Signal Processing 97 (2014) 110–116

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Additive models and separable pooling, a new look at structural similarity

Boban P. Bondzulic [a,*], Vladimir S. Petrovic [b]

[a] University of Defence in Belgrade, Military academy, Generala Pavla Jurisica Sturma 33, 11000 Belgrade, Serbia
[b] University of Manchester, Imaging Science and Biomedical Engineering, Oxford Road, Manchester M13 9PT, UK

ARTICLE INFO

Article history:
Received 18 March 2013
Received in revised form
29 September 2013
Accepted 18 October 2013
Available online 31 October 2013

Keywords:
Image quality assessment
Structural similarity index
Visual importance pooling

ABSTRACT

Objective quality metrics predict perceived quality of image signals computationally and can: (i) *benchmark* and *monitor* compression and processing algorithms and (ii) *optimise* their performance for a given application (content, bandwidth, packet loss…). Structural similarity, represented by the well known SSIM index is a framework for objective assessment of image quality well known for its relative simplicity and robustness. Despite its practical appeal, SSIM's performance level, measured as agreement with subjective quality scores, lags more complex state-of-the-art metrics. We present a new look into structural similarity that uses an additive model and a spatial pooling approach that decouples individual structural comparisons and utilises the quality driven aggregation paradigm. We apply this new approach to both baseline intensity SSIM and gradient SSIM (GSSIM) frameworks and show, through extensive evaluation on four publicly available subjective datasets that it provides considerably more ordered (linear) relationship between objective and subjective quality for a variety of input conditions. We demonstrate that newly formulated structural similarity metrics using this approach are capable of equal or even better performance than more complex state-of-the-art objective metrics in the process lending support to the theory that humans base their opinion on the worst sections of the observed signal.

1. Introduction

Recent years have seen tremendous growth in transmission of remotely acquired digital imagery. While subjecting images to distortions such as lossy compression that degrades quality to reduce the bandwidth needed to transmit them and packet loss due to error-prone channels, it is beneficial for a transmission system to quantify these quality degradations so that it can maintain, control and possibly enhance the quality of its output [1].

The most relevant way of assessing quality of transmitted images is direct rating by humans as their ultimate users. However, subjective trials are impractical, inherently offline and slow. Objective metrics instead predict perceived quality computationally and can be practically employed to: (i) *benchmark* and *monitor* compression and processing algorithms and (ii) *optimise* their performance for a given application (content, bandwidth, packet loss…) [2]. To be relevant objective metrics need to demonstrate agreement with (mean) subjective opinions of human observers [3].

Structural similarity index (SSIM) [4] is an approach that has come closest to becoming a *de facto* standard for objective assessment of image quality due to its relative simplicity and robust performance. Since natural images are structured and human visual system (HVS) highly

* Corresponding author. Tel.: +381 64 143 82 79.
E-mail addresses: bondzulici@yahoo.com (B.P. Bondzulic), v.petrovic@manchester.ac.uk (V.S. Petrovic).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.10.020

adapted for extracting such information, SSIM uses a basic idea of comparing structural information between original and degraded signals through localised comparisons of luminance, contrast and structure [4,5]. The three comparison models are then combined into a single quality score (SSIM).

Since its introduction SSIM has become a foundation for a wide range of objective signal quality metrics [6–21]. Into this crowded field we present a new look into structural similarity by deconstructing the basic model and proposing a new additive integration model (AM) and a spatial pooling approach that decouples individual structural comparisons and utilises the quality driven pooling paradigm. We apply this new approach to baseline intensity SSIM [4] and gradient SSIM (GSSIM) [7] frameworks and show through an extensive evaluation on over 1600 subjectively rated test images in four publicly available datasets that it results inconsiderably better behaved metrics with nearly linear relationship with subjective quality for a variety of input conditions. We demonstrate that new structural similarity metrics are capable of equal or even better performance than more complex state-of-the-art objective metrics. In the process, our results lend support to the theory that humans base their opinion on the worst sections of the observed signal. A high degree of inconsistency in SSIM performance quoted in the literature was a partial motivation for the analysis presented in our work.

In the following, Section 2 introduces the structural similarity index for image quality estimation and its evolution over the years. Proposed quality guided similarity pooling and additive integration models are outlined in Sections 3 and 4. These novel formulations are evaluated in Section 5 and we discuss the results and conclude in Section 6.

2. Structural similarity

Various structural similarity models between original and derived (degraded) signals have been in use for two decades to assess image and video quality [22,23] and multisensor image fusion [24]. They gained wider recognition with formulation of the structural similarity index (SSIM) model for image quality assessment [4,5]. Under the assumption that HVS is highly adapted for extracting structural information from a scene, a framework for quality assessment based on the degradation of structural information between two non-negative signals $\mathbf{x}$ and $\mathbf{y}$ can be constructed through direct comparisons of their – luminance (l), contrast (c) and structure (s) defined as:

$$l_{\mathbf{x},\mathbf{y}} = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{1}$$

$$c_{\mathbf{x},\mathbf{y}} = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2}$$

$$s_{\mathbf{x},\mathbf{y}} = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3}$$

μ_x and μ_y represent means, σ_x^2 and σ_y^2 variances and σ_{xy} is the correlation between $\mathbf{x}$ and $\mathbf{y}$. Constants C_1, C_2, and C_3 are set to avoid instability when the denominator is close to zero. All three comparison models effectively measure similarity between the signals and rise to 1 as their differences decrease (they are exactly 1 for $\mathbf{x}=\mathbf{y}$) and conversely decrease as differences between the signal samples increases. They thus reflect on the quality of representation of $\mathbf{x}$ by $\mathbf{y}$ and vice versa, (measurement is symmetric) and are combined into a single structural similarity measure using a multiplicative SSIM model [4]:

$$SSIM_{\mathbf{x},\mathbf{y}} = l_{\mathbf{x},\mathbf{y}}^{\alpha} \cdot c_{\mathbf{x},\mathbf{y}}^{\beta} \cdot s_{\mathbf{x},\mathbf{y}}^{\gamma} \tag{4}$$

where α, β and γ define relative importance of the three components, in a common form $\alpha=\beta=\gamma=1$. As $s_{\mathbf{x},\mathbf{y}}$ can be go down to –1 (for inverted signals), $SSIM_{\mathbf{x},\mathbf{y}}$ is in the range of -1 to 1 ($\mathbf{x}=\mathbf{y}$).

When applied to image signals *SSIM* is evaluated at each location (pixel) in the scene by applying an 11×11 Gaussian kernel window to evaluate local statistics (μ and σ) [4]. Universal Image Quality Index – UIQI [5] uses a square sliding window instead of the Gaussian. This provides a set of localised estimates of structural similarity between the two images, $SSIM_{x,y}$ at $\forall$x,y. A global structural similarity index between the two images is then obtained by taking the mean of local SSIM estimates. The influence of window size and constants C_1, C_2 and C_3 in image quality assessment is analysed in [25]. In [26] the predictive performance of the three SSIM components, l, c and s, (1–3), individually and in pairwise combinations was investigated for quality evaluation of common image artefacts and found that ignoring the luminance comparison produces no drop in metric performance.

A number of modifications have been proposed to improve SSIM's correlation with subjective ratings [6–21]. Multi-Scale Structural Similarity index (MS-SSIM) [6] applies SSIM at a number of resolutions while Chen et al. [7] developed Gradient-based Structural Similary (GSSIM), based on edge information as the most important image structure information. In [8], images are not compared directly, but their similarity is measured by SSIM between feature maps (corner, edge and symmetry maps). In [9] the structure term is replaced by additional terms that depend on region statistics.

SSIM is a local quality/distortion measure and better spatial pooling is one potential improvement [10]. Various strategies were proposed to combine local estimates: using visual fixation and quality based weighting [11], region type weighting [12], information content weighting [13], and pooling using actual visual attention information (eye tracks) [14]. In terms of applications SSIM has been applied to assess quality of colour images [15], multisensor fusion [16] and video [17–21] as well as to computer vision problems [2,27]. A fast SSIM implementation can be found in [28], and transform-based implementations are given in Refs. [29,30]. A relationship between SSIM and MSE is analysed in [31] while its potential in theoretical development and applications is highlighted in [32].The performance of SSIM on different image and video datasets is evaluated in [13,19,20,33,34,35].