

NO-REFERENCE SHARPNESS METRIC BASED ON LOCAL GRADIENT ANALYSIS

Christoph Feichtenhofer, 0830377

Supervisor:

Univ. Prof. DI Dr. techn. Horst Bischof

*Inst. for Computer Graphics and Vision
Graz University of Technology, Austria*



**Bachelor Thesis Telematics
Graz, October 17, 2011**

Abstract

A no-reference perceptual sharpness metric is proposed for sharpness assessment as overall quality indicator in image and video. We also introduce an alternative sharpness metric, able to assess perceived image sharpness, insensitive to local blur distortions. The essential idea of our work is to analyze the edge spread in images and the affect on human blur perception.

Evaluation is performed by using several quality databases and appendant subjective scores. By comparing with state-of-the-art no-reference sharpness metrics we show the advantage of our approach. The proposed algorithms correlate well with the subjective quality ratings. On uniformly blurred content the metric is even competitive to well-established full-reference metrics. We demonstrate stable results for images with diverse content and show the superiority of the proposed metric compared to latest no-reference metrics in the literature. Another advantage of our approach is the low computational complexity, making real-time video sharpness estimation possible. The experiments were conducted on distorted images, including Gaussian blur, JPEG-2000 compression and white Gaussian noise.

The research and implementation of this work has been carried out at JOAN-NEUM RESEARCH Forschungsgesellschaft mbH, DIGITAL- Institute for Information and Communication Technologies. Prototyping was performed in MATLAB and a method to measure the sharpness in video frames was implemented in C++.

Keywords: *sharpness, blur, no-reference, sharpness metric, blur metric, image/video quality assessment, perceptual, human visual system, visual quality*

Contents

1	Introduction	2
2	Overview of Existing Metrics	4
2.1	Proposed Basic Metrics	4
2.1.1	Autocorrelation Metric	4
2.1.2	Variance Metric	5
2.2	Existing No-reference Sharpness Metrics	5
2.2.1	Spatial Domain Metrics	5
2.2.2	Frequency Transform Domain Metrics	6
2.3	Issues of Existing No-reference Sharpness Metrics	7
3	Proposed Sharpness Metric Based on Local Gradient Analysis	10
3.1	Relevant Feature Extraction	10
3.2	Measuring Edge Widths	13
3.3	Subpixel Accurate Local Extrema Estimation	15
3.4	Modeling the Human Perception of Acutance	16
3.5	Block Based Calculation	18
4	Experiments	20
4.1	Subjective Quality Assessment	20
4.2	Subjective Visual Quality Assessment Databases	20
4.2.1	LIVE Quality Assessment Database - LIVE[15]	20
4.2.2	TID2008 Database [16]	21
4.2.3	Database Comparison	21
4.3	Performance Measures	23
4.4	Content Dependency	25
4.5	Noise Susceptibility	27
4.6	Blocking Susceptibility	27
4.7	Performance Results	29
4.7.1	LIVE Database	30
4.7.2	TID2008 Database	31
5	Conclusion	32
6	Acknowledgment	32

1 Introduction

Research in image/video quality assessment (QA) has emerged in the last years, substantial effort has been dedicated to the development of new image QA methods. There are numerous applications for automatic machine vision quality estimators. A possible field of application is the media production, delivery and archiving process, where humans have to monitor the contents' visual quality. The Human Visual System (HVS) is able to quickly judge the quality of visual content even if the original image or video is not present.

Subjective quality assessments provide the most reliable results, because in many applications the end user is the human observer. In practice, these methods are too costly and time-consuming. It is obvious that objective techniques are needed to predict the video/image quality automatically. So called perceptual metrics try to estimate the quality as perceived by an averaged viewer.

Objective metrics can be categorized, based on the presence of a distortion-free original image for comparison. Full-reference metrics need the distorted image and the original image as input. Common metrics are the Mean Squared Error (MSE) or the Peak Signal-to-Noise Ratio (PSNR). Those metrics are simple to calculate, however they do not represent the perceived visual quality very well. Many approaches have been developed, involving the characteristics of the HVS. A perceptual alternative of those two basic methods is the Structural SIMilarity index (SSIM) [1], exhibiting consistency with the qualitative visual appearance. Several full-reference quality assessment algorithms are evaluated in [2]. Furthermore an extensive classification, review and performance comparison of recent video QA Methods, following the ITU-T recommendations for objective quality evaluation, can be found in [3].

Reduced-reference QA methods work with only partial information of the original visual signal available, i.e. a reduced feature dataset. Obviously in most of the practical applications the reference signal is unavailable. Hence no-reference metrics are needed to blindly quantify the image distortions, if no knowledge of the original image is available.

Numerous video QA methods and metrics have been proposed in the literature, including several full-reference approaches, showing promising correlation with the human visual perception. Contrary most no-reference QA techniques do not correlate well with perceived quality, and therefore continue to be areas of active research. A typical application for image quality metrics

is the media production and delivery process [4]. Broadcasters have to monitor video quality while recording, after editing or compression and before transmission. At recording no original reference image is available, therefore no-reference metrics could execute this job. After media editing or compression a full-reference metric could be applied to check the quality before transmission. If real-time assessment is required, e.g. at the receiver, with no original video data available, no-reference methods with low-computational complexity are needed.

No-Reference metrics are typically designed for a specific type of distortion (e.g. blur, noise, ringing, blocking), while full-reference metrics are able to assess the effect of several distortions. Various distortion types may occur during acquisition, processing, transmission and storage of digital content.

Our work concentrates on blur assessment in images, the most common distortion type in digital image/video processing. To extend this for digital video, blur can be estimated in every frame, or rather in every n -th frame. Objective QA algorithms face many challenges. One of them is to be independent from image content. Another one is to distinguish between image features and impairments. Artifacts can occur in all technical image/video processing stages applied to the analogue content up to its digitization. A detailed overview of different impairments, generated in video processing cycle can be found in [4].

Finally the visual content is interpreted by the HVS in a way that can not be modelled computational (at least at the current state of research). The HVS is able to estimate blur well. Although, due to the high complexity of the HVS and cognitive aspects of the brain, it is not shown how this mechanism works. The HVS recognizes image impairments, by establishing a reference to the observers' knowledge about natural images. Human observers detect blur so well, because of their imagination of the sharp image in contrast to the blurred one.

We consider sharpness as an effect of blurring, it is multiplicative inversely proportional to image blur. Two metrics, differing in the feature extraction process are presented. One estimates perceptual sharpness with respect to the overall image quality, the other assesses perceptual sharpness regardless of perceptive quality. The advantage of the latter method is a more precise prediction of the relative sharpness. A possible application would be a process, where sharpness is just altered slightly, e.g. broadcasters investigating, if the received video content has been upscaled.

Our method follows the idea of measuring the edge widths, by analyzing the spread of the edge slopes, proposed by Marziliano *et. al* in 2002 [5]. We have chosen this idea because of its simplicity and low computational complexity, to make real-time sharpness estimation possible. Actually many

sharpness/blur metrics (e.g [6, 7, 8, 9]) base on this method.

The proposed metric is compared against four no-reference sharpness metrics which are well known in the literature. For performance comparison two publicly available subjective image databases are used. A subjective evaluation is done by humans rating the content. The result is a mean opinion score (MOS), representing the quality rated by different subjects.

This thesis is organized as in the following. Section 2 provides an overview of recent sharpness metrics. We also present two basic methods using local statistical features, followed by a description of the methods we used for evaluation. In section 2.3, some issues of current no-reference metrics are demonstrated. Section 3 gives a definition of our sharpness metric based on local gradient analysis. Experimental results, using different quality assessment databases with a large set of subjective scores, are provided in section 4. The thesis is concluded in section 5.

2 Overview of Existing Metrics

Several objective no-reference metrics exist in the literature. Some are transform based, knowing that sharp edges increase the high frequency components, while blurring acts like a low pass filtering in the frequency domain. Spatial domain methods extract local or global image features and evaluate specific statistical properties, e.g. variance, autocorrelation or edge spread. Before using the edge spread as main feature, we also tested other local spatial features for sharpness measurement.

In this section we will first present these basic methods using autocorrelation and variance features. Followed by a description of the existing no-reference metrics used for comparison to our main approach, which is presented in section 3. A more comprehensive review of existing objective no-reference sharpness metrics can be found in [8].

2.1 Proposed Basic Metrics

2.1.1 Autocorrelation Metric

Autocorrelation can be used to measure the similarity of the image with itself. The autocorrelation function allows to find repeating patterns in images. Blurred images (with large smooth regions) have high autocorrelation values, while sharp images contain less correlated patches. A very basic method to determine the sharpness in an image is to compare neighboring image patches. Blurring leads to an increase in correlation between adjacent pixels.

Our method divides the image into 16×16 blocks and computes the cross-correlation of every 16×16 block with its' 8-blocks-neighborhood.

2.1.2 Variance Metric

The observation of sharp images having greater variation in pixel values than blurred ones can be used to measure the sharpness. Another benefit of variance-features is the robustness to noise. In our implementation we divide the image into 16×16 blocks first, then the local variance of each block is calculated by

$$\tilde{\sigma}^2 = \sum_{x=0}^n \sum_{y=0}^m (B(x, y) - \bar{B})^2 \quad (1)$$

\bar{B} denotes the mean of the block B . A sharp block tends to have high variance, while smooth, blurred blocks exhibit low variance. The global sharpness of the image is determined by averaging the local variances. Only 15% of the sharpest blocks have influence on the result.

2.2 Existing No-reference Sharpness Metrics

We compare the performance of our metric with popular no-reference sharpness metrics, two located in the frequency domain and two spatial domain metrics.

One metric is a transform based method, proposed by Bovik *et al.* [10], using statistics of the discrete wavelet transform (DWT) coefficients in natural images, to produce quality scores for JPEG2000 compressed images. A more recent sharpness metric, located in the wavelet transform domain, analyzing the local phase coherence (LPC) of complex wavelet coefficients is proposed in [11]. Ferzli *et al.* [8] and Narvekar *et al.* [9] propose spatial domain sharpness metrics, based on a concept of just noticeable blur. Their algorithms base on the analysis of edges and adjacent regions in images proposed by Marziliano *et al.* [7].

2.2.1 Spatial Domain Metrics

A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)

The JNB is conceptualized by reverse engineering on the human visual system. Using studies of human blur perception, Ferzli *et al.* [8] model the

degree of blurring in an image patch by

$$P_{BLUR}(e_i) = 1 - \exp\left(-\left|\frac{w(e_i)}{w_{JNB}(e_i)}\right|^\beta\right). \quad (2)$$

$w(e_i)$ is the measured edge spread as presented in [7], $w_{JNB}(e_i)$ is the "just noticeable blur" width, which is obtained in subjective experiments by using a blurred edge with different Gaussian standard deviations. By blurring the edge width increased and a blurring was noticed by the subjects, in dependence to the local edge contrast C , at following widths:

$$w_{JNB} = \begin{cases} 5, & \text{if } C \leq 50, \\ 3, & \text{if } C > 50. \end{cases} \quad (3)$$

Ferzli *et al.* approximate the human foveal region with 64×64 image blocks. They calculate P_{BLUR} for every block, if more than 0.2% of the total number of the pixels in the block were classified as edges. The probability of detecting a blur distortion in each of those blocks, P_{BLUR} , is used in a Minkowski metric to derive an overall image sharpness.

A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)

This method extends the JNB Metric, by only using blocks with

$$CPBD = P(P_{BLUR} \leq P_{JNB}). \quad (4)$$

Hence, only image patches which a human subject would not recognize as blurred are used for this metric.

2.2.2 Frequency Transform Domain Metrics

Blurring can be considered as a low pass filtering of the image, higher spectral components are suppressed in blurred images. Sharp images tend to have more fine detail, resulting in an increase of the high frequency components. Although noise also increases the high frequency components. This oppositional effect makes methods based on this property very susceptible to noise. One requirement for our metric is a low noise susceptibility. For that reason we locate our metric in the spatial domain.

No-Reference Quality Assessment Using Natural Scene Statistics: JPEG2000 (NSS)

Bovik *et al.* [10] propose a method to assess image quality of JPEG2000 compressed images. This compression is based on the DWT. Due to quantization small DWT coefficients become zero, resulting in ringing and blurring artifacts. High compression causes a higher probability of zero coefficients in different subbands. Bovik *et al.* gather statistics of the DWT coefficients of JPEG2000 compressed (natural scene) images. The main parameters are C , the wavelet coefficients' magnitude and P , the linear predicted magnitude of the coefficient. They expect natural images to have significant P and C values, which get 'insignificant' by the compression process. The method is especially designed for JPEG2000 compressed images, hence it exhibits a low performance for other blur types.

No-Reference Image Sharpness Assessment Based on Local Phase Coherence Measurement (LPC-SI)

Assuming that blur causes a disruption of local phase i.e. it yields to a loss of phase coherence. Wang *et. al* [12] demonstrate that precisely localized features, e.g. sharp edges, cause a strong local phase coherence in the complex wavelet transform domain. Typical blurring means a convolution with a low pass filter, averaging out rapid changes in intensity. A convolution with such a filter wont change the phase information in the global Fourier domain. Wang *et. al* show, that *local* phase information changes. An auspicious sharpness metric, based on LPC was proposed recently [11]. It uses the effect of the LPC exhibiting a constant relationship near sharp image features, e.g. edges. This relationship is disrupted by image impairments that affect sharpness.

2.3 Issues of Existing No-reference Sharpness Metrics

The main concern in existing no-reference sharpness metrics is the high susceptibility to altering image content, demonstrated for the JNB metric in figure 1. This issue is also shown in figure 2 where the correlation of the

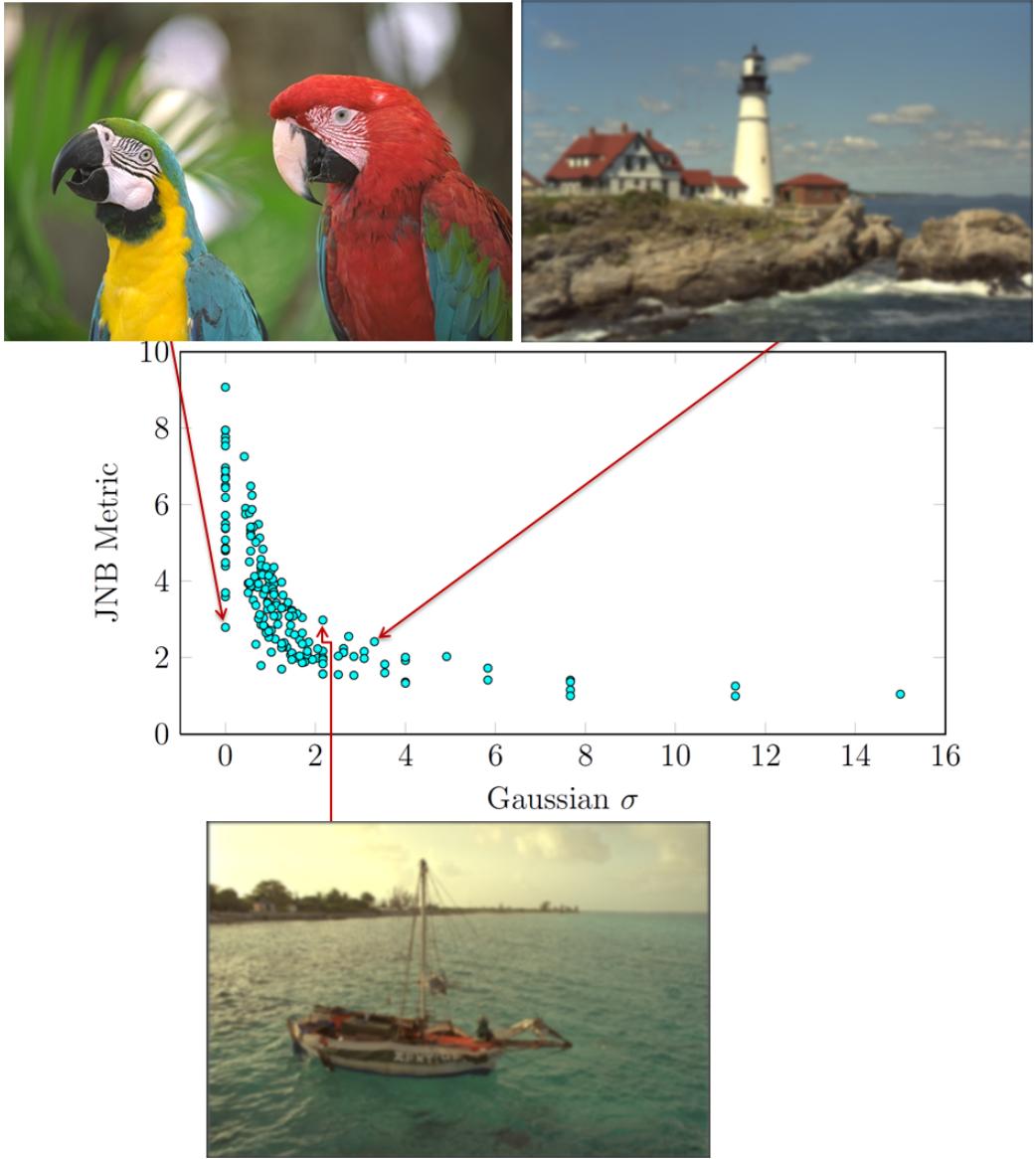


Figure 1: 29 reference images of the LIVE image quality database, low-pass filtered by using a circularly symmetric 2-D Gaussian kernel with a standard deviation of σ , resulting in 174 images. Each data point represents an image. The high variance of the JNB Metric for diverse image content is shown. The untouched parrots image ($\sigma_1 = 0$) is assessed with approximately the same sharpness as the $\sigma_2 = 2.17$ sailing image and the $\sigma_3 = 2.73$ blurred lighthouse image. Our proposed metric assesses a sharpness of $s_1 = 0.68$, $s_2 = 0.16$ and $s_3 = 0.12$, respectively

JNB, as well as CPBD Metric with the subjective tests are shown. The CPBD Metric achieves better performance than the JNB Metric, but also struggles with altering image content. Another disadvantage of the CPBD method is the rapid descent of the metric for increasingly blurred content. More precisely, in Figure 10 on page 26 it is shown that the CPBD Metric converges to zero for nearly all blurred LIVE database images with $\sigma > 2$ pixels. This is because no more image patches with $P_{BLUR} \leq P_{JNB}$ exist in the image, i.e. no more patches, which human observers would consider as 'not blurred'. P_{JNB} denotes the degree of blurriness, which is just noticed as blurred by a human observer. Hence the metric can not be used for medium to highly blurred images.

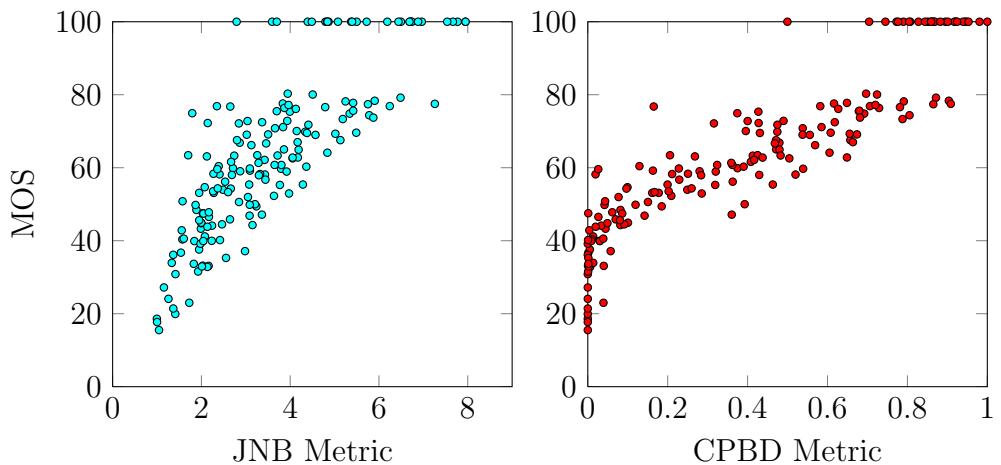


Figure 2: Scatter plots of JNB and CPBD Metric and mean opinion scores of the LIVE Gaussian blur database. A high variance in the 29 reference images ($MOS = 100$) can be observed in both plots. The CPBD metric shows a better performance, though it non-linearly saturates to zero for images with a MOS lower than $\approx 40\%$.

The same issue can be found in the no-reference sharpness metric, based on local phase coherence (LPC-SI). A similar saturation is visualized in the scatter plot in [11], on page 2436, figure 3. Both, the CPBD and the LPC-SI metric, assess a sharpness of zero for images with subjective mean opinion scores of $\approx 40\%$.

As these metrics provide the best results for no-reference sharpness/blurriness measurement at the current state of research, it is obvious that a more accurate method for sharpness estimation in image and video is needed.

3 Proposed Sharpness Metric Based on Local Gradient Analysis

3.1 Relevant Feature Extraction

Our assumption of all image distortions being uniformly distributed, in every color channel, is sustained by the effect shown in [13]. The MSE vectors, each vector having 3 (R,G,B) dimensions, of the proposed *full-reference* MSE-metric [13] concentrate around a $R = G = B$ line. Therefore we apply our metric in the luminance component only. Note that all luminance values in this document refer to 8 bit images, i.e. 256 grey-levels.

It is important to locate appropriate measuring points, representing the relevant edges in the image. The main idea is to measure the spread of all edges detected by use of a Sobel filter (horizontal S_x and vertical S_y) [14]. If no edges are detected, e.g. for blackframes in video, a sharpness value of zero is assigned.

$$G_x = S_x * I = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * I \quad (5)$$

$$G_y = S_y * I = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I \quad (6)$$

By combining these two results, we get the magnitude of the gradient, i.e. the rate of intensity change at each point in the image.

$$G = \sqrt{G_x^2 + G_y^2} \quad (7)$$

A threshold is applied to the magnitude. For performance increase the square root wont be applied and the threshold is squared.

A good predictor for human sharpness perception can be obtained, if the sensitivity threshold of the Sobel method is chosen adaptively, so only the most significant edges in the image are measured. By taking the mean value of the magnitude into account, the threshold can be computed by

$$thresh_{adaptive} = \alpha \sqrt{\bar{G}}, \quad (8)$$

where \bar{G} denotes the mean of the gradients magnitude and α is a scaling factor. In our experiments we set the parameter to $\alpha = 2$. The adaptive threshold leads to a focus on edges with high gradient magnitude (in relation to the magnitudes' mean). If the metric is used as an overall quality predictor, all reasonable edges have to be analyzed. The reason is that artifacts, induced by image compression, e.g. blur and ringing in JPEG2000, do not occur homogeneous, i.e. the whole image is not degraded uniformly.

Therefore a constant Sobel threshold of $thresh = 2.3$ is applied for our second metric. The result of the Sobel filtering and thresholding of a JPEG2000 compressed image can be seen in figure 3. Even with high compression some regions in the image stay sharp, e.g. the roof of the house in figure 3(a). So a human observer would typically rate the perceived sharpness of the image as 'fair', due to its sharp regions. While when assessing the perceived quality of image 3(a), one would judge it as 'poor'. This is because of humans tend to assess the sharpness based on the sharpest region in the image. Hence the threshold after applying the Sobel filter is an important parameter for our metric and we derive 2 metrics:

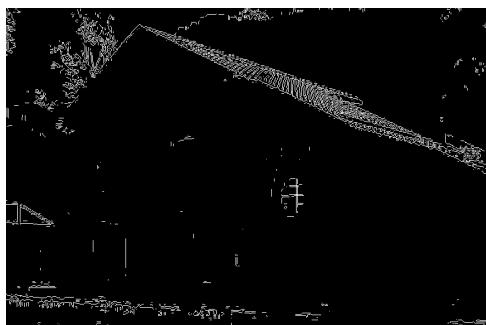
- Metric 1 - predictor of perceived image sharpness.
→ Adaptive Sobel thresholding (8)
- Metric 2 - predictor of overall perceived image quality, characterized by the severity of image blur.
→ Constant Sobel thresholding

After thresholding a thinning process is performed, to get a binary edge image, providing the measuring points for our metric. The thinning process can be considered as a non-maximum suppression, where the local neighborhood across the magnitude is analyzed. Because edges are expected to continue along an outline, values smaller than the neighbors are preferably suppressed if the non-maximum points reside perpendicular to the edge direction, rather than parallel to the edge.

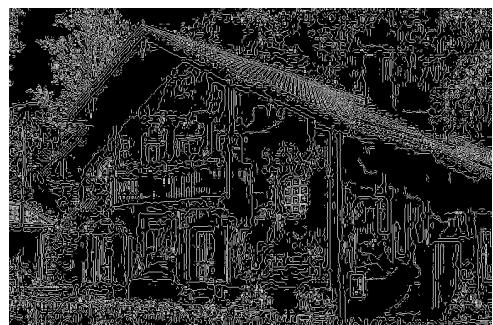
We apply a hysteresis threshold, with a lower threshold set to $\frac{1}{3} \cdot thresh$. The hysteresis thresholding immediately rejects values below the lower threshold, but accepts points between the two limits, if they are connected to pixels with strong magnitude. Figure 3(b) shows the gradients magnitude, after application of an adaptive threshold (8) and thinning. In figure 3(c) a constant threshold was applied before thinning.



(a)



(b)



(c)

Figure 3: Painted house image of LIVE image quality database, demonstrating the effect of different gradient magnitude thresholds: (a) JPEG2000 compression with a bitrate of 0.24 bits/pixel. (b) thresholded G using equation (8). (c) G after applying a constant threshold of 2.3. Image (a) has a mean opinion score of 44%, rated by 54 observers. Assessed sharpness is 0.64 by Metric 1 and 0.33 by Metric 2.

3.2 Measuring Edge Widths

Before measuring the edge widths, the gradients' direction has to be determined. For that purpose the horizontal and vertical pixel differences are calculated. The resulting images $I_x = \frac{\partial I}{\partial x}$ and $I_y = \frac{\partial I}{\partial y}$ form the gradient vectors in each point of the image. The angle of the gradient is calculated by

$$\Phi = \arctan \left(\frac{I_y}{I_x} \right). \quad (9)$$

At every measuring point the closest local extrema to the edge pixel are found by iterating along the gradient, perpendicular to the edge. The edge width is defined by the pixels between the local minimum and maximum.

Gradients are measured horizontal and vertical, with an angle tolerance of $\Delta\phi$. Measuring the width of the diagonal edges resulted in a performance decrease. The reason is assumed to be the larger quantization step-width ($\sqrt{2}$) of diagonal pixels, compared to a spacing of one between pixels in horizontal/vertical direction.

The final edge width is computed by

$$w = \frac{w_{up} + w_{down}}{\cos(\Delta\phi)}, \quad (10)$$

where w_{up}, w_{down} are the pixels between the detected edge and the local maximum or minimum, respectively. $\Delta\phi$ denotes the angle difference between the gradients' direction and the measured direction, e.g. if we measure a vertical edge with an angle of $\phi = 93^\circ$, then $\Delta\phi = 3^\circ$. Though gradients with $\Delta\phi > \Delta\phi_{max}$ are not measured at all. In our experiments we set this parameter to $\Delta\phi_{max} = 8^\circ$. We ignore all edges at the image borders (i.e. indent of 32 pixels). Additionally a width is rejected, if the corresponding gradient starts or ends at an image border. The iteration process, starting at the edge

pixel detected by the Sobel operator, is illustrated in figure 4. The figure also shows, that we allow minor intensity changes (maximum luminance change of 2) against the gradients' slope, two for each measurement. Additionally, after the gradients' slope changes, the added width is rejected if the next extrema is found close(≤ 2 pixels) to the first extrema.

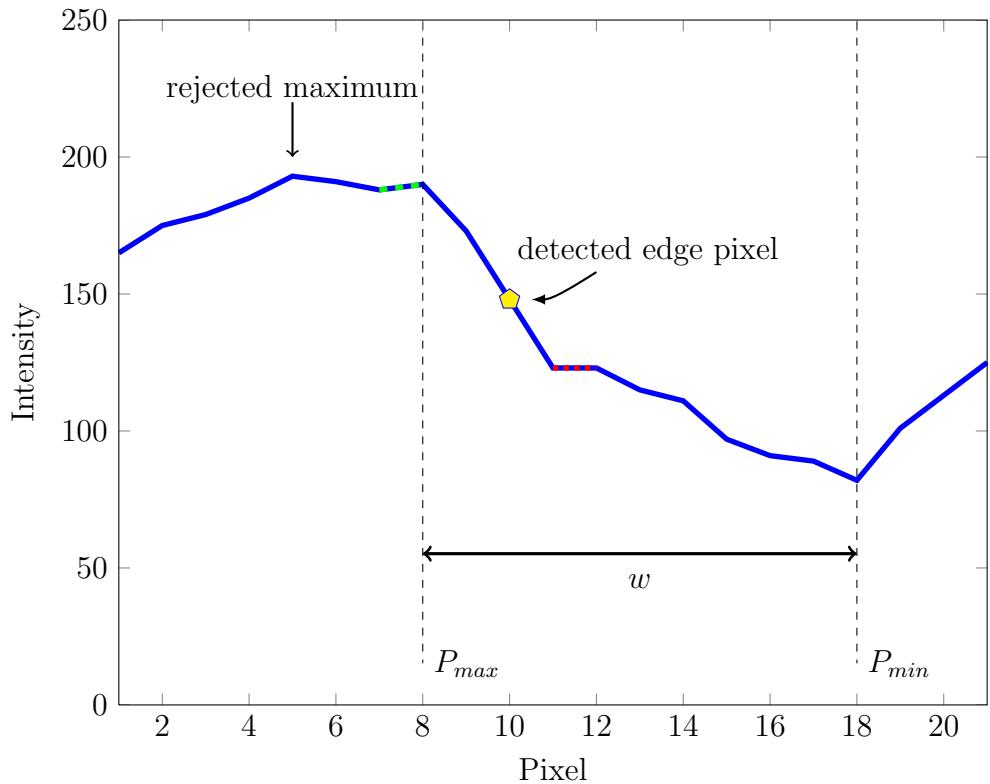


Figure 4: Image intensity curve along the gradient. P_{max} , P_{min} represent the detected local maxima, w the edge width measured. The dotted red line between pixel 11 and 12 denotes no intensity change along the gradient. The green loosely dotted line denotes a intensity change against the gradients direction.

3.3 Subpixel Accurate Local Extrema Estimation

The spacing of one between each pixel limits the precision of the local extrema estimation. So we added a subpixel-accurate approximation of the local extrema by fitting a polynomial of degree 2 and use of the first derivative. This procedure is illustrated in figure 5 for the pixel intensities of 190, 220, 210 and 100, 220, 210 respectively.

It is illustrated that the difference between the maximum pixel and the estimated maximum increases proportional to the local slope of the gradient ($d_2 > d_1$). After testing this feature, our results showing the correlation to the mean opinion score of different human subjects, diminished. This is explained by the human visual system perceiving edges with a high change of intensity (e.g. 100 to 220) sharper than low changes in intensity (e.g 190 to 220).

Hence we did not add d to our edge widths, but *subtracted* it, making a higher local gradient slope result in a smaller (sharper) width. This modification showed a higher correlation with the human perception of sharpness.

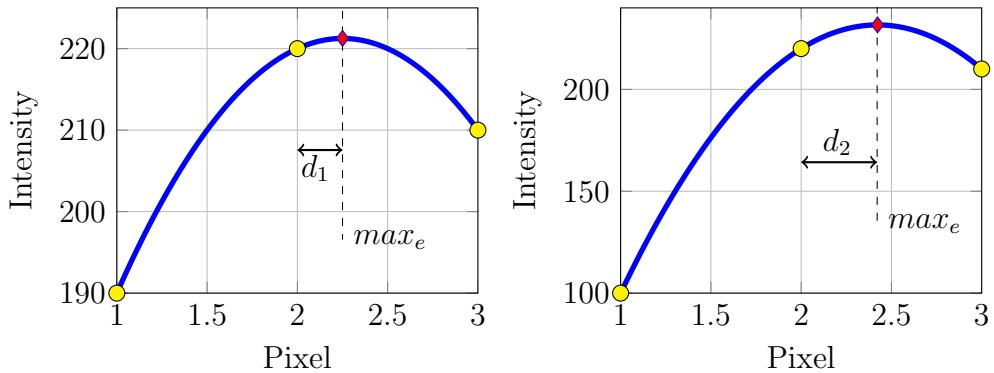


Figure 5: Subpixel-accurate extrema estimation, three pixel neighborhood. d_1 and d_2 are the differences from the maximum pixel to the estimated maximum max_e . The maximum approximated by an interpolated polynomial. Note that max_e is always closer to the highest of the two neighbor pixel values.

3.4 Modeling the Human Perception of Acutance

Motivated by the observation of the HVS perceiving edges with high contrast as sharper, we refine our metric by taking the gradients slope into account. The slope is determined by

$$\text{slope} = \frac{I_{\max} - I_{\min}}{w}. \quad (11)$$

Where I_{\max} and I_{\min} denote the local extrema along the gradient and w is the width, i.e. the amount of pixels between I_{\max} and I_{\min} .

The measured edge widths are decreased in proportion to the slope by

$$w_{\text{new}} = \begin{cases} w - \frac{\text{slope}}{\delta}, & \text{if } w > 2, \\ w & \text{otherwise,} \end{cases} \quad (12)$$

δ specifies the impact of the gradients' slope. In our experiments we used values in the range of $\delta \in [150, 800]$ and for all results shown in this paper we used the parameter $\delta = 500$.

The smallest possible edge width occurs if the gradients' local minimum and local maximum are neighbor pixels ($w = 1$). To imply the Just Noticeable Difference (JND) of the HVS, we recalculated the widths only for $w > 2$. The JND is defined as the smallest difference between two levels of a particular sensory stimulus (e.g. edge width, in respect to the two intensity levels) to produce a noticeable variation in sensory experience. In [8] a subjective test procedure was realized, estimating that an edge is perceived as blurred by the HVS if the width is greater than 5 pixels for a contrast smaller than 50, and 3 pixels for a higher contrast. For that reason and because of the slope being contrast-dependent, we only recalculated edges with a minimum spread of $w > 2$.

Overall the attempt resulted in an increase in correlation with the subjective ratings, especially the *linear* relationship between our metric and the subjective scores elevated. But also the variance of the measured sharpness for images with diverse content (but the same sharpness) increases slightly, this is illustrated in figure 6. If robustness to alterable content is the main requirement, this feature should not be added to the metric.

We have also experimented with the slope of a fitted first order polynomial to all values of the gradient. This yielded to similar results, at higher computational complexity.

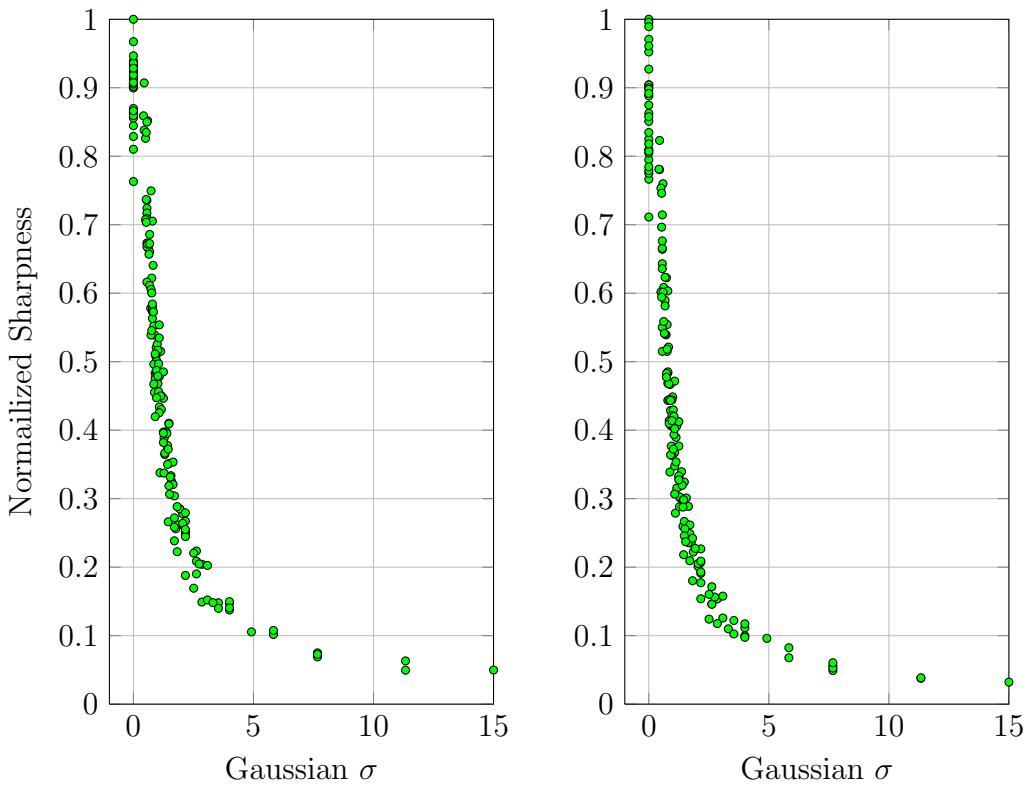


Figure 6: Scatter plot, showing the determined sharpness of the 174 Gaussian blurred images of the LIVE Database (including 29 reference images with $\sigma = 0$). The left scatter plot shows the metric without taking acutance into account. In the right plot local edge contrast change was considered as described in sections 3.3 and 3.4. The parameter $\delta = 300$ was used.

3.5 Block Based Calculation

The image is divided into 32×32 blocks to quantize the focused region and to deduce local sharpness values. The representative edge width for each block is calculated by averaging all measured widths in a block. Block B is only processed further if the following condition is fulfilled:

$$w_{sum} \geq w_{min} \quad (13)$$

where w_{sum} is the sum of all widths in B and w_{min} is a threshold, we tested with $w_{min} \in [2, 5]$, and used a value of 2 for the final metric. This constraint requires more than one narrow edges to be measured in a block. Sharp edges are able to generate a higher relative error due to the quantization step of one. Hence (13) increases the robustness to noise.

The local sharpness in all of the image blocks is then converted to a final sharpness value by selecting only the sharpest blocks. The average widths of all n blocks, meeting requirement (13), are sorted and only the k sharpest blocks are used to determine the sharpness. Parameter k is set to $k = \lceil 0.15 \cdot n \rceil$ and $k = \lceil 0.45 \cdot n \rceil$ for Metric 1 and Metric 2, respectively. Focusing on the sharpest blocks ignores out of focus regions and increases perceptual sensing.

The overall image sharpness is finally deduced by inverting the mean of the k sharpest average block widths

$$s = k \sum_{i=1}^k \frac{1}{\bar{w}_i}, \quad (14)$$

k represents the number of sharpest blocks and \bar{w}_i is the average of all measured edge spreads in block B_i .

Figure 7(a) illustrates the local sharpness estimation of the 512×768 monarch image, obtained from the LIVE image quality database, having different blur amounts in figure 7(c)-(d). The results of the proposed metric, as well as the mean opinion scores are listed in Table 1.

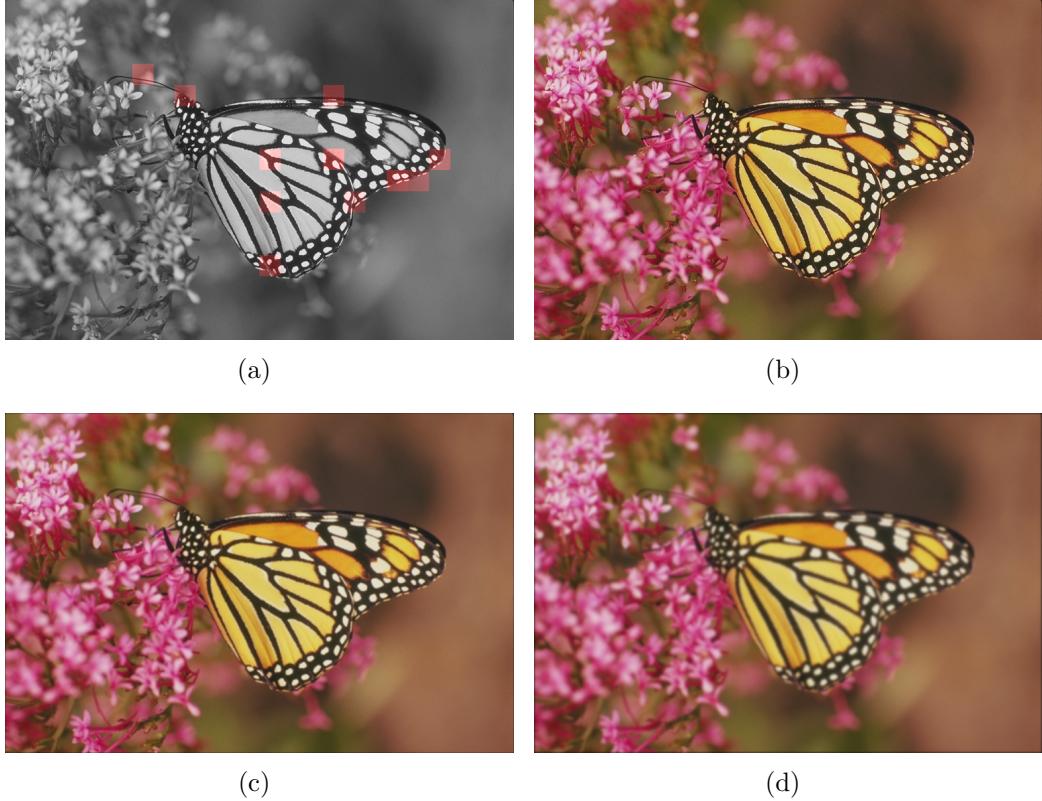


Figure 7: Localized block based sharpness estimation, the intensity of the red blocks in (a) correspond to the sharpness of the underlying image patches. (b) Original monarch image convoluted by using a 2-D Gaussian kernel with standard deviation σ (d) 0.9 and (d) 1.85, respectively. The corresponding mean opinion scores of 24 observers are listed in table 1.

	σ_{blur}	Metric 1	MOS
(b)	0	0.59	100
(c)	0.9	0.31	76.76
(d)	1.85	0.17	58.18

Table 1: Performance of the proposed metric on the LIVE monarch image. A monotonic decrease of the metric, with increasing blur amount σ_{blur} is shown.

4 Experiments

Due to subjective experiments being costly and time consuming, the values provided in two well known quality assessment databases are used for evaluation. We compared our results with several other novel metrics. The algorithm performances were either taken from the correlating papers or the publicly available implementations.

4.1 Subjective Quality Assessment

For comparison of the different metrics, subjective quality scores are needed. In this work we used the Mean Opinion Score (MOS) and Differential Mean Option Score (DMOS) values, provided in the LIVE Database [15]. The MOS is determined by showing the subjects the original and the distorted image. Then the user has to rate the quality on a 5 to 1 scale: excellent, good, fair, poor and bad. The more ratings a human subject performs, the more constant his rates get. Although the tests should not take too long, because of lack of concentration. DMOS is the difference between the MOS of the original image and the MOS of the degraded image. Typically subject rejection is performed on the raw opinion scores, if more than k evaluations of a subject were outliers.

4.2 Subjective Visual Quality Assessment Databases

The reference images, used for a quality assessment database, need to fulfill some requirements. Typically the images have to provide a high variety in image content. Particularly local image characteristics, like contrast and intensity should vary a lot. For example some images should contain a high amount of different textures, smooth regions or sharp edges. A short description of the two databases, used to test the metrics' performance, is provided in this section.

4.2.1 LIVE Quality Assessment Database - LIVE[15]

Prof. Alan C. Bovik, The University of Texas at Austin, USA

The database consists of 29 images taken primarily from the *Kodak Lossless True Color Image Suite* test set. The reference images were distorted by JPEG2000, JPEG, White Gaussian Noise (WGN) and Gaussian blur. All distortions were applied in the RGB components. Our performance evaluation

was realized on the 174 Gaussian blurred images, filtered by using a circular-symmetric 2-D Gaussian kernel of standard deviation σ . We also evaluated the metrics' susceptibility to noise, using the 174 WGN images, distorted with WGN of standard deviation σ . Because of the JPEG2000 compression causing mainly blurring and ringing artifacts, the 227 JPEG2000 images, compressed with different bitrates, were also used to measure the performance. 24, 23 and 54 observers participated in the subjective tests, for the blurred, noisy and JPEG2000 compressed images, respectively.

4.2.2 TID2008 Database [16]

Nikolay Ponomarenko, National Aerospace University of Ukraine

This database consists of 25 reference images, distorted in 17 different types, with 4 different levels for each type of distortion. This results in 1700 test images, therefore it is the largest database for evaluation of quality metrics. As reference images, similar to the LIVE database, the images of the *Kodak* test set were used. The images were cropped to a size of 512×384 pixels. Additionally an artificial image was added, containing objects with different characteristics and texture fragments.

Mean opinion scores have been obtained in experiments with more than 800 observers participating. The relevant distortions for our experiments were Gaussian blurred images and JPEG2000 images.

4.2.3 Database Comparison

While the TID2008 database offers scores, determined by plenty of subjective experiments, the experiments of LIVE were conducted under normalized viewing conditions. The viewing distance was kept at about 2-2.5 times of the screen height, where 21-in CRT monitors displaying at a resolution of 1024×768 pixels were used [15].

The LIVE database also provides a higher number of distortion intensities, as shown in figure 8. Where the woman-hat image of the LIVE database and the cropped version of the image, used in the TID2008 database, are shown. The images are blurred with different blur amounts. Mean opinion scores were conducted in various subjective experiments. Note that LIVE database provides difference scores (DMOS), and therefore associates the score for the original image with the scores of the degraded versions.



Figure 8: Blurred woman-hat image of LIVE database (a)-(f) and TID2008 database (g)-(j) with subjective MOS. (MOS = 100 - DMOS for LIVE)

4.3 Performance Measures

We have chosen the evaluation procedures suggested in the Video Quality Experts Groups' (VQEG) [17] work. They suggest a non-linear pre-mapping of the predictors' results to the subjective ratings. Unlike than in other public evaluations, we kept the linear correlation coefficient in our evaluation to provide a more critical judgement. The following performance measures are used to evaluate the proposed method.

- Pearson Correlation Coefficient

The **LCC** describes the linear dependence of two variables (i.e. the objective and subjective scores) and therefore it is a sign for how *accurate* a prediction is. The LCC between two data sets x and y is given by

$$LCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (15)$$

For the following performance measures a prior non-linear mapping on the predicted scores is performed.

To get a linear relationship between the objective scores and the subjective ratings a non-linear mapping function, as suggested by the VQEG [17], is used. In order to make our metric comparable with current state-of-the-art no-reference metrics, evaluated in [9], the four-parameter logistic function of [9] was applied for the non-linear fitting procedure

$$MOS_{pi} = \frac{\beta_1 - \beta_2}{1 + e^{(\frac{metric_i - \beta_3}{|\beta_4|})}}. \quad (16)$$

MOS_{pi} is the predicted metric value $metric_i$, after non-linear regression analysis. The model parameters $\beta_1 - \beta_4$ are chosen for a best fit to the corresponding subjective MOS scores.

- Nonlinear Pearson Correlation Coefficient

NLCC is the linear correlation coefficient, after non-linear regression analysis. Therefore it also provides an indication of prediction *accuracy*.

- Spearman Rank-order Correlation Coefficient

Expressing the *monotonicity* of the prediction, the **SROCC** tries to describe the relationship between the objective/subjective results, by use of a monotonic function. The Spearman coefficient ignores the relative distance between the data, only the ordering of the prediction is checked. Hence the SROCC can be considered as a LCC between the ranks of the two data sets.

- Root Mean Square Error and Mean Absolute Error are given by

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2} \quad (17)$$

and

$$MAE = \frac{1}{n} \sum_{i=1}^n |\gamma_i - \hat{\gamma}_i| \quad (18)$$

The **RMSE** is the square root of the sum of squared residuals. A residual of a sample is the difference between the sample value γ_i (e.g MOS) and the predicted value $\hat{\gamma}_i$. While the **MAE** is a linear score, weighting individual differences equal in the average, the RMSE adds high weights to large errors, i.e. outliers. MAE and RMSE also describe the predictions' *accuracy*.

- Outlier Ratio (**OR**), is the fraction of predictions outside of the interval $[MOS - 2\sigma, MOS + 2\sigma]$ to the total number of predictions. *MOS* is the mean opinion score and σ is the standard deviation of the opinion scores for a single image. The OR expresses the *consistency* of the prediction.

4.4 Content Dependency

Being robust to alterable content is one of the most important requirements for image sharpness metrics. A full robustness for a metric means, that an undistorted image A , blurred by σ_{gblur} , and image B , blurred by σ_{gblur} , result in the same sharpness s . Figure 9 shows several images, with different contents. In figure 10 the 174 images of the LIVE Gaussian blur database and the corresponding sharpness, determined by different metrics, is shown. A content-dependence-comparison of the proposed metric (without consideration of acutance), as well as CPBD and JNB metrics is illustrated. All of the 174 LIVE Gaussian blur images are measured and plotted in 10. Many different images have similar blur amounts σ . And, of course, all of the 29 reference images have a blur value of $\sigma = 0$. The Sum of Absolute Distances (SAD) to an interpolated polynomial, after normalizing the metric scores to $[0, 1]$, is given as a performance measure.

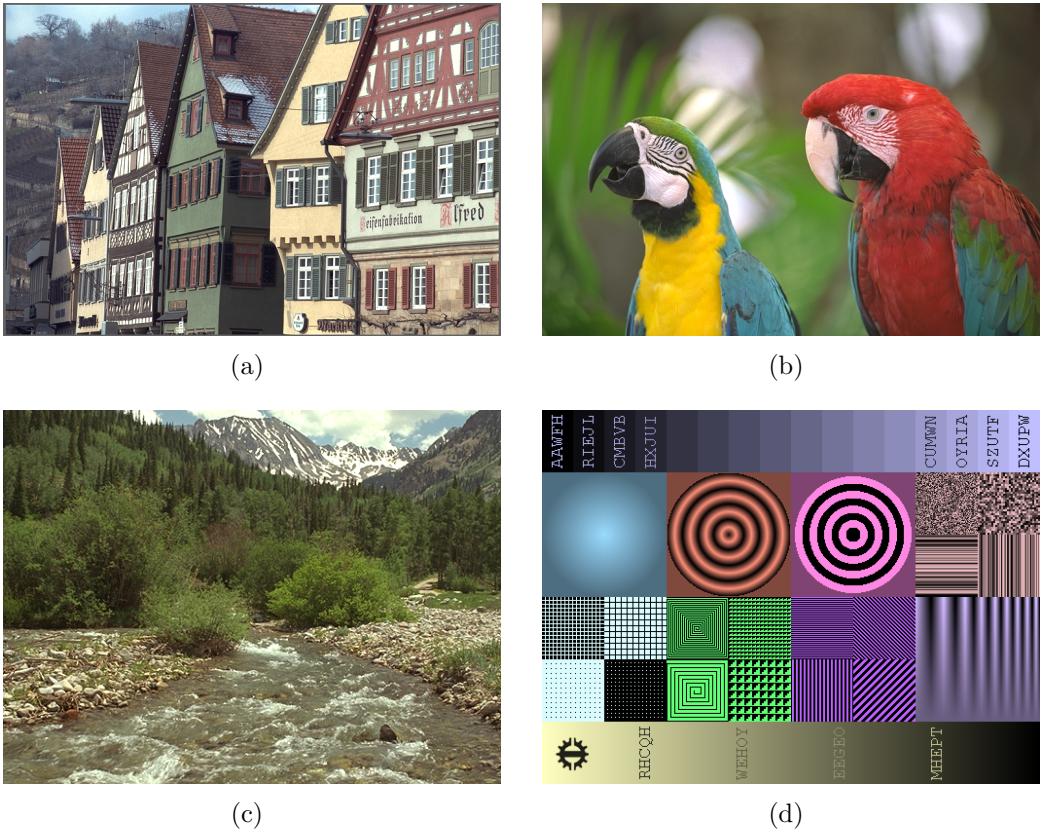


Figure 9: Diverse content in LIVE database (a), (b) and TID2008 database (c), (d) is an artificial image of TID2008 with different texture fragments.

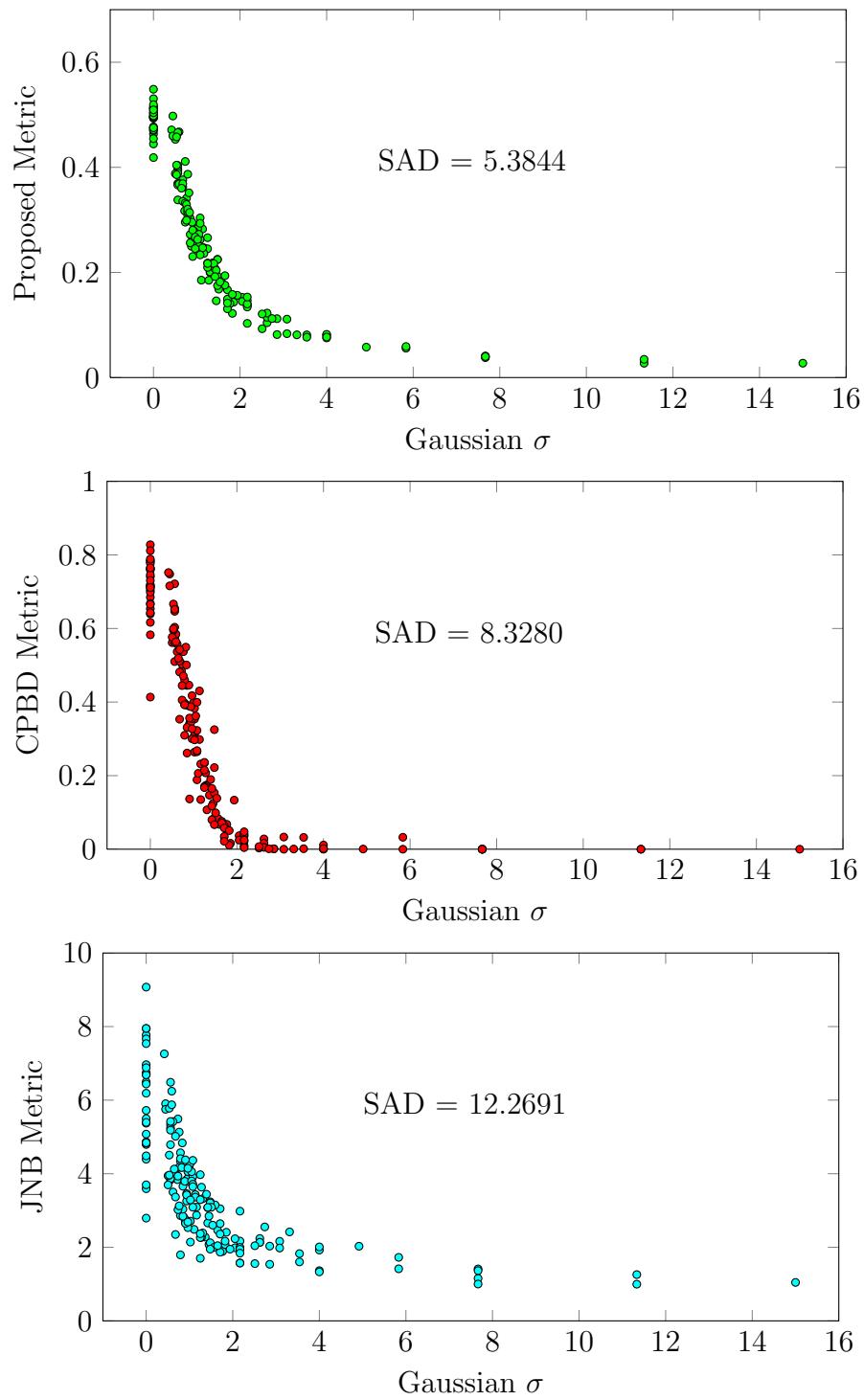


Figure 10: LIVE gblur measures with SAD to an interpolated polynomial.

4.5 Noise Susceptibility

To assess the susceptibility to noise, the proposed metric was tested on the 174 White Gaussian Noise images of the LIVE database. Figure 11 illustrates the WGN LIVE images and the measured sharpness. A noise robust metric would predict the same sharpness for every image.

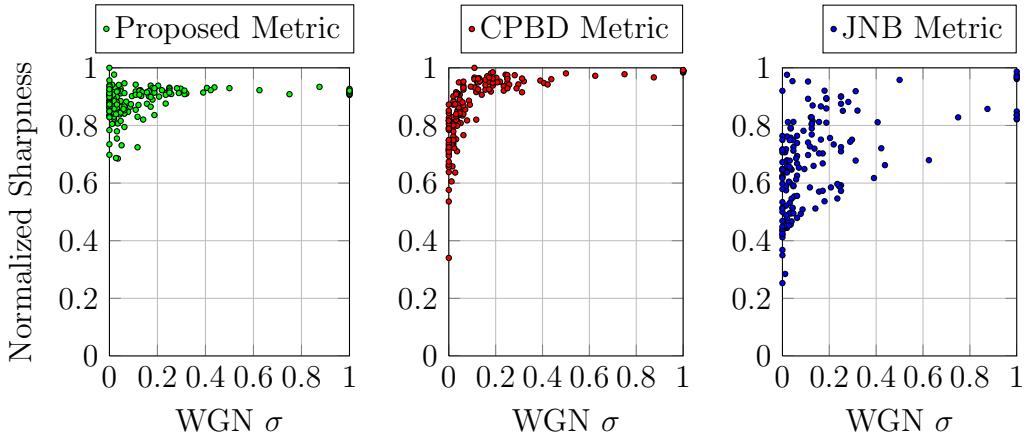


Figure 11: LIVE noise measures, 174 distorted images with WGN of standard deviation σ . It is shown that the values of all metrics increase with increasing noise level. The reason is that sharp edges are introduced by noise.

4.6 Blocking Susceptibility

A common distortion in video/image compression, is blocking. Blocking artifacts are artificial discontinuities in images generated by block-wise quantization in several block-based coders (e.g. JPEG, MPEG-2, H.264). Those impairments increase the measured sharpness by adding artificial edges to the images. Due to our focus on only the sharpest edges in an image, blocking artifacts can be very falsifying, especially for heavily blurred images.

Blocking impairments typically appear in a regular form (e.g. in a grid of 8x8 pixels). In our tests we created artificial 8x8 blocks in the LIVE Gaussian blur test set by adding random intensity $\Delta i \in [-10, 10]$ to each pixel in the block. Experiments showed that only at high blurring ($\sigma_{gblur} \geq 5$) our metric gets susceptible to blocking impairments. The reason is blocking-edges mainly affecting our result. If robustness to blockiness is vitally important,

a better result could be achieved by using a blocking detection method and then simply omit the measured edge widths at local blocking borders. Alternatively a blocking metric, e.g [18], could be used to measure the magnitude of extracted blocking impairments. Thus the metrics sharpness can be decreased in respect to the assessed blockiness.

4.7 Performance Results

The proposed sharpness metric is compared with the existing metrics in section 2, also the results of the Marziliano's and our basic metrics are listed. We calculate the LCC, NLCC, SROCC, OR, RMSE and MAE for performance comparison. Higher LCC, NLCC and SROCC indicate a higher correlation between the metrics' values and the subjective scores, while lower OR, RMSE and MAE indicate a higher prediction accuracy. Two subjective databases, listed in section 4.2, are used. The OR can not be calculated for the TID2008 database, because no standard deviation of the subjective ratings is available. The different RMSE and MAE scales are due to the altering scale of the estimated MOS, e.g. the LIVE database provides MOS ranging from 0 to 100, while TID2008-MOS are between 0 and 9. The proposed metric demonstrates good and stable performance and is the best performer on most of the datasets. Figure 12 shows the scatter plot of MOS versus metric prediction, an almost linearly relation is given.

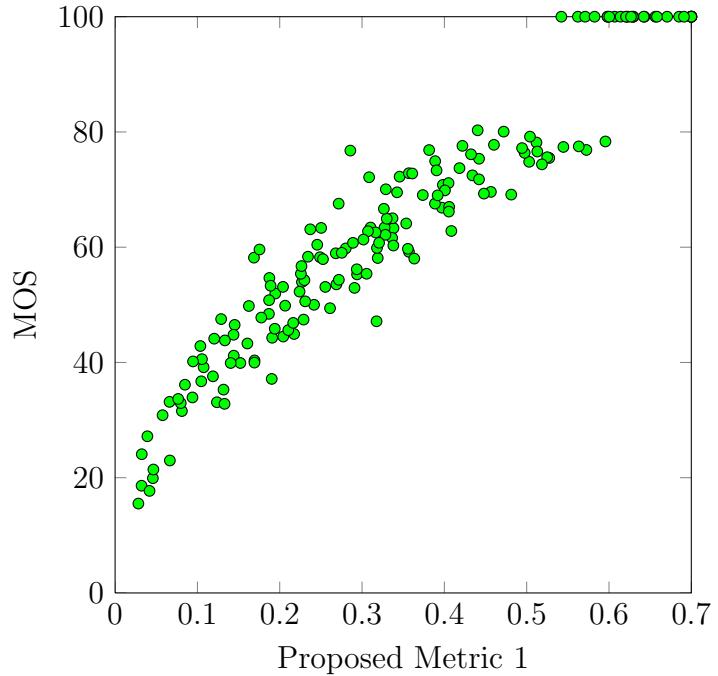


Figure 12: Scatter plot of the proposed sharpness metric on the LIVE Gaussian blur database (no non-linear mapping applied). Each data point represents a test image, with a mean subjective score (y-axis) and an objective metric output (x-axis).

As noted in section 3, Proposed Metric 2 distinguishes from Metric 1 by a constant gradient magnitude threshold and a higher usage of image blocks for the final sharpness computation.

4.7.1 LIVE Database

LIVE Gaussian Blur Dataset						
Metric	LCC	NLLCC	SROCC	OR	RMSE	MAE
Proposed Metric 1	0.9600	0.9627	0.9626	0.0460	5.8831	4.6482
Proposed Metric 2	0.9389	0.9449	0.9452	0.0862	7.1221	5.6271
CPBD [9]	0.9127	0.9127	0.9431	0.1609	8.8889	6.8227
LPC-SI [8]	N/A	0.9239	0.9368	0.1724	8.525	6.9335
JNB [8]	0.8222	0.8428	0.8423	0.2356	11.7061	9.2404
NSS jp2k metric [10]	0.1801	0.1975	0.3377	0.5000	21.3215	16.9850
Marziliano <i>et al.</i> [7]	-	0.8597	0.8659	0.2184	11.1106	8.2743
Variance, 2.1.2	0.5916	0.0213	0.5893	0.5000	21.7502	17.4564
Autocorrelation, 2.1.1	0.3968	0.3771	0.4683	0.5057	20.1444	15.8333
SSIM, <i>full-reference</i> [1]	N/A	0.9487	0.9519	N/A	5.8225	N/A
LIVE JPEG2000 Dataset						
	LCC	NLCC	SROCC	OR	RMSE	MAE
Proposed Metric 2	0.8701	0.8974	0.9012	0.2203	10.7647	8.7701
CPBD [9]	0.8658	0.8823	0.8859	0.2511	11.4825	9.0782
LPC-SI [8]	N/A	0.4233	0.388	0.6167	22.1024	18.4596
JNB [8]	0.7002	0.7081	0.7159	0.4449	17.2251	13.9603
NSS jp2k metric [10]	0.9137	0.9209	0.9153	0.1894	9.5080	8.3387
Marziliano <i>et al.</i> [7]	N/A	0.7815	0.7744	0.2184	15.2196	0.4670
Variance, 2.1.2	0.1137	0.1137	0.1413	0.6123	24.3951	21.1076
Autocorrelation, 2.1.1	0.3649	0.3264	0.3643	0.6123	19.3194	19.3194

Table 2: Performance comparison on the LIVE database, the boldface entries indicate the best performer. As shown, our metric performs superior on uniformly blurred content, even comparable to recent *full-reference* image quality metrics. An evaluation of *full-reference* image quality assessment algorithms, is provided in [2]. As shown the structural similarity (SSIM) index, comparing local patterns in images, produces a NLCC of 0.9487.

4.7.2 TID2008 Database

TID2008 Gaussian Blur Dataset				
Metric	NLCC	SROCC	RMSE	MAE
Proposed Metric 1	0.8537	0.8492	0.6112	0.4927
Proposed Metric 2	0.8278	0.8359	0.6584	0.5170
CPBD [9]	0.8235	0.8412	0.6657	0.5173
LPC-SI [8]	0.8113	0.803	0.6778	0.5202
JNB [8]	0.6931	0.6667	0.8459	0.6529
NSS jp2k metric [10]	0.3367	0.2761	1.1049	0.9237
Marziliano <i>et al.</i> [7]	0.709	0.7165	0.8176	0.6466
Variance, 2.1.2	0.4427	0.4502	1.0522	0.8439
Autocorrelation, 2.1.1	0.3275	0.4307	1.1088	0.9182

TID2008 JPEG2000 Dataset				
Metric	NLCC	SROCC	RMSE	MAE
Proposed Metric 2	0.9301	0.9287	0.7169	0.5652
CPBD [9]	0.9223	0.925	0.7406	0.5831
LPC-SI [8]	0.7952	0.7295	1.1621	0.9356
JNB [8]	0.8798	0.8789	0.9111	0.7213
NSS jp2k metric [10]	0.3222	0.3999	1.8521	1.6030
Marziliano <i>et al.</i> [7]	0.8667	0.8694	0.9561	0.7127

Table 3: Performance comparison on TID database. The proposed metrics exhibit a good performance. Note that the JPEG2000 quality metric of [10] shows a low correlation for this database. The reason is assumed to be the training of this QA algorithm on half of the LIVE data set.

The difference in performance on the databases can be attributed to the variance in the distortion strengths, contents or subjective judgement. More precisely, the better performance of the metrics on the LIVE database can be explained by the wider range of distortion strengths, making subjective and objective prediction easier. An example is shown in figure 8.

5 Conclusion

An effective sharpness estimation algorithm, utilizing HVS characteristics, is presented. We have shown that the proposed metric outperforms existing no-reference sharpness metrics. The effectiveness of the proposed method is validated by subjective tests on Gaussian-blurred and JPEG2000 compressed images. The metric resolves some issues in existing metrics, major advantages are a more accurate sharpness prediction and a lower susceptibility to diverging image content. As evaluated in [2] our metric is even comparable to *full-reference* image quality assessment algorithms for uniformly blurred content.

For applications without reference to the HVS, e.g. an objective sharpness comparison of different images/videos, the features in section 3.3 and 3.4 may be omitted. Hence more consistent results for diverging image/video content can be achieved. The presented method also exhibits a low computational complexity, making real-time video sharpness assessment possible (i.e. the C++ implementation needs $\approx 25ms$ for a standard-definition-frame). Furthermore unlike than in recent sharpness metrics, it is possible to effectively use the metric for highly blurred content. Possible applications are the detection of out of focus blur, due to camera defocus, or motion blur, caused by relative motion between the camera and the scene. In our experimental results the metric shows a consistently decrease for Gaussian blur up to a σ of 15 pixels, as well as a high accuracy, monotonicity and consistency.

We confute an argument of Hsin *et al.* [19], that the edge width may not be accurately measured in moderately blurred content. The experimental results show the superiority of our method compared to recent metrics. Future work includes a weighting method for the measured edge widths and the submission of a four page excerpt of this thesis to the ICIP 2012.

6 Acknowledgment

The author would like to thank his colleagues at JOANNEUM RESEARCH, DIGITAL for the good collaboration, especially Peter Schallauer and Hannes Fassold for their support and feedback. Thanks as well to Prof. Horst Bischof and Georg Thallinger for making this work possible. The research in this work has been partially supported by the PrestoPRIME project (FP7-ICT-231161), funded by the 7th Framework program of the European Union.

References

- [1] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 –612, april 2004. [2](#), [30](#)
- [2] H. Sheikh, M. Sabir, and A. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *Image Processing, IEEE Transactions on*, vol. 15, no. 11, pp. 3440 –3451, nov. 2006. [2](#), [30](#), [32](#)
- [3] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165 –182, june 2011. [2](#)
- [4] P. Schallauer, H. Fassold, M. Winter, W. Bailer, G. Thallinger, and W. Haas, “Automatic content based video quality analysis for media production and delivery processes,” in *SMPTE Technical Conference and Exhibition*, oct. 2009. [3](#)
- [5] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “A no-reference perceptual blur metric,” in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 3, 2002, pp. III–57 – III–60 vol.3. [3](#)
- [6] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, “A no-reference quality metric for measuring image blur,” in *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, vol. 1, july 2003, pp. 469 – 472 vol.1. [4](#)
- [7] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “Perceptual blur and ringing metrics: application to jpeg2000,” *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163 – 172, feb. 2004. [4](#), [5](#), [6](#), [30](#), [31](#)
- [8] R. Ferzli and L. Karam, “A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb),” *Image Processing, IEEE Transactions on*, vol. 18, no. 4, pp. 717 –728, april 2009. [4](#), [5](#), [16](#), [30](#), [31](#)
- [9] N. D. Narvekar and L. J. Karam, “A no-reference image blur metric based on the cumulative probability of blur detection (cpbd),” *Image Processing, IEEE Transactions on*, vol. 20, no. 9, pp. 2678 –2683, sept. 2011. [4](#), [5](#), [23](#), [30](#), [31](#)

- [10] H. Sheikh, A. Bovik, and L. Cormack, “No-reference quality assessment using natural scene statistics: Jpeg2000,” *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1918 –1927, nov. 2005. [5](#), [7](#), [30](#), [31](#)
- [11] R. Hassen, Z. Wang, and M. Salama, “No-reference image sharpness assessment based on local phase coherence measurement,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 2434 –2437. [5](#), [7](#), [9](#)
- [12] Z. Wang, E. P. Simoncelli, and H. Hughes, “Local phase coherence and the perception of blur,” in *in Adv. Neural Information Processing Systems (NIPS03)*. MIT Press, 2004, pp. 786–792. [7](#)
- [13] F. Zhang, S. Li, and L. Ma, “Limitation and challenges of image quality measurement,” in *Proc. SPIE Visual Communications and Image Processing Conf.*, 2010, pp. 1–8. [10](#)
- [14] I. E. Sobel, “Camera models and machine perception,” Ph.D. dissertation, Stanford, CA, USA, 1970. [10](#)
- [15] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, “Live image quality assessment database release 2.” [Online]. Available: <http://live.ece.utexas.edu/research/quality> [1](#), [20](#), [21](#)
- [16] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, *TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics*, vol. 10, pp. 30 –45, nov. 2009. [Online]. Available: www.ponomarenko.info/tid2008.htm [1](#), [21](#)
- [17] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment.” 2000. [Online]. Available: <http://www.vqeg.org/> [23](#)
- [18] Z. Wang, A. Bovik, and B. Evan, “Blind measurement of blocking artifacts in images,” in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 3, 2000, pp. 981 –984 vol.3. [28](#)
- [19] C. Hsin, J.-W. Jang, S.-J. Shin, and S.-H. Chen, “A no-reference objective image sharpness metric based on a filter bank of gaussian derivative wavelets,” in *Multimedia Technology (ICMT), 2011 International Conference on*, july 2011, pp. 3362 –3365. [32](#)