STATISTICAL EVALUATION OF VISUAL QUALITY METRICS FOR IMAGE DENOISING

Karen Egiazarian[1], Mykola Ponomarenko[1],
Vladimir Lukin[2], and Oleg Ieremeiev[2]

([1]) Laboratory of Signal Processing, Tampere University of Technology, Finland

([2]) Department of Signal Transmitting, Receiving and Processing, National Aerospace University, Ukraine

Outline

1. Description of reference and distorted images of FLT database

2. Verification of modern metrics using MOS of FLT database
- *Image quality assessment based on FLT database*

3. Description of the proposed DSI metric
- *Introduction of new perceptual image quality metrícs*

Visual Quality Metrics for image denoising

Modern full-reference image visual quality metrics are not verified on their suitability for image denoising task, because

- the differences between denoised images are very small,
- distortions typical for the task (e.g. smoothed low-contrast textures, blurred noise-like textures) are not included in widely used image databases such as LIVE and TID2013.

Main goal of the paper

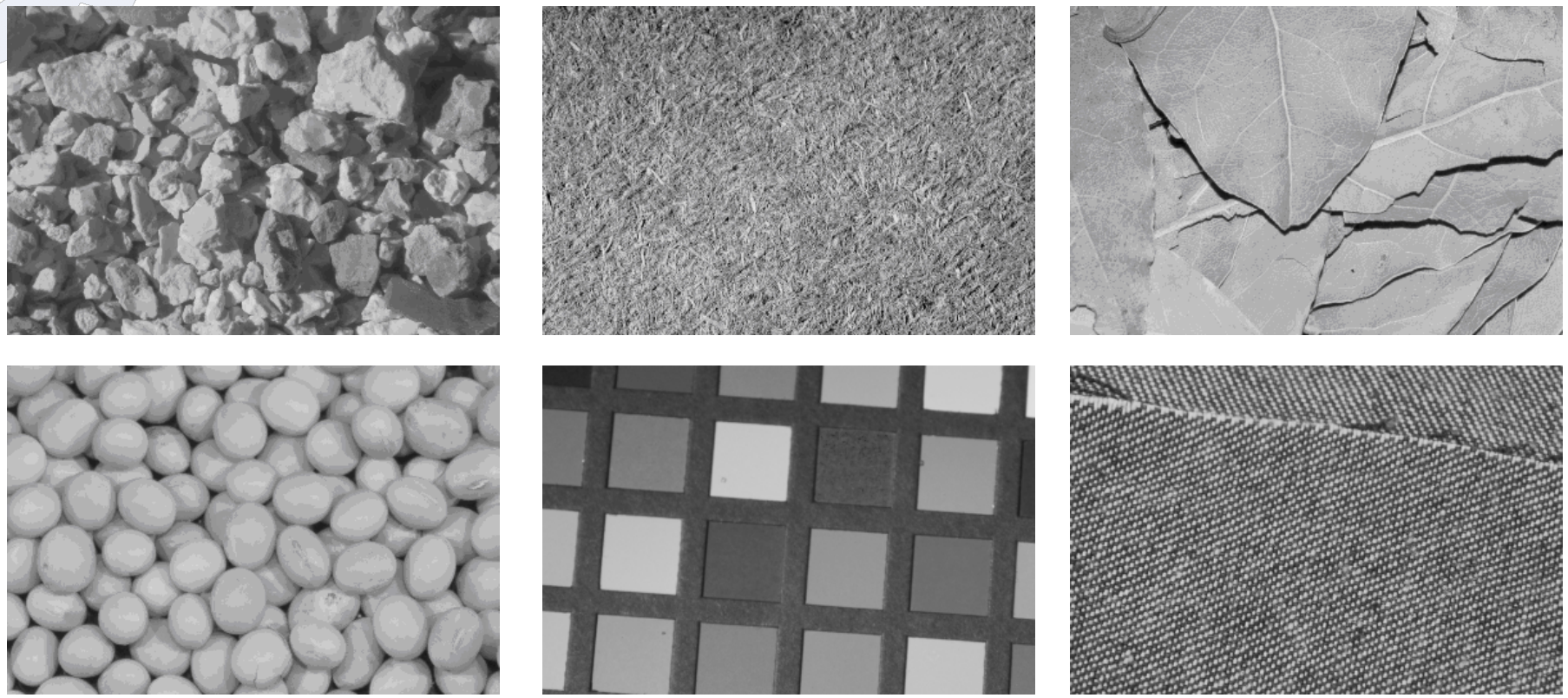
- Design of FLT databases consisting of filtered (denoised images)

  FLT (FiLTering) is a new image database of filtered images with various level of noise attenuation

This makes difficult to perform visual quality assessment for modern image quality metrics.

Reference images

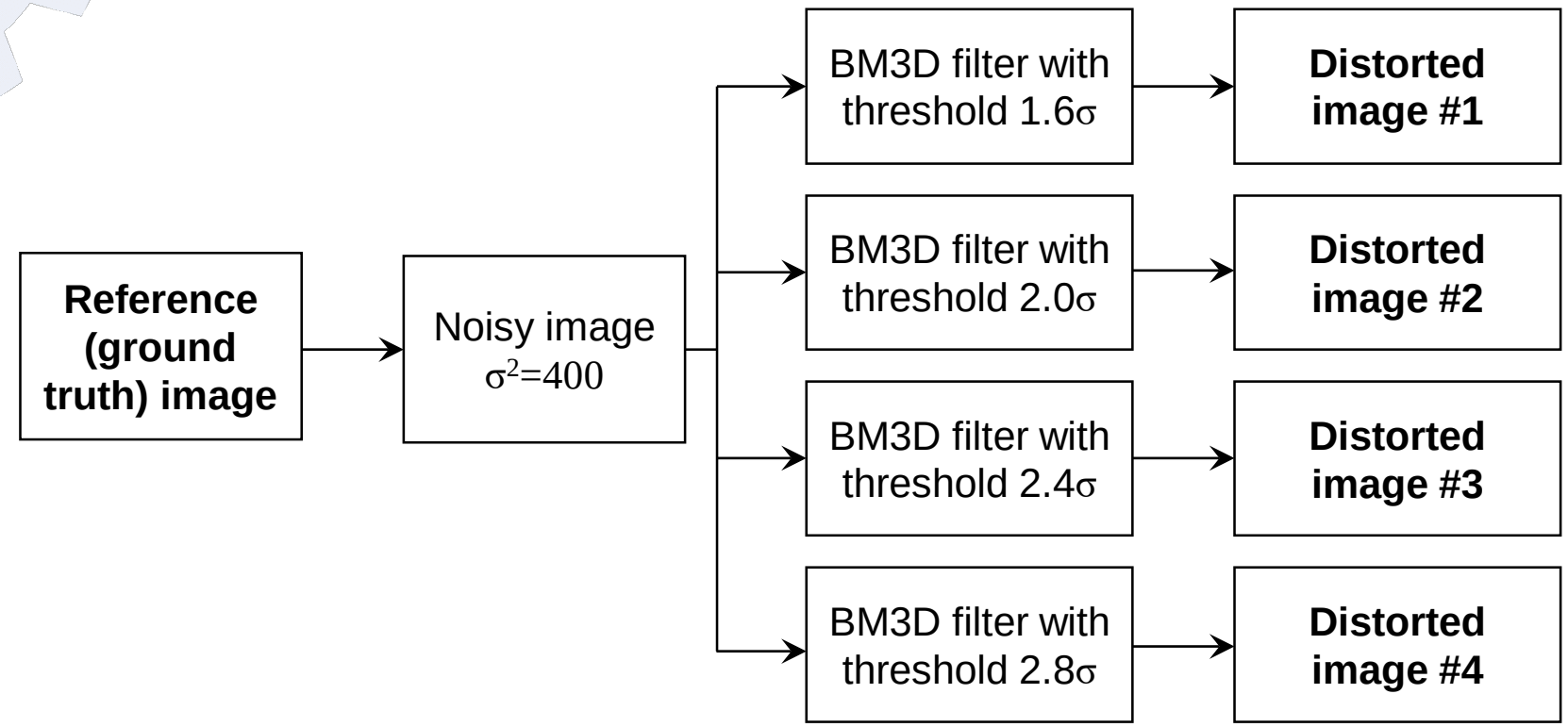
The reference images of FLT database are 75 grayscale texture images of resolution 384x256 pixels selected from Amsterdam Library of Textures:



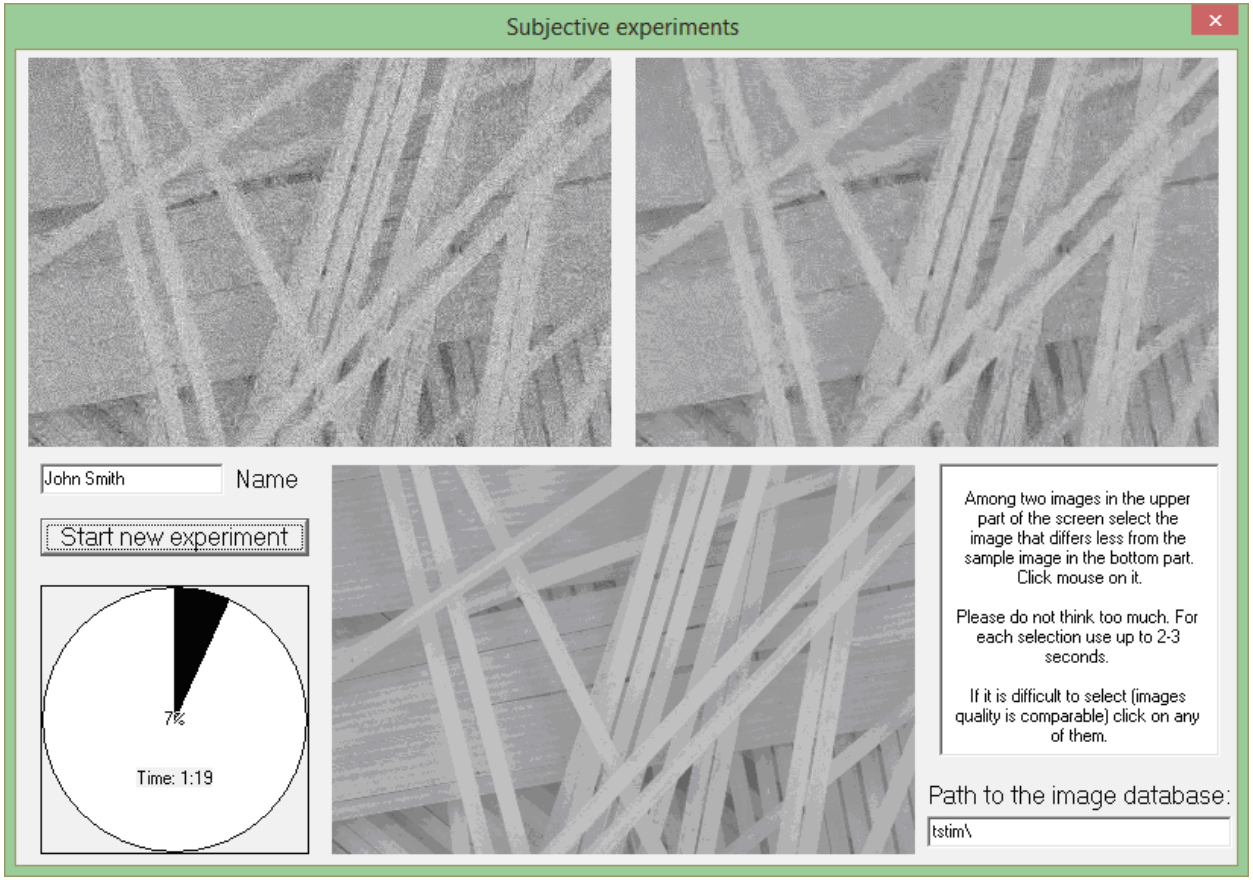
Examples of reference test images in FLT database

Distorted images

For each reference image 4 distorted images have been created in the following scheme:

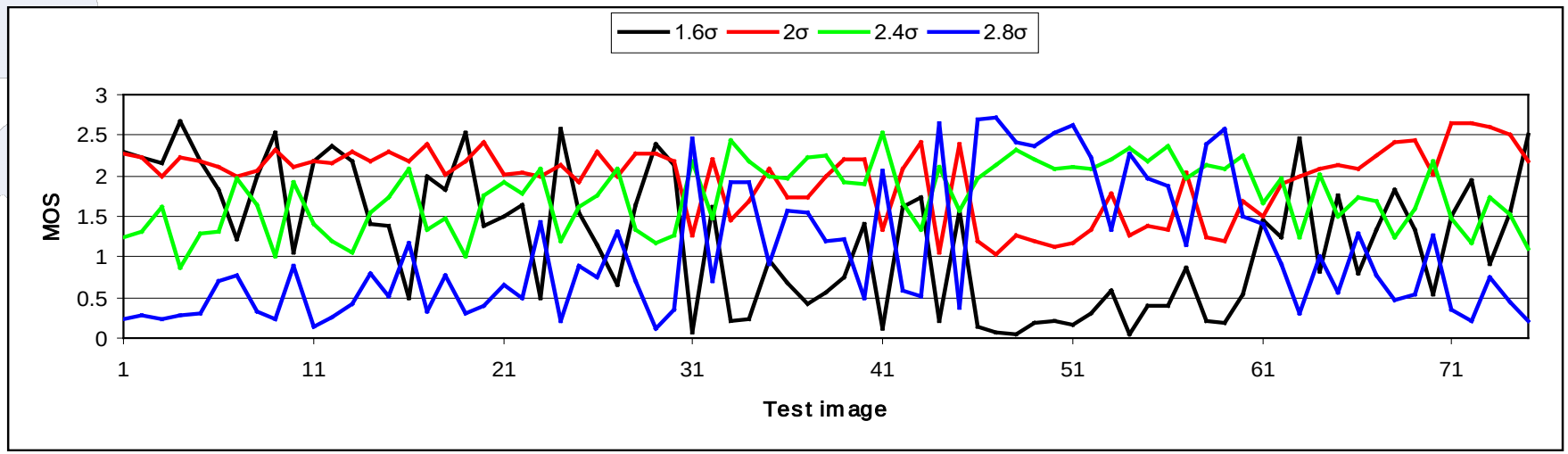


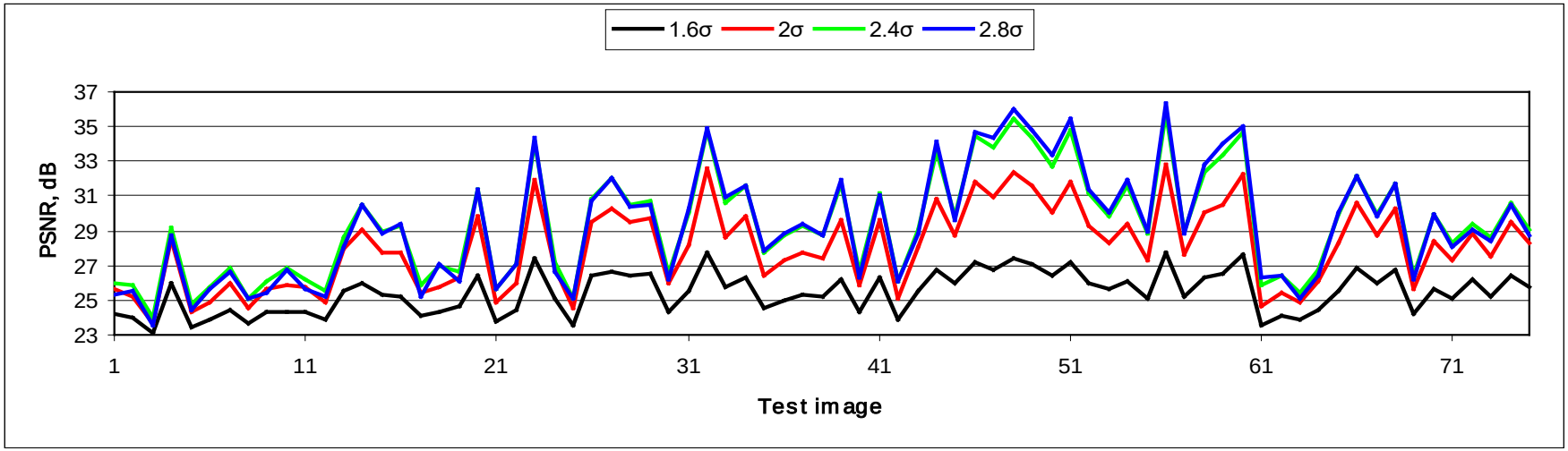
Obtaining of Mean Opinion Scores



Screenshot of software used in experiments for MOS obtaining (47 observers have been involved)

PSNR is totally useless for this task



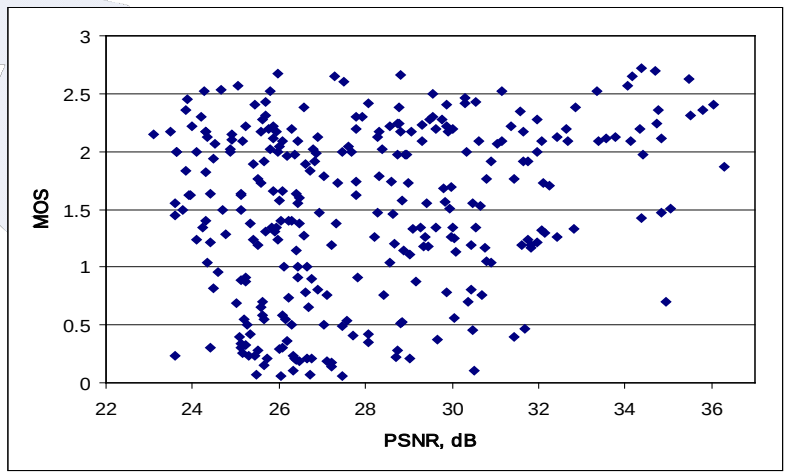


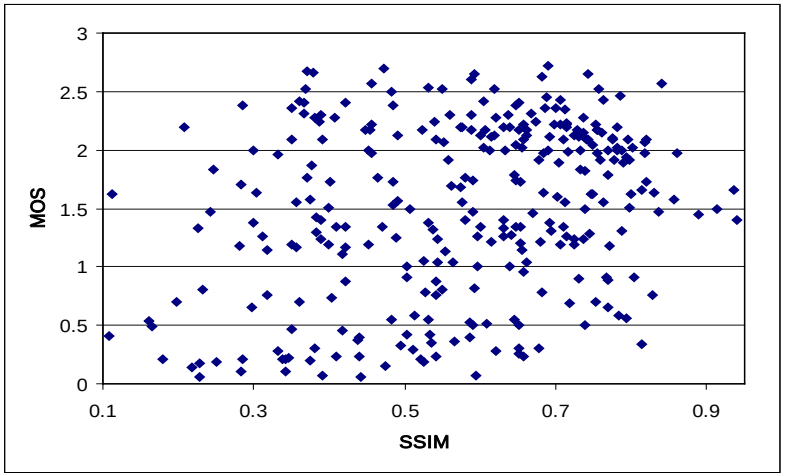
Using obtained MOS for verification of effectiveness of image visual quality metrics

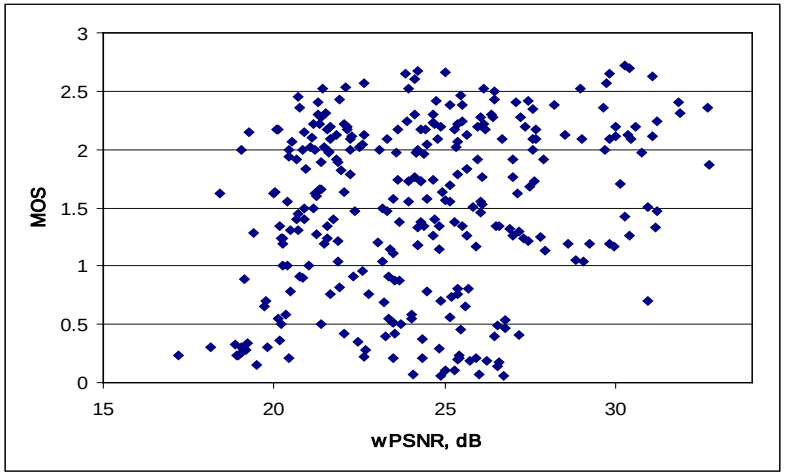
Spearman correlation between MOS and the metrics

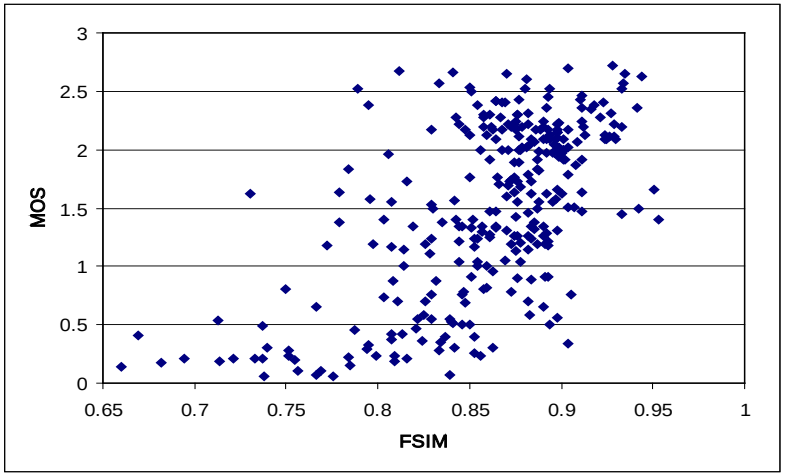
#	Metric	Subset 1 (16 images, noise-like textures)	Subset 2 (26 images , low-contrast textures)	Subset 3 (33 images , regular textures)	**Full set (75 images)**
1	**wPSNR**	0.83	0.74	0.76	**0.77**
2	**MSDDM**	0.70	0.64	0.85	**0.74**
3	**FSIM**	0.75	0.58	0.79	**0.71**
4	**GSM**	0.71	0.61	0.77	**0.70**
5	**SR-SIM**	0.66	0.53	0.82	**0.69**
6	**GMSD**	0.61	0.55	0.82	**0.68**
7	**SSIM**	0.58	0.32	0.73	**0.55**
8	**MSSIM**	0.58	0.21	0.72	**0.51**
9	**SFF**	0.21	0.22	0.69	**0.42**
10	**PSNR-HVSM**	-0.17	-0.12	0.72	**0.23**
11	**PSNR**	-0.55	-0.31	0.62	**0.05**

Scatter-plots for MOS and compared metrics









Description of proposed DSI metric

Let's start from the following metric:

MSDDM metric (Ponomarenko, Jin, Lukin and Egiazarian, 2011) is defined as:

$$MSDDM = -\frac{1}{QW}\sum_{i=1}^{Q}\sum_{j=1}^{W}\left(\sqrt{D_{ij}} - \sqrt{D_{ij}^{d}}\right)^{2}$$

where Q x W is an image size, D_{ij} and D^{d}_{ij} are dissimilarity maps of reference and distorted images.

Description of proposed DSI metric

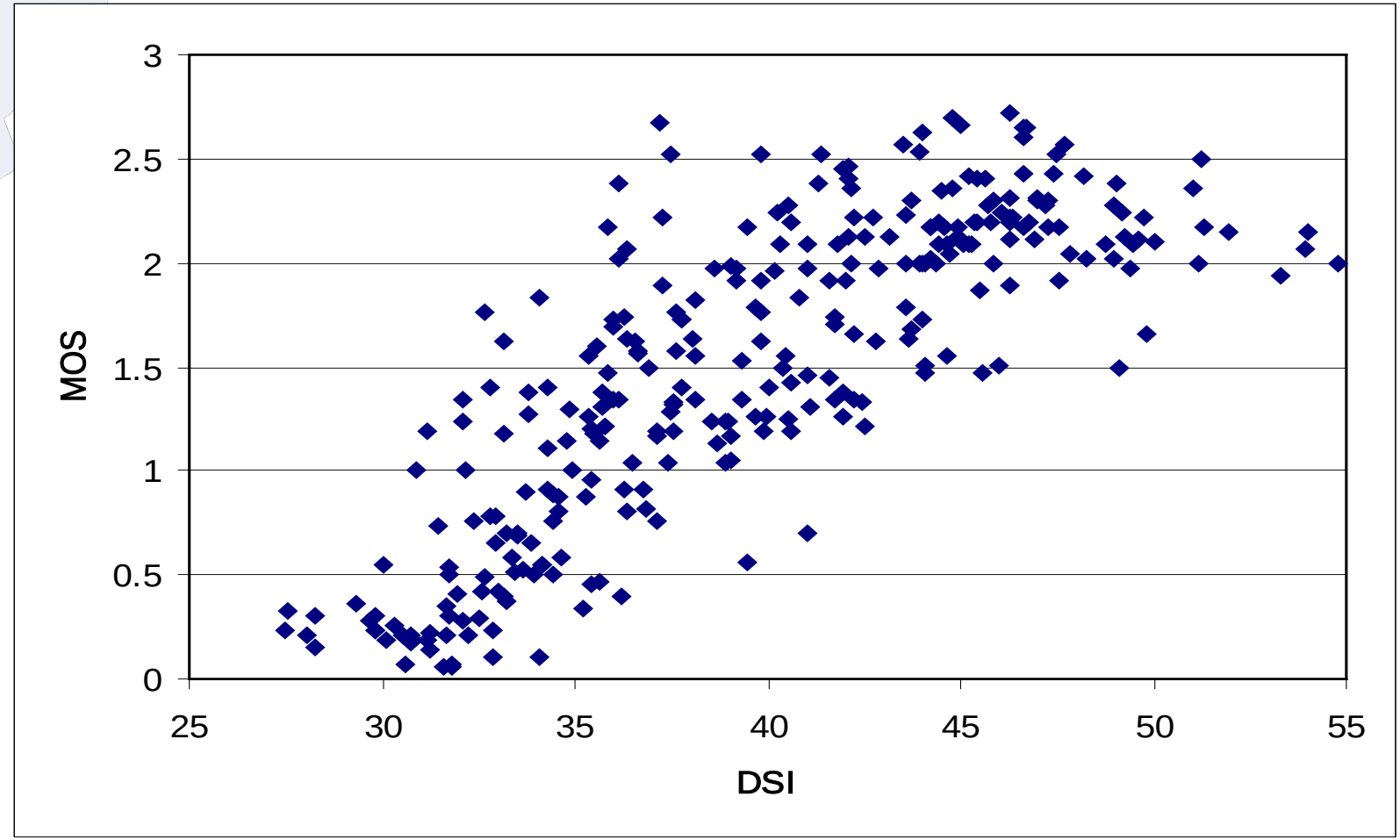
We modify this metric by decreasing each element of the sum on a value proportional to $\sqrt{D_{ij}^{d}}$ (masking effect of non-predictable energy of distorted image for the pixel (i,j)), resulting in the following DisSimilarity Index (DSI):

$$DSI = -\frac{1}{QW}\sum_{i=1}^{Q}\sum_{j=1}^{W}\max\left(0, \left|\sqrt{D_{ij}} - \sqrt{D_{ij}^{d}}\right| - \frac{\sqrt{D_{ij}^{d}}}{4.5}\right)^{2}$$

where DSI is the proposed modification, 4.5 is the correcting factor obtained as a result of the optimization process.

In this paper, for calculation of the proposed DSI metric, we have used blocks of size 5x5 pixels as well as an area for blocks similarity calculation of 19x19 pixels.

Scatter-plots for MOS and DSI metric



Comparison of suitability of visual quality metrics for the task of image denoising

Spearman correlation between MOS and the metrics

#	Metric	Subset 1 (16 images, noise-like textures)	Subset 2 (26 images , low-contrast textures)	Subset 3 (33 images , regular textures)	**Full set (75 images)**
1	**DSI**	0.85	0.82	0.80	**0.82**
2	**wPSNR**	0.83	0.74	0.76	**0.77**
3	**MSDDM**	0.70	0.64	0.85	**0.74**
4	**FSIM**	0.75	0.58	0.79	**0.71**
5	**GSM**	0.71	0.61	0.77	**0.70**
6	**SR-SIM**	0.66	0.53	0.82	**0.69**
7	**GMSD**	0.61	0.55	0.82	**0.68**
8	**SSIM**	0.58	0.32	0.73	**0.55**
9	**MSSIM**	0.58	0.21	0.72	**0.51**
10	**SFF**	0.21	0.22	0.69	**0.42**
11	**PSNR-HVSM**	-0.17	-0.12	0.72	**0.23**
12	**PSNR**	-0.55	-0.31	0.62	**0.05**

Summary

- A new test image database FLT, consisting of BM3D-filtered images with various parameters, is proposed.
- Relatively small visual difference between filtered images having close filter parameters makes visual quality assessment difficult for existing subjective quality metrics, resulting in SROCC less than 0.9 between them and human perception (MOS).

Summary

- A new image visual quality metric DSI (modification of the MSDDM metric) is proposed. It is shown that DSI provides better correspondence to HVS for FLT database than any other considered metrics. DSI takes into account a masking effect of non-predictable energy of image regions, proving correctness of the hypothesis formulated in (Ieremeiev, Ponomarenko, Lukin, Astola and Egiazarian, 2017).

- The FLT database together with the Matlab realization of DSI metric are freely available at http://ponomarenko.info/flt.htm.