INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/AG 05 MPEG VISUAL QUALITY ASSESSMENT

ISO/IEC JTC 1/SC 29/WG 04 M 68223
July 2024, Sapporo, JP

Status	Input document
Title	IV-SSIM: adapting structural similarity to immersive video
Source	Poznan University of Technology, Poznań, Poland
Authors	Jakub Stankowski, Weronika Nowak, Adrian Dziembowski

1. Abstract

This document describes a new objective quality metric designed for immersive video applications – IV-SSIM. IV-SSIM is an evolution of IV-PSNR. IV-SSIM combines the advantages of IV-PSNR and metrics based on the structural similarity of images, being able to properly mimic the subjective quality perception of immersive video with its characteristic distortions induced by the reprojection of pixels between multiple views.

Effectiveness of IV-SSIM was compared in two experiments, using results of the MIV CfP [WG11 N18353] and a commonly-used (non-immersive) image quality database – TID2013. It was compared to 15 state-of-the-art full-reference objective quality metrics.

An efficient implementation of the IV-SSIM metric is included in the QMIV software proposed for WG04 [M68224].

2. SSIM vs. immersive video

SSIM is an efficient objective quality metric for general purposes. For immersive video, however, where many artifacts are induced by reprojection of pixels between different views, its performance is worse. Similarly to PSNR and other pixel-based metrics, it is sensitive to even slightest shifts of objects:

No shift (compared to input view)	2-pixel shift	Significant shift
		
SSIM_Y: 0.9954 IV-SSIM: 0.9973	SSIM_Y: 0.9893 IV-SSIM: 0.9973	SSIM_Y: 0.9805 IV-SSIM: 0.9862

Development of a negligible-shift-independent quality metric is possible due to changing the pixel-based behavior of the metric. Instead of comparing a pixel of image I to the colocated pixel of image J, the pixel of image I is being compared to the most similar pixel within a colocated neighborhood of image J:

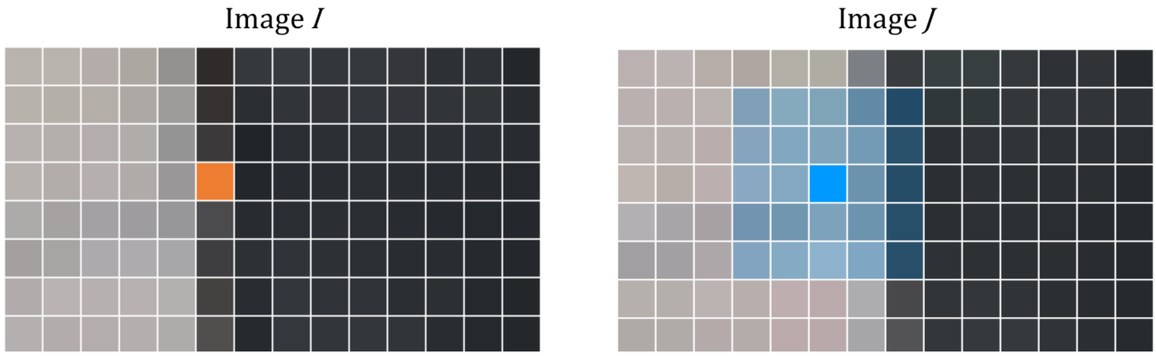


Figure. Pixel-based analysis: orange pixel in image I is compared to the colocated opaque blue pixel in image J; Proposed analysis: orange pixel in image I is compared to all blue pixels in image J (5×5 neighborhood of the colocated pixel), the difference is calculated between the value of the orange pixel and the most similar pixel within the blue block.

Moreover, SSIM cannot properly assess the quality of a view with globally changed color (a typical case for views rendered using input cameras with different color characteristics):

Reference input view	View rendered using inputs with different color characteristics	View rendered using color-corrected inputs
		
	SSIMY: 0.9216 IV-SSIM: 0.9900	SSIMY: 0.9107 IV-SSIM: 0.9905

Similarly to IV-PSNR, IV-SSIM includes calculation of the global color offset between two compared images:

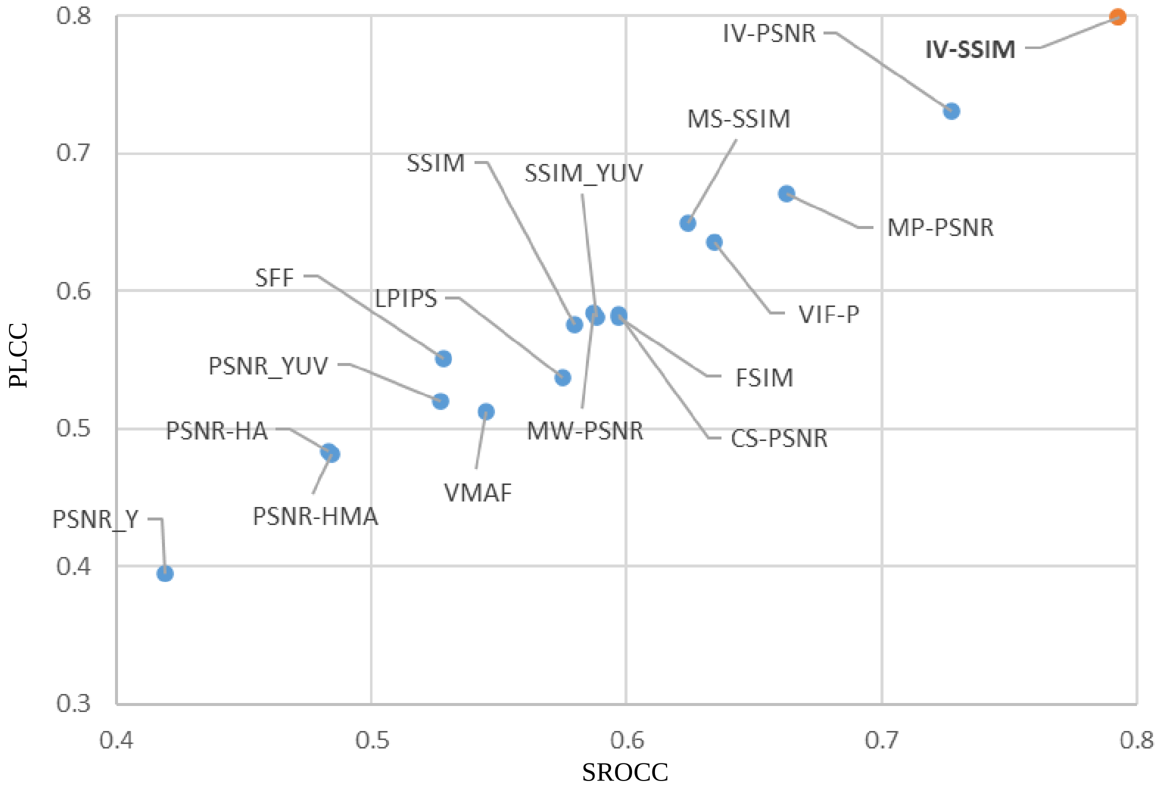
$$s_c^{I \to J} = \frac{1}{W_c \cdot H_c} \sum_{y=0}^{H_c-1} \sum_{x=0}^{W_c-1} \left(\boldsymbol{I}_c(x,y) - \boldsymbol{J}_c(x,y) \right).$$

This shift is then added to each pixel of image J.

3. IV-SSIM vs. SSIM

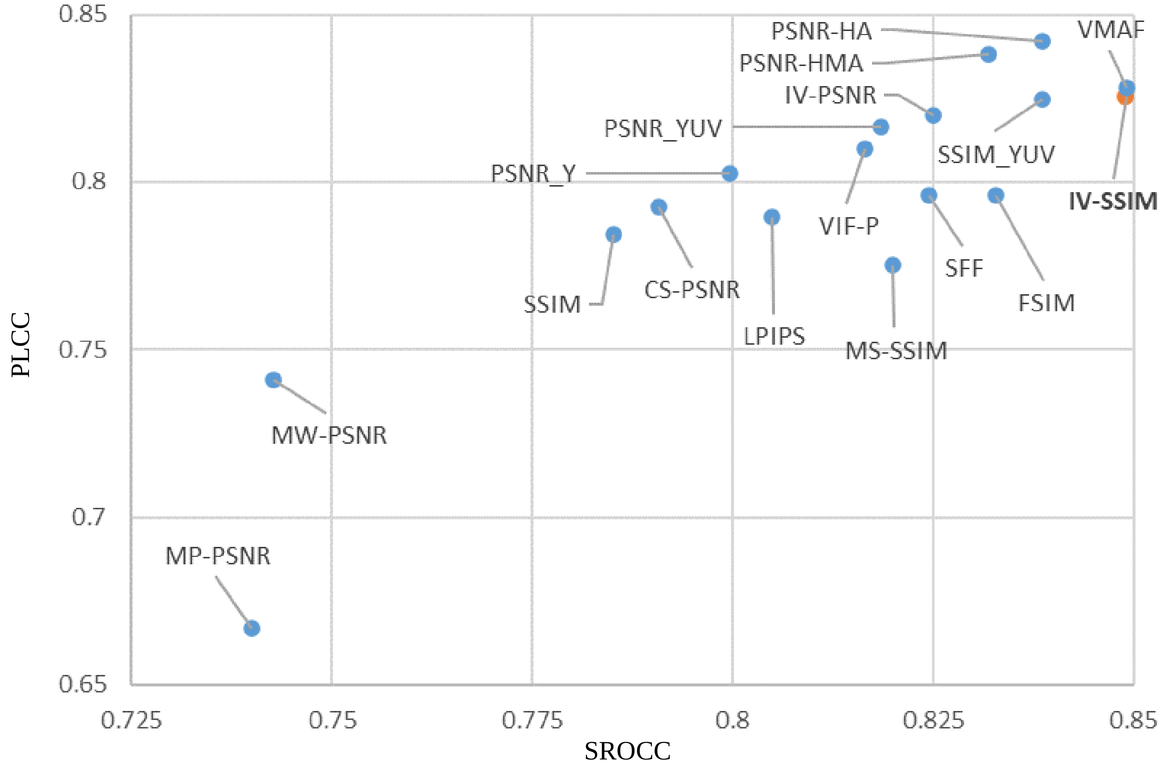
SSIM	IV-SSIM
Local image statistics	
$\boldsymbol{\mu}_c^I(x,y) = \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} [\boldsymbol{\omega}(i-x,j-y) \cdot \boldsymbol{I}_c(i,j)]$	$\boldsymbol{\mu}_c^I(x,y) = \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} [\boldsymbol{\omega}(i-x,j-y) \cdot \boldsymbol{I}_c(i,j)]$
$\boldsymbol{\sigma}_c^I(x,y) = \sqrt{\sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} \left[\boldsymbol{\omega}(i-x,j-y) \cdot \left(\boldsymbol{I}_c(i,j)\right)^2\right] - \left(\boldsymbol{\mu}_c^I(x,y)\right)^2}$	$\boldsymbol{\sigma}_c^I(x,y) = \sqrt{\sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} \left[\boldsymbol{\omega}(i-x,j-y) \cdot \left(\boldsymbol{I}_c(i,j)\right)^2\right] - \left(\boldsymbol{\mu}_c^I(x,y)\right)^2}$
$\boldsymbol{\mu}_c^J(x,y) = \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} [\boldsymbol{\omega}(i-x,j-y) \cdot \boldsymbol{J}_c(i,j)]$	$\boldsymbol{\mu}_c^J(x,y) = \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} [\boldsymbol{\omega}(i-x,j-y) \cdot \boldsymbol{J}_c(i',j')]$
$\boldsymbol{\sigma}_c^J(x,y) = \sqrt{\sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} \left[\boldsymbol{\omega}(i-x,j-y) \cdot \left(\boldsymbol{J}_c(i,j)\right)^2\right] - \left(\boldsymbol{\mu}_c^J(x,y)\right)^2}$	$\boldsymbol{\sigma}_c^J(x,y) = \sqrt{\sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} \left[\boldsymbol{\omega}(i-x,j-y) \cdot \left(\boldsymbol{J}_c(i',j')\right)^2\right] - \left(\boldsymbol{\mu}_c^J(x,y)\right)^2}$
$\boldsymbol{\sigma}_c^{I,J}(x,y) = \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} [\boldsymbol{\omega}(i-x,j-y) \cdot \boldsymbol{I}_c(i,j) \cdot \boldsymbol{J}_c(i,j)] - \boldsymbol{\mu}_c^I(x,y) \cdot \boldsymbol{\mu}_c^J(x,y)$	$\boldsymbol{\sigma}_c^{I\to J}(x,y) = \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} [\boldsymbol{\omega}(i-x,j-y) \cdot \boldsymbol{I}_c(i,j) \cdot \boldsymbol{J}_c(i',j')] - \boldsymbol{\mu}_c^I(x,y) \cdot \boldsymbol{\mu}_c^J(x,y)$
Calculation of i', j'	
	$i' = i + s_x(i,j)\,, \quad j' = j + s_y(i,j)$
	$s_x(x,y), s_y(x,y) \in [-B,B] \cap \mathbb{Z} \quad \ni$ $\left\lvert \boldsymbol{I}_C(x,y) - \boldsymbol{J}_C\left(x + s_x(x,y), y + s_y(x,y)\right)\right\rvert = \min_{w\in[-B,B],\, h\in[-B,B]} \lvert \boldsymbol{I}_C(x,y) - \boldsymbol{J}_C(x+w, y+h)\rvert$
Three image properties	
$\boldsymbol{L}_c^{I,J}(x,y) = \frac{2 \cdot \mu_c^I(x,y) \cdot \mu_c^J(x,y) + C_1}{\mu_c^I(x,y)^2 + \mu_c^J(x,y)^2 + C_1}$	$\boldsymbol{L}_c^{I\to J}(x,y) = \frac{2 \cdot \mu_c^I(x,y) \cdot \left[\mu_c^J(x,y) + s_c^{I\to J}\right] + C_1}{\mu_c^I(x,y)^2 + \left[\mu_c^J(x,y) + s_c^{I\to J}\right]^2 + C_1}$
$\boldsymbol{C}_c^{I,J}(x,y) = \frac{2 \cdot \sigma_c^I(x,y) \cdot \sigma_c^J(x,y) + C_2}{\sigma_c^I(x,y)^2 + \sigma_c^J(x,y)^2 + C_2}$	$\boldsymbol{C}_c^{I\to J}(x,y) = \frac{2 \cdot \sigma_c^I(x,y) \cdot \sigma_c^J(x,y) + C_2}{\sigma_c^I(x,y)^2 + \sigma_c^J(x,y)^2 + C_2}$
$\boldsymbol{S}_c^{I,J}(x,y) = \frac{\sigma_c^{I,J}(x,y) + C_3}{\sigma_c^I(x,y) \cdot \sigma_c^J(x,y) + C_3}$	$\boldsymbol{S}_c^{I\to J}(x,y) = \frac{\sigma_c^{I\to J}(x,y) + C_3}{\sigma_c^I(x,y) \cdot \sigma_c^J(x,y) + C_3}$
Calculation of the global offset between images I and J	
	$s_c^{I\to J} = \frac{1}{W_c \cdot H_c} \sum_{y=0}^{H_c-1} \sum_{x=0}^{W_c-1} (\boldsymbol{I}_c(x,y) - \boldsymbol{J}_c(x,y))$
Local quality scores	
$\boldsymbol{Q}_c^{I,J}(x,y) = \left[\boldsymbol{L}_c^{I,J}(x,y)\right]^\alpha \cdot \left[\boldsymbol{C}_c^{I,J}(x,y)\right]^\beta \cdot \left[\boldsymbol{S}_c^{I,J}(x,y)\right]^\gamma$	$\boldsymbol{Q}_c^{I\to J}(x,y) = \left[\boldsymbol{L}_c^{I\to J}(x,y)\right]^\alpha \cdot \left[\boldsymbol{C}_c^{I\to J}(x,y)\right]^\beta \cdot \left[\boldsymbol{S}_c^{I\to J}(x,y)\right]^\gamma$
Global quality score	
$IV\text{-}SSIM_c^{I,J} = \frac{1}{W_c \cdot H_c} \sum_{y=0}^{H_c-1} \sum_{x=0}^{W_c-1} \boldsymbol{Q}_c^{I,J}(x,y)$	$IV\text{-}SSIM_c^{I\to J} = \frac{1}{W_c \cdot H_c} \sum_{y=0}^{H_c-1} \sum_{x=0}^{W_c-1} \boldsymbol{Q}_c^{I\to J}(x,y)$
$IV\text{-}SSIM_{YUV}^{I,J} = \frac{IV\text{-}SSIM_Y^{I,J} \cdot w_Y + IV\text{-}SSIM_U^{I,J} \cdot w_U + IV\text{-}SSIM_V^{I,J} \cdot w_v}{w_Y + w_U + w_V}$	$IV\text{-}SSIM_{YUV}^{I\to J} = \frac{IV\text{-}SSIM_Y^{I\to J} \cdot w_Y + IV\text{-}SSIM_U^{I\to J} \cdot w_U + IV\text{-}SSIM_V^{I\to J} \cdot w_v}{w_Y + w_U + w_V}$
	$IV\text{-}SSIM_{YUV}^{I,J} = \min\left(IV\text{-}SSIM_{YUV}^{I\to J}, IV\text{-}SSIM_{YUV}^{J\to I}\right)$
$IV\text{-}SSIM_{video}^{I,J} = \frac{1}{F} \sum_{f=0}^{F-1} IV\text{-}SSIM_{YUV}^{I,J}(f)$	$IV\text{-}SSIM_{video}^{I,J} = \frac{1}{F} \sum_{f=0}^{F-1} IV\text{-}SSIM_{YUV}^{I,J}(f)$
Some constants	
$\boldsymbol{\omega}$ – 11x11 Gaussian mask	$\boldsymbol{\omega}$ – 11x11 Gaussian mask
$C_1 = (K_1 \cdot (2^b - 1))^2, \quad C_2 = (K_2 \cdot (2^b - 1))^2, \quad C_3 = \frac{C_2}{2}$	$C_1 = (K_1 \cdot (2^b - 1))^2, \quad C_2 = (K_2 \cdot (2^b - 1))^2, \quad C_3 = \frac{C_2}{2}$
$K_1 = 0.01, \quad K_2 = 0.03$	$K_1 = 0.01, \quad K_2 = 0.03$
	$B = 2$
$\alpha = \beta = \gamma = 1$	$\alpha = \beta = \gamma = 1$
$w_Y = 4, \quad w_U = 1, \quad w_V = 1$	$w_Y = 4, \quad w_U = 1, \quad w_V = 1$

4. Effectiveness in immersive video coding (MIV CfP)

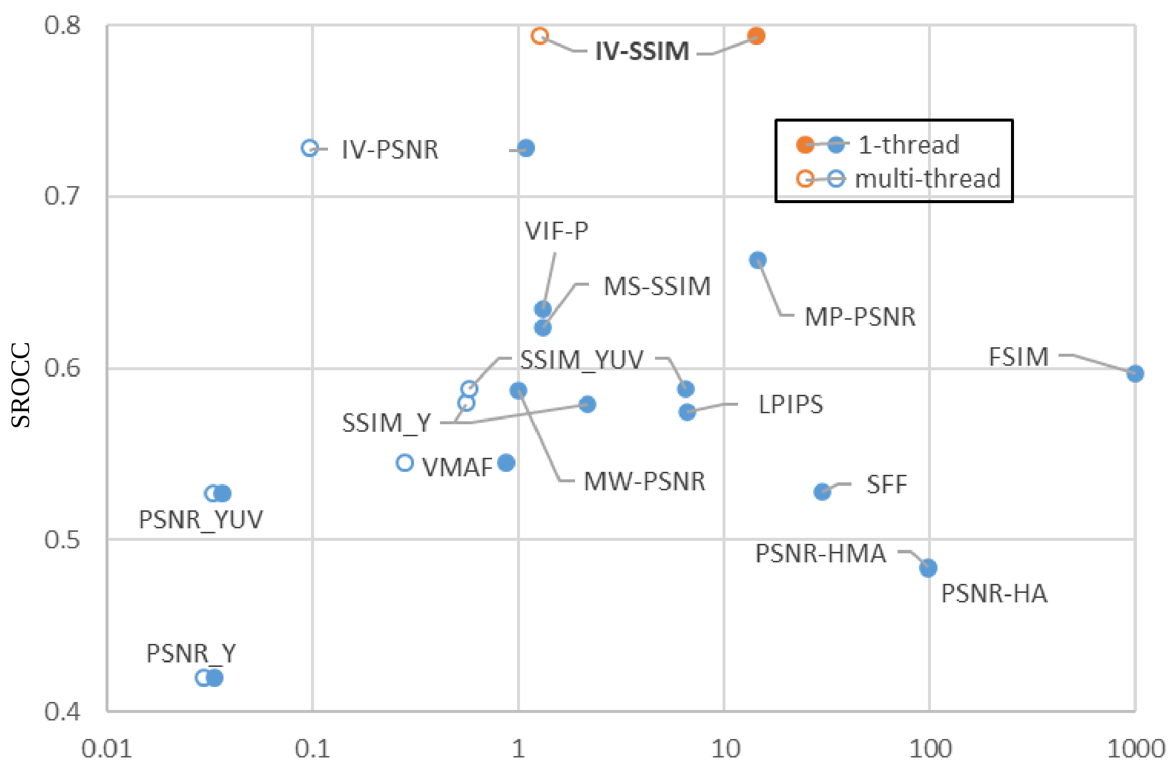


Correlation metrics:
PLCC – Pearson linear correlation coefficient
SROCC – Spearman's rank correlation coefficient

5. Effectiveness in general applications (TID2013)



6. Computational time



Computational time for single 4K×4K frame [s]

Calculations were performed on AMD Ryzen 9 5900X (12 cores).
Times for PSNR, SSIM, IV-PSNR, and IV-SSIM were obtained using the QMIV framework.
Other metrics were evaluated using publicly available implementations.

7. Recommendations

We encourage the Group to test the IV-SSIM metric for evaluating its effectiveness in various applications.

8. Acknowledgement

This work was supported by Ministry of Science and Higher Education of Republic of Poland.