Improving the robustness to image scale of the total variation of difference metric

Marius Pedersen and Ivar Farup

***Abstract*— Objective image quality assessment has received a lot of attention in the last decades, and it is still an unsolved challenge. One of the problems with many existing image quality metrics is that they suffer from scale differences, i.e. images have been rated similar by observers but according to the image quality metrics the images are different. We propose a normalization step as a solution to this problem for one of the state-of-the-art metrics, the Total Variation of Difference (TVD) metric. The normalization is similar to Michelson contrast, and experimental results show that the proposed normalization significantly increases the performance of the TVD metric.**

I. INTRODUCTION

Objective image quality assessment has been an active area of research for many decades[1], [2]. These objective image quality metrics have the goal of being correlated with perceptual image quality. However, the search for an image quality metric correlated with the percept is still on-going.

One direction that several researchers have taken to achieve image quality metrics predicting perceived image quality is to develop models that simulate the human visual system [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], and use these models as a pre-filtering stage to image quality estimation. It has been shown that image quality metrics incorporating models of the human visual system are more robust than those who do not [13].

Pedersen et al. [10] proposed in 2011 a metric based on perceptual contrast filtering. The method was based on previous work by Peli [14] on contrast filtering and color total variation [15]. The contrast filtering was further refined through the use of wavelets by Pedersen and Farup [16]. However, wavelets are not good at representing discontinuities, which are common in natural images. Therefore, Pedersen et al. [17] extended the filtering by replacing wavelets with contourlets. In 2014 Pedersen [18] proposed the last enhancement to the filtering method by using shearlets, which provided the most accurate filtering compared to the other methods. The filtering was used in the Total Variation of Difference (TVD) metric, and it showed improved performance over metric without simulation of the human visual system. The TVD metric also has the advantage of being possible to optimize, so that it can be used to enhance image quality for different applications, such as image compression and denoising.

Many metrics today, including TVD, suffers from image scale differences [19], [20], [21], [1], [22], [18], [21]. Scale differences occur because images have been rated similar by observers but according to the image quality metrics the images are different. This results in many metrics having a low performance when evaluated on entire databases, but in fact they might perform well for one single image (i.e. one original with different distortions) or for a single distortion (i.e. different original images with the same distortion). It is therefore wanted to have a metric that is robust to scale differences, which is the focus of this paper. We propose a normalization step for image quality metrics, more specifically the TVD metric, to make the metric more robust to scale differences. Increased robustness should also lead to an increased performance of the image quality metric.

The paper is organized as follows; first relevant background, then we introduce the proposed normalization to deal with image scale differences, followed by the experimental evaluation. Results are presented next, before the conclusion and future work.

II. BACKGROUND

The TVD metric follows a general framework for low-level based image quality metrics, where the original and reproduction is transformed into a suitable color space, before the images are filtered to simulate the human visual system, the difference between the original and reproduction is calculated, and at last the quality values are pooled into a single value.

A. Color space transformation

The original and reproduction are transformed to the Ybr color space. Starting with a sRGB input image, which is linearized to remove the gamma correction, and further transformed to CIEXYZ. The next step consists of transforming from CIEXYZ to a new RGB space, where the primaries are defined to match the wavelengths of the monochromatic gratings used in the measurements of the chromatic contrast senstivity functions [23]. Red is defined as 602 nm, green is defined as 526 nm, and blue is defined as 470 nm. Finally, the last transformation is from the new RGB space to Ybr:

$$Y = Y_r R + Y_g G + Y_b B, \tag{1}$$

where R, G, and B are the new linear RGB values, Y_r, Y_g, and Y_b are the $\bar{y}$ values from the color matching functions for the red, green, and blue channels, such that Y is the CIE luminance. The color channels are defined as:

$$b = \frac{Y_b B}{Y} \quad \text{and} \quad r = \frac{Y_r R}{Y}.$$

Marius Pedersen and Ivar Farup is with the Faculty of Computer Science and Media Technology, Norwegian University of Science and Technology (NTNU).

With the images transformed into this color space we can apply the contrast sensitivity functions directly, since the channel contrasts correspond to the contrast definitions used in the measurement of the chromatic contrast sensitivity functions. Also, the channels in Ybr are decorrelated.

B. Simulation of the human visual system

In order to simulate the human visual system, the work by Peli [14] is adopted. The image is decomposed and the contrast between the bandpass and lowpass information is found. The decomposition is carried out using shearlets [24], and is done for each color channel. Pedersen [18] suggested to decompose the images at four scales, with 2, 3, and 4 shear levels, and standard parameters otherwise. We use the same parameters. The decomposition results in a set of low pass filtered coefficients (LL), and many sets of highpass coefficients from different levels, shearings, and cones. The highpass coefficients are filtered with contrast sensitivity functions; a luminance contrast sensitivity function is applied to the achromatic channel (Y) and chrominance contrast sensitivity functions to the chromatic channels (b and r). Since shearlets are divided into different orientations, the luminance contrast sensitivity function from [25] that include orientation dependence and surround illumination was used:

$$CSF_L(u) = \frac{C\exp\left(-0.0016u^2\left(1+100/L\right)^{0.08}\right)}{\sqrt{\left(1+\frac{144}{X_0^2}+0.64\left(1+3\sin^2(2\varphi)\,u^2\right)\left(\frac{63}{L^{0.83}}+\frac{1}{1-\exp^{-0.02u^2}}\right)\right)}}, \quad (2)$$

where C is set to 3700, u is the spatial frequency in cycles per degree, L is the luminance in cd/m^2, and X_0^2 is the angular object area in square degrees, and φ is the orientation angle in degrees. f is a multiplicative correction factor used to account for that the visibility of an object can change depending of the surround [25]:

$$f = \exp\left(-\frac{\ln^2\left(\frac{L_s}{L}\left(1+\frac{144}{X_0^2}\right)^{0.25}\right)-\ln^2\left(\left(1+\frac{144}{X_0^2}\right)^{0.25}\right)}{2\ln^2(32)}\right), \quad (3)$$

where L is the luminance of the object, L_s is the surround luminance, and X_0^2 is the object area in square degrees of visual angle.

For the chrominance contrast sensitivity function the functions from Johnson and Fairchild [4] is applied, one for the red-green channel and one for the blue-yellow channel:

$$CSF_C = \alpha_1\exp(-\beta_1 u^{\gamma_1}) + \alpha_2\exp(-\beta_2 u^{\gamma_2}), \quad (4)$$

where u is defined as cycles per degree and the parameters for the red-green and blue-yellow channels are the same as used by Johnson and Fairchild [4], which are found in Table I.

The luminance contrast sensitivity function (CSF_L) is applied to the luminance channel (Y), and the chrominance contrast sensitivity functions (CSF_C) are applied to the chrominance channels (b and r) for each level, cone, and shearing.

TABLE I
PARAMETERS FOR THE CHROMINANCE CONTRAST SENSITIVITY FUNCTIONS FROM JOHNSON AND FAIRCHILD [4].

Parameter	red-green channel	blue-yellow channel
α_1	109.14130	7.032845
β_1	-0.00038	-0.000004
γ_1	3.42436	4.258205
α_2	93.59711	40.690950
β_2	-0.00367	-0.103909
γ_2	2.16771	1.648658

The contrast sensitivity functions are not normalized, and therefore applied directly at the given scale.

We denote $l'_j(x,y)$ as the contrast filtered LL band and $h_{\psi\tau j}(x,y)$ denote the highpass bands at level j for shearing ψ and cone τ. LL is not filtered at the lowest level N, and therefore $l'_N(x,y) = l_N(x,y)$. $a_{\psi\tau j}(x,y)$ denotes the contrast sensitivity filtered version of $h_{\psi\tau j}(x,y)$. The contrast filtered octave bands are defined as follows:

$$h'_{\psi\tau j}(x,y) = \begin{cases} h_{\psi\tau j}(x,y) & \text{if } a_{\psi\tau j}(x,y) > l'_j(x,y) \\ 0 & \text{else} \end{cases} \quad (5)$$

Then the filtered information in ψ shears (i.e. orientations) and τ cones enables the reconstruction of the image for the lowpass filtered version at the next level. In addition, the TVD metric incorporates contrast masking by using an extended intra channel masking model accounting for local activity [26]. This model accounts for the effect that the detectability of one stimulus is influenced by the presence of another stimulus.

C. Quality calculation and pooling

After simulation of the human visual system, the original and reproduction are converted to the log-compressed OSA-UCS color space [27], which is proven to correlate well with calculated differences [28]. The final quality calculation is calculated using the following equation:

$$\text{TVD} = \sqrt{\sum_j\left(\int_\Omega \mid \nabla L_{O_j} - \nabla L_{R_j} \mid dA\right)^2} + \lambda\int_\Omega\sqrt{\sum_j (L_{O_j}-L_{R_j})^2}dA, \quad (6)$$

where $\sqrt{\sum_i(\int_\Omega \mid \nabla L_{O_j} - \nabla L_{R_j} \mid dA)^2}$ is the color total variation term, while $\lambda\int_\Omega\sqrt{\sum_j(L_{O_j}-L_{R_j})^2}dA$ is the color difference term. L_O is the original filtered image and L_R the filtered reproduction. Ω is the image domain, λ is the weighting parameter for the color difference term, and j denotes the color channel. The first term is similar to the color total variation defined by Blomgren and Chan [15], except that the gradient of the difference between the original and reproduction is taken. For the second term, this is the Euclidean color difference.

III. Proposed approach to deal with image scale differences

As mentioned above image scale differences is a problem for many image quality metrics. We propose a modification to the TVD metric in order to make it more robust against scale differences. We will focus on the total variation calculation of the TVD metric (first term in Equation 6), and compare our proposed metric to the total variation calculation of TVD (hereafter denoted as TVD-TV). Normalization of the quality calculation is an approach to deal with scale differences, and we propose to normalize the total variation by the gradient of the sum of the images.

Our normalization is similar to Michelson contrast [29], where the gradient of the difference is divided by the sum of the gradients of the original and reproduction. The final TVD calculation is the following:

$$\text{TVD-TV-Norm} = \sqrt{\sum_j \left(\int_\Omega \frac{|\nabla L_{O_j} - \nabla L_{R_j}|}{|\nabla L_{O_j} + \nabla L_{R_j}|} dA \right)^2}. \quad (7)$$

The proposed approach also has the advantage of being possible to minimize, and can therefore be used in many different applications to optimize quality, such as for image compression, denoising, halftoning, or gamut mapping.

IV. Experimental Setup

We calculate the TVD-TV-Norm with the proposed normalization (Equation (7)) and TVD-TV without the normalization (First term in Equation (6)) on on the Colour Image Database:Image Quality (CID:IQ) [30], [31]. CID:IQ contains 23 original images (Figure (1)), all which have been modified with six distortions; JPEG2000 compression, JPEG compression, blur, Poisson noise, ΔE gamut mapping, and SGCK gamut mapping. The original images modified with these distortions in five levels from low quality to high quality. 17 observers participated in the experiment, which was carried out at two viewing distances; 50 cm and 100 cm. The level of ambient illumination in the experiment was approximately 4 lux. The chromaticity of the white displayed on the color monitor was D65 and luminance level of the monitor was 80 cd/m^2. All settings are suitable for the sRGB color space. The controlled conditions of the subjective experiments allow accurate simulation of the human visual system, which is a core component of the TVD metric.

A. Performance measures

The performance of each metric is calculated as the correlation between subjective scores and the values calculated by the metric. We have used two standard types of correlation. The Pearson's correlation coefficient, which assumes a normal distribution in the uncertainty of the data values and that the variables are ordinal. The Spearman's rank-correlation coefficient, which is a non-parametric measure of association based on the ranks of the data values, that describes the relationship between the variables without making any assumptions about the frequency distribution.

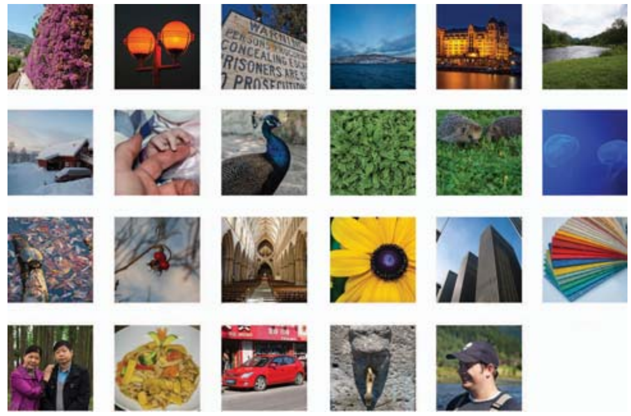

Fig. 1. Images in CID:IQ. The 23 original images have been distorted over 5 levels with six different distortions (JPEG compression, JPEG2000 compression, Gaussian blur, Poisson noise, ΔE gamut mapping, and SGCK gamut mapping.

The relationship between the metrics and subjective scores are not necessarily linear. Therefore, we investigate the correlation using non-linear regression by applying a mapping function [32]:

$$f(x) = \theta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\theta_2 (x - \theta_3)}} \right) + \theta_4 X + \theta_5, \quad (8)$$

where θ_i, $i = 1, 2, 3, 4$, and 5 are the parameters to be fitted. Initial parameters for the fitting are max(subjective scores), min(subjective scores), $median$(metric scores), 0.1, and 0.1. 95% confidence intervals are calculated using Fisher's z transformation [33].

V. Results

The results for non-linear Pearson correlation can be seen in Figure 2 for 50 cm viewing distance, and in Figure 3 for 100 cm viewing distance. For 50 cm viewing distance (Figure 2) the proposed TVD-TV-Norm is statistically significantly better in four out of six distortions given the 95% confidence interval, and have similar performance in the two other distortions. For the full database the proposed TVD-TV-Norm is statistically significantly better than TVD-TV. Similar results are found for 100 cm (Figure 3), where the proposed TVD-TV-Norm is statistically significantly better in three out of six distortions, and have similar performance in the two other distortions. For the full database the proposed TVD-TV-Norm is statistically significantly better than TVD-TV.

The results for Spearman correlation can be seen in Figure 4 for 50 cm viewing distance, and in Figure 5 for 100 cm viewing distance. For three of the six distortions the proposed TVD-TV-Norm is statistically significantly better than TVD-TV, in the other three distortions it has comparable performance. For the full database the proposed method is significantly better. For 100 cm viewing distance the results are similar to 50 cm, with significantly better performance in three distortions and better performance for the full database.

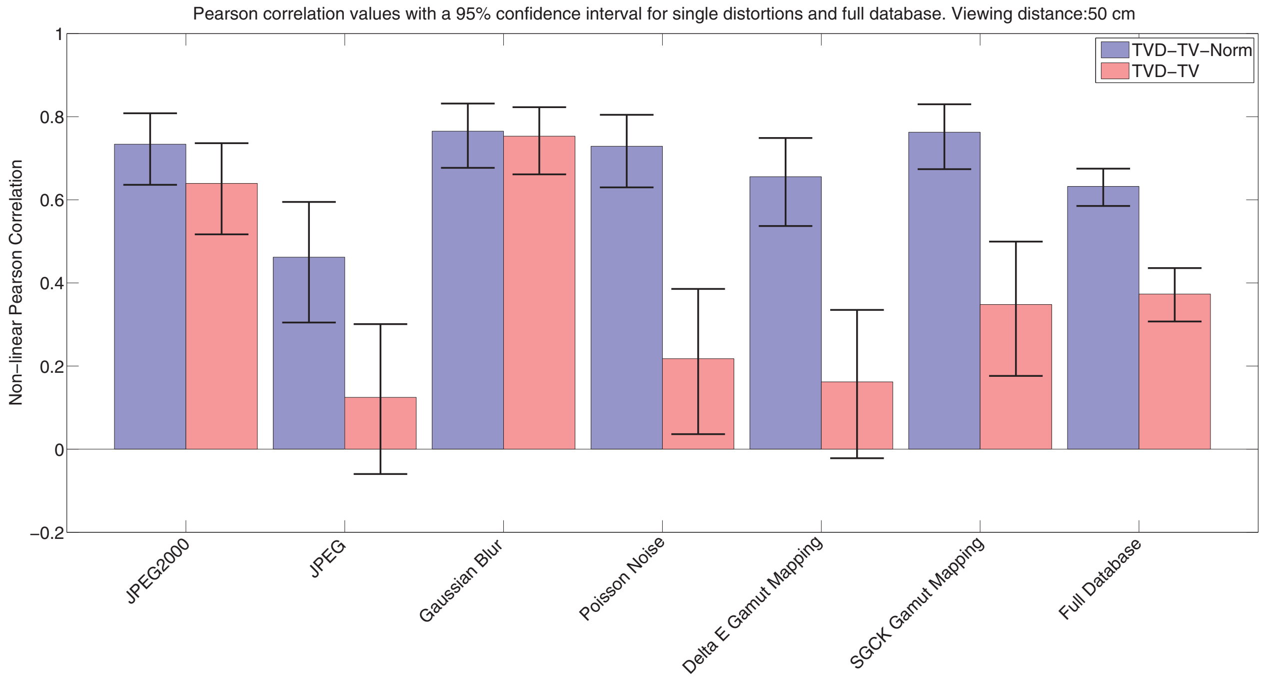

Fig. 2. Non-linear Pearson correlation values for single distortions and full database for the proposed TVD-TV-Norm and the previous TVD-TV for 50 cm viewing distance. The proposed metric is statistically significantly better in four out of six distortions (JPEG, Poisson noise, *Delta*E gamut mapping, and SGCK gamut mapping), and have similar performance in the two other distortions (JPEG2000 and Gaussian blur). For the full database the proposed TVD-TV-Norm is statistically significantly better than TVD-TV. The errorbars indicate the 95% confidence interval.

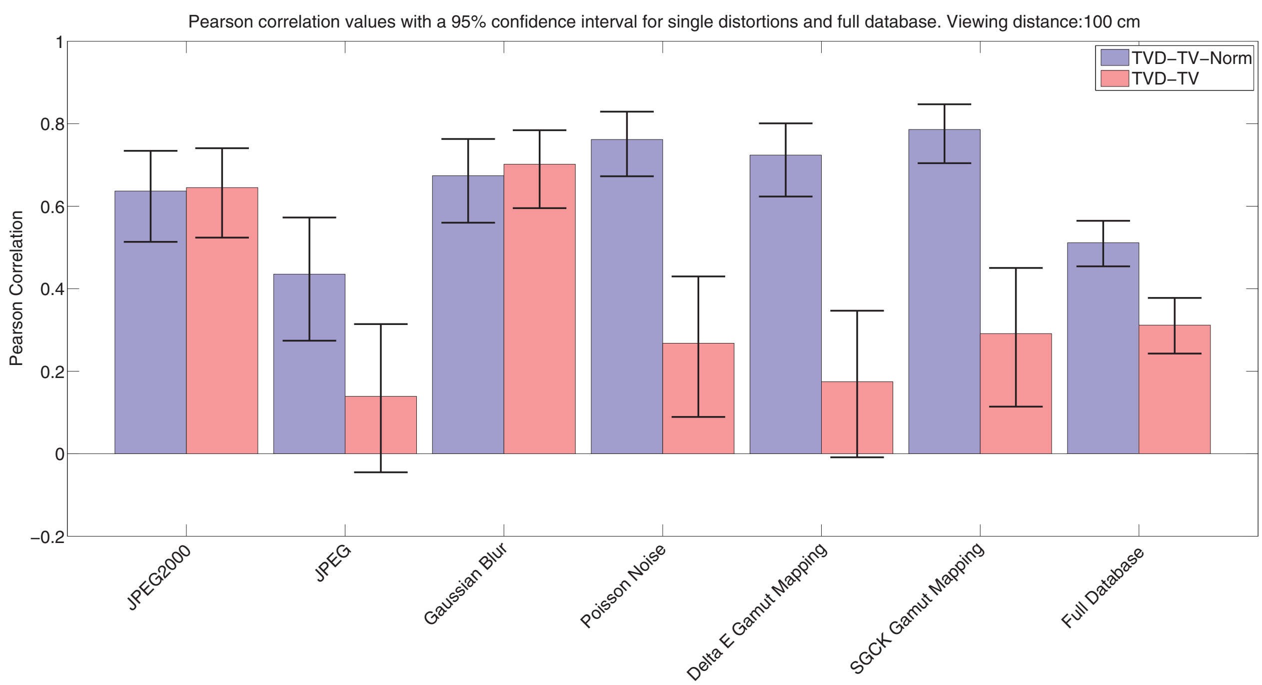

Fig. 3. Non-linear Pearson correlation values for single distortions and full database for the proposed TVD-TV-Norm and the previous TVD-TV for 100 cm viewing distance. The proposed metric is statistically significantly better in three out of six distortions, and have similar performance in the two other distortions. For the full database the proposed TVD-TV-Norm is statistically significantly better than TVD-TV. The errorbars indicate the 95% confidence interval.

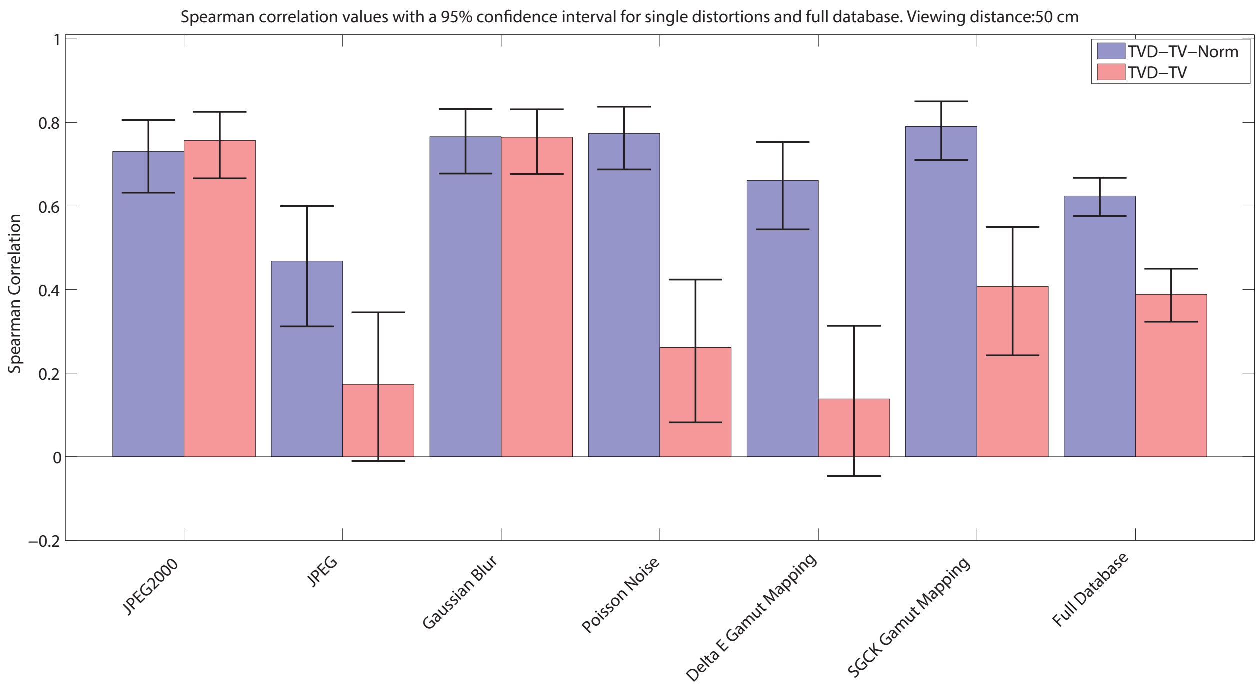

Fig. 4. Spearman correlation values for single distortions and full database for the proposed TVD-TV-Norm and the previous TVD-TV for 50 cm viewing distance. The proposed TVD-TV-Norm is significantly better than TVD-TV in three of the six distortions, and significanly better for the full database. The errorbars indicate the 95% confidence interval.

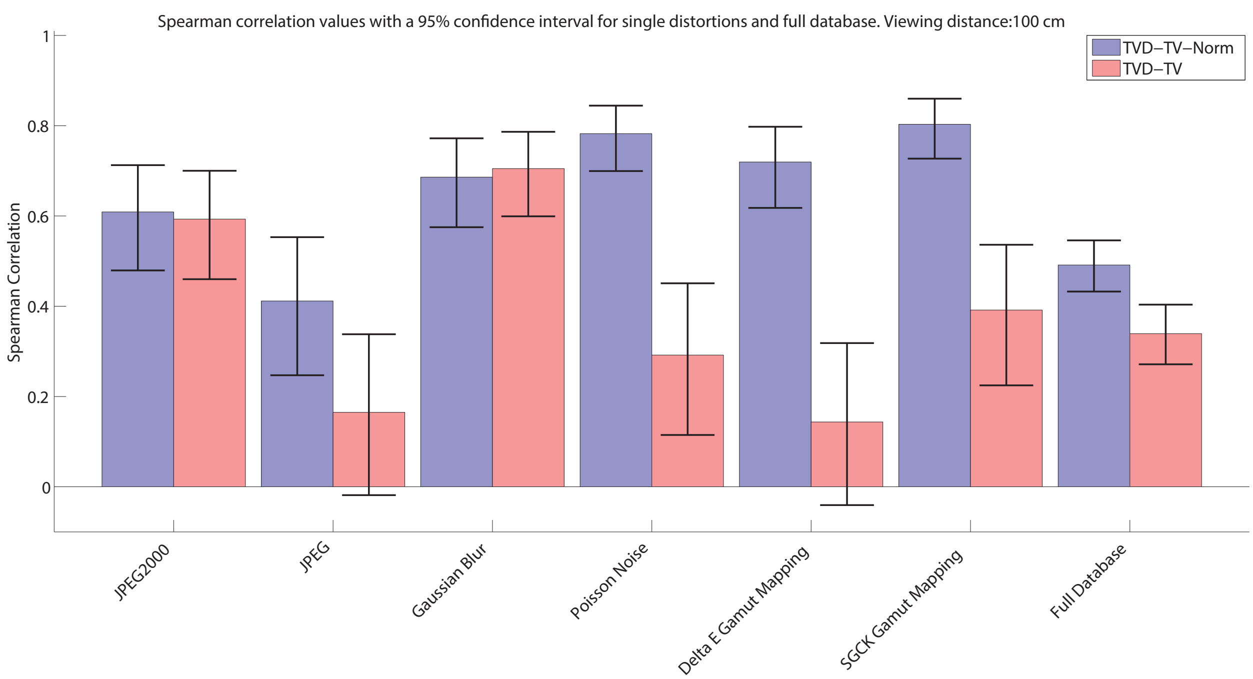

Fig. 5. Spearman correlation values for single distortions and full database for the proposed TVD-TV-Norm and the previous TVD-TV for 100 cm viewing distance. The proposed TVD-TV-Norm is significantly better than TVD-TV in three of the six distortions, and significanly better for the full database. The errorbars indicate the 95% confidence interval.

Overall, the proposed TVD-TV-Norm has a more stable performance (similar correlation coefficients) for the different distortions than TVD-TV. Also, for the full database for both 50 cm and 100 cm viewing distance we see that the proposed TVD-TV-Norm has a better distribution in the regression plot. This is also confirmed by the increased correlation values. This indicates that the normalization step results in better robustness to scale differences.

VI. CONCLUSIONS

We have proposed a normalization step to make the Total Variation of Difference (TVD) image qualtiy metric more robust to scale differences. The proposed approach, named TVD-TV-Norm, is evaluated on the CID:IQ database, and the results show that it is more robust to scale differences. The performance of TVD-TV-Norm is also significantly better or similar to the previous metric. For the full database, the normalization step produces significantly better results.

Future work includes evaluation of TVD-TV-Norm against other image quality metrics. Since the proposed metric can be optimized, we also plan to use it to enhance image quality in different applications, such as image compression or denoising.

ACKNOWLEDGMENT

This research has been funded by the Research Council of Norway through project no. 221073 'HyPerCept Colour and quality in higher dimensions'.

REFERENCES

[1] M. Pedersen and J. Y. Hardeberg. Full-reference image quality metrics: Classification and evaluation. *Found. Trends. Comp. Graphics and Vision*, 7(1):1–80, 2012.

[2] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[3] G. M. Johnson and M. D. Fairchild. On contrast sensitivity in an image difference model. In *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pages 18–23, Portland, OR, Apr 2002.

[4] G. M. Johnson and M. D. Fairchild. Darwinism of color image difference models. In *Color Imaging Conference*, pages 108–112, Scottsdale, AZ, Nov 2001.

[5] Z. Wang and J. Y. Hardeberg. Development of an adaptive bilateral filter for evaluating color image difference. *Journal of Electronic Imaging*, 21(2):023021–1–023021–10, 2012.

[6] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin. On between-coefficient contrast masking of DCT basis functions. In *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07*, pages 1–4, Scottsdale, AZ, Jan 2007.

[7] Z. Wang and J. Y. Hardeberg. An adaptive bilateral filter for predicting color image difference. In *Color Imaging Conference*, pages 27–31, Albuquerque, NM, USA, Nov 2009. IS&T/SID.

[8] X. Zhang and B. A. Wandell. A spatial extension of CIELAB for digital color image reproduction. In *Soc. Inform. Display 96 Digest*, pages 731–734, San Diego,CA, 1996.

[9] M. Pedersen and J. Y. Hardeberg. A new spatial hue angle metric for perceptual image difference. In *Computational Color Imaging*, volume 5646 of *Lecture Notes in Computer Science*, pages 81–90, Saint Etienne, France, Mar 2009. Springer Berlin / Heidelberg.

[10] M. Pedersen, G. Simone, M. Gong, and I. Farup. A total variation based color image quality metric with perceptual contrast filtering. In *International conference on Pervasive Computing, Signal Processing and Applications*, Gjøvik, Norway, Sep 2011.

[11] M. Pedersen and J. Y. Hardeberg. A new spatial filtering based image difference metric based on hue angle weighting. *Journal of Imaging Science and Technology*, 56:50501–1–50501–12(12), September 2012.

[12] G. Simone, C. Oleari, and I. Farup. Performance of the euclidean color-difference formula in log-compressed OSA-UCS space applied to modified-image-difference metrics. In *11th Congress of the International Colour Association (AIC)*, Sydney, Australia, Sep/Oct 2009.

[13] P. Le Callet and D. Barba. A robust quality metric for color image quality assessment. In *International Conference on Image Processing (ICIP)*, volume 1, pages 437–440, Barcelona, Spain, Sep 2003. IEEE.

[14] E. Peli. Contrast in complex images. *J. Opt. Soc. Am. A*, 7:2032–2040, 1990.

[15] P. Blomgren and T. F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, Mar 1998.

[16] M. Pedersen and I. Farup. Simulation of image detail visibility using contrast sensitivity functions and wavelets. In *Color and Imaging Conference*, pages 70–75, Los Angeles, CA, November 2012.

[17] M. Pedersen, X. Liu, and I. Farup. Improved simulation of image detail visibility using the non-subsampled contourlet transform. In *Color and Imaging Conference*, pages 191–196, Albuquerque, NM, USA, Nov. 2013. IS&T.

[18] M. Pedersen. An image difference metric based on simulation of image detail visibility and total variation. In *Color and Imaging Conference*, pages 37–42, Boston, Ma, Nov 2014.

[19] M. Pedersen. Full-reference image quality metrics: Classification and evaluation. CIE Reportership DR R8-10, CIE, 2015-06-19 2015.

[20] M. Pedersen. Evaluation of 60 full-reference image quality metrics on the CID:IQ. In *International Conference on Image Processing*, pages 1588 – 1592, Quebec, Canada, September 2015. IEEE.

[21] M. Pedersen and J. Y. Hardeberg. Rank order and image difference metrics. In *4th European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, pages 120–125, Terrassa, Spain, Jun 2008. IS&T.

[22] C.G. Zewdie, M. Pedersen, and Z. Wang. A new pooling strategy for image quality metrics: Five number summary. In *Visual Information Processing (EUVIP), 2014 5th European Workshop on*, pages 1–6, Dec 2014.

[23] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359:381–400, 1985.

[24] D. Labate, W-Q. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. In M. Papadakis, A. F. Laine, and M. A. Unser, editors, *Wavelets XI*, volume 5914, pages 59140U–59140U–9, San Diego, CA, Jul. 2005. SPIE.

[25] P. G. J. Barten. Formula for the contrast sensitivity of the human eye. In Y. Miyake and D. R. Rasmussen, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5294, pages 231–238, December 2003.

[26] M. Nadenau. *Integration of human color vision models into high quality image compression*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2000.

[27] C. Oleari, M. Melgosa, and R. Huertas. Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space. *Journal of the Optical Society of America A*, 26(1):121–134, 2009.

[28] D. R. Pant and I. Farup. Riemannian formulation and comparison of color difference formulas. *Color Research & Application*, 37(6):429–440, Dec 2012.

[29] A. Michelson. *Studies in Optics*. University of Chicago Press, 1927.

[30] X. Liu, M. Pedersen, and J.Y. Hardeberg. CID:IQ - a new image quality database. In A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, editors, *Image and Signal Processing*, volume 8509 of *Lecture Notes in Computer Science*, pages 193–202. Springer, Cherbourg, France, Jul. 2014.

[31] Colourlab image database: Image quality (CID:IQ). Available at `www.colourlab.no/cid`, 2014.

[32] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. Image Processing*, 15(2):430–444, 2006.

[33] Video Quality Experts Group. Final report from the video quality experts group: Validation of reduced-reference and no-reference objective models for standard definition television, phase I. Technical report, International Telecommunication Union, 2009.