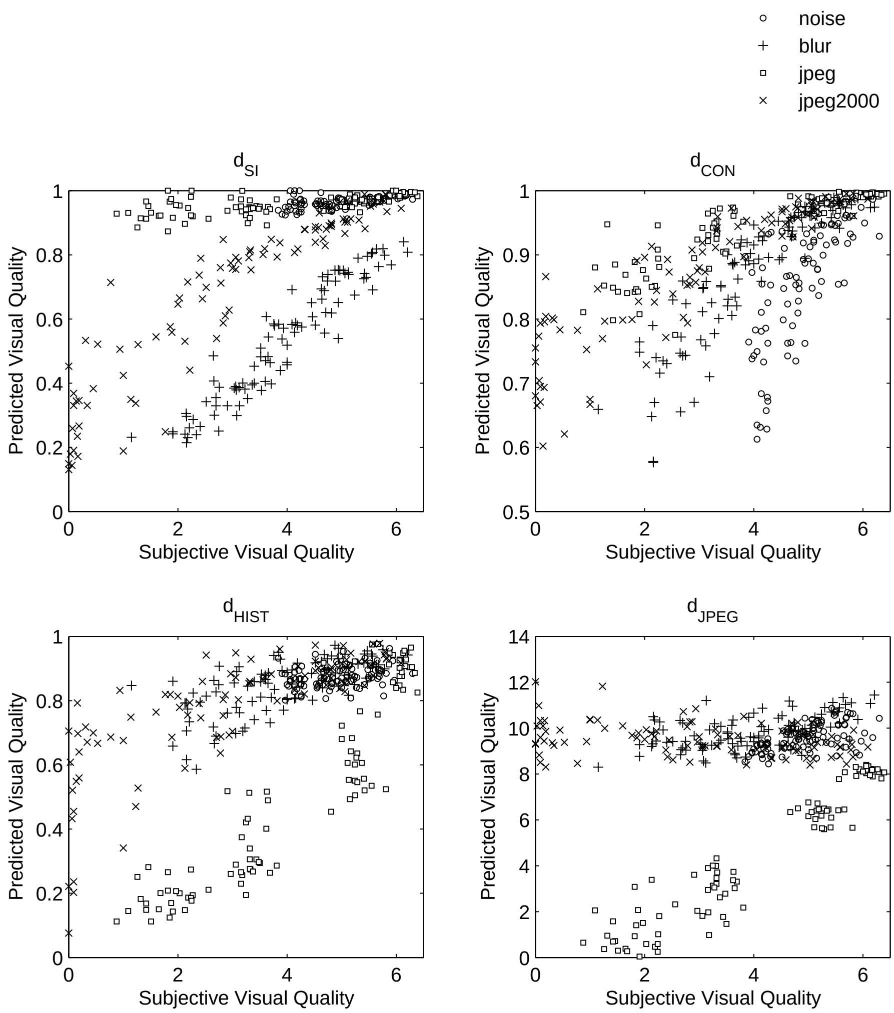

Fig. 1. Scatter plot comparison of the distortion-selective quality metrics d_{SI}, d_{HIST}, d_{CON} and d_{JPEG}. Each sample point represents one image in the TID2008 database.

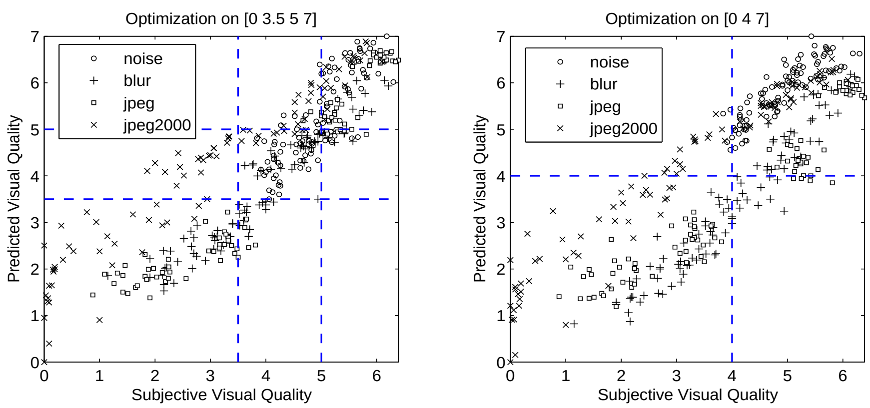

Fig. 2. Comparison of the combination method for two different partitions. We see that the metric on the left has a higher correlation.