

A HUMAN VISUAL SYSTEM- BASED 3D VIDEO QUALITY METRIC

Amin Banitalebi¹, Mahsa Pourazad^{1,2}, Panos Nasiopoulos¹

¹University of British Columbia, Vancouver, Canada

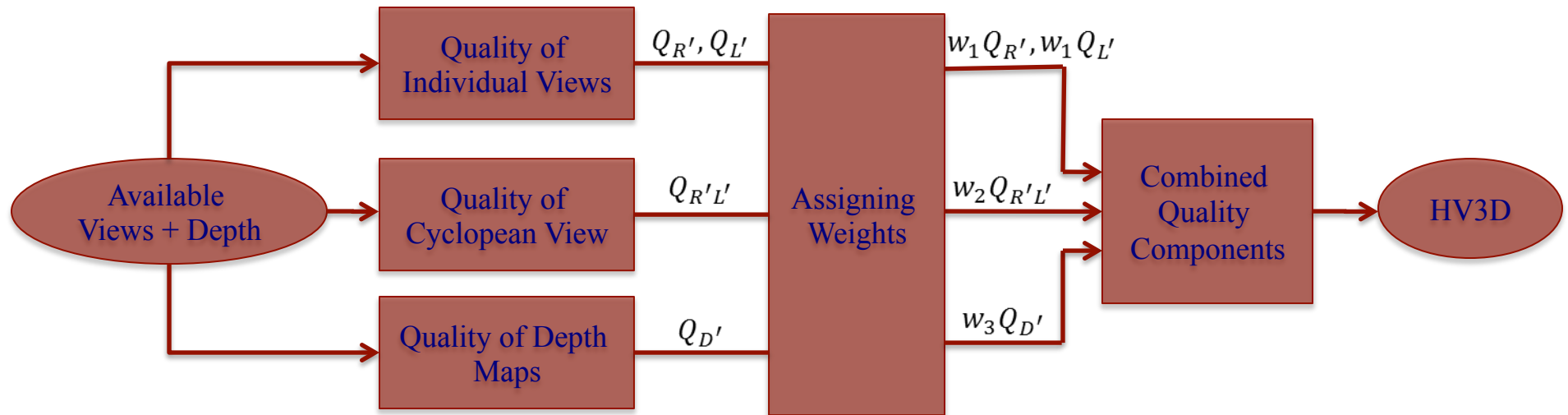
²TELUS Communications Inc., Canada



a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA



Human Visual System-based Quality Metric for 3D (HV3D):



$$HV3D = w_1 Q_{R'} + w_1 Q_{L'} + w_2 Q_{R'L'} + w_3 Q_{D'}$$

Quality of individual views:

■ HV3D: Formulation

$$HV3D = \underbrace{w_1 Q_{R'}}_{\text{Right View}} + \underbrace{w_1 Q_{L'}}_{\text{Left View}} + w_2 Q_{R'L'} + w_3 Q_{D'}$$

■ Quality of individual views

$$w_1 Q_{R'} = w_1 VIF(Y_R, Y_{R'}) + w_4 VIF(U_R, U_{R'}) + w_4 VIF(V_R, V_{R'})$$

$$w_1 Q_{L'} = w_1 VIF(Y_L, Y_{L'}) + w_4 VIF(U_L, U_{L'}) + w_4 VIF(V_L, V_{L'})$$

VIF: Visual information fidelity

Quality of Cyclopean View:

■ HV3D: Formulation

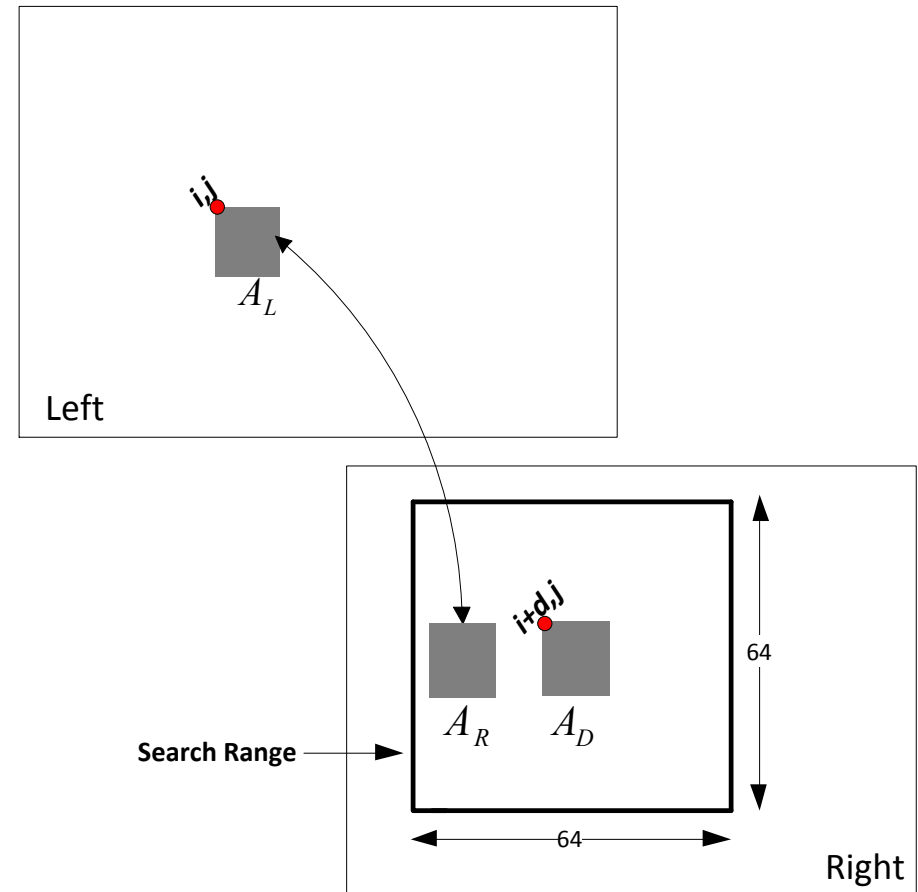
$$HV3D = w_1 Q_{R'} + w_1 Q_{L'} + w_2 Q_{R'L'} + w_3 Q_{D'}$$

■ Quality of cyclopean view

$$Q_{R'L'} = VIF(D, D')^\beta \sum_{i=1}^N \frac{SSIM(IDCT(XC_i), IDCT(XC'_i))}{N}$$

Modeling the Cyclopean View:

- HVS fuses the left and right views of a scene into a single cyclopean view
- Modeling the cyclopean view:
 - **Step1: Finding matching blocks**



Modeling the Cyclopean View:

■ Step 2: Fuse the information of right and left views:

- apply the 3D-DCT transform
- keep the the low frequencies (a 16×16 DCT-block)
(human visual system is more sensitive to low freq.)

■ Step 3: Include the effect of sensitivity to contrast:

- apply Contrast Sensitivity Function (CSF) modeling mask to 3D-DCT coefficients:
 - Coefficients of CSF are inversely proportional to the JPEG Q-table

$$XC = \sum_{i=1}^{16} \sum_{j=1}^{16} C_{i,j} X_{i,j}$$

Cyclopean view

CSF

low freq. coef.
of 3D-DCT

Quality of Cyclopean View:

$$Q_{R'L'} = VIF(D, D')^\beta \sum_{i=1}^N \frac{SSIM(IDCT(XC_i), IDCT(XC'_i))}{N}$$

- SSIM: measure structural similarities between two cyclopean views (not geometric distortions)
- VIF: measure geometric distortions between right and left images (within the cyclopean view) from the depth map

Quality of Depth Map:

■ HV3D: Formulation

$$HV3D = w_1 Q_{R'} + w_1 Q_{L'} + w_2 Q_{R'L'} + w_3 Q_{D'}$$

- The quality of the depth map becomes more important if there are several different depth levels in the scene.

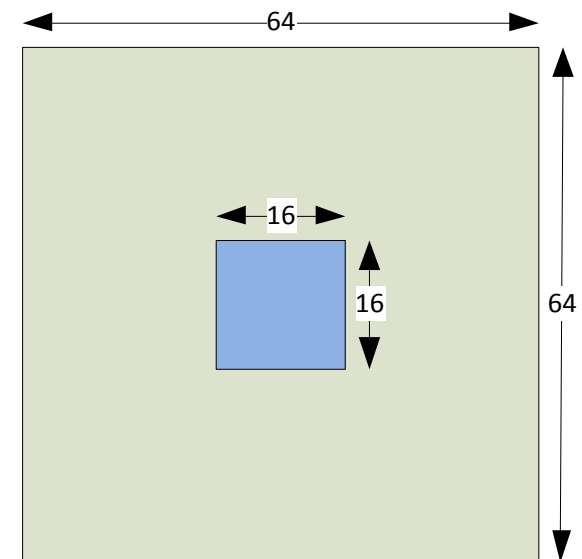
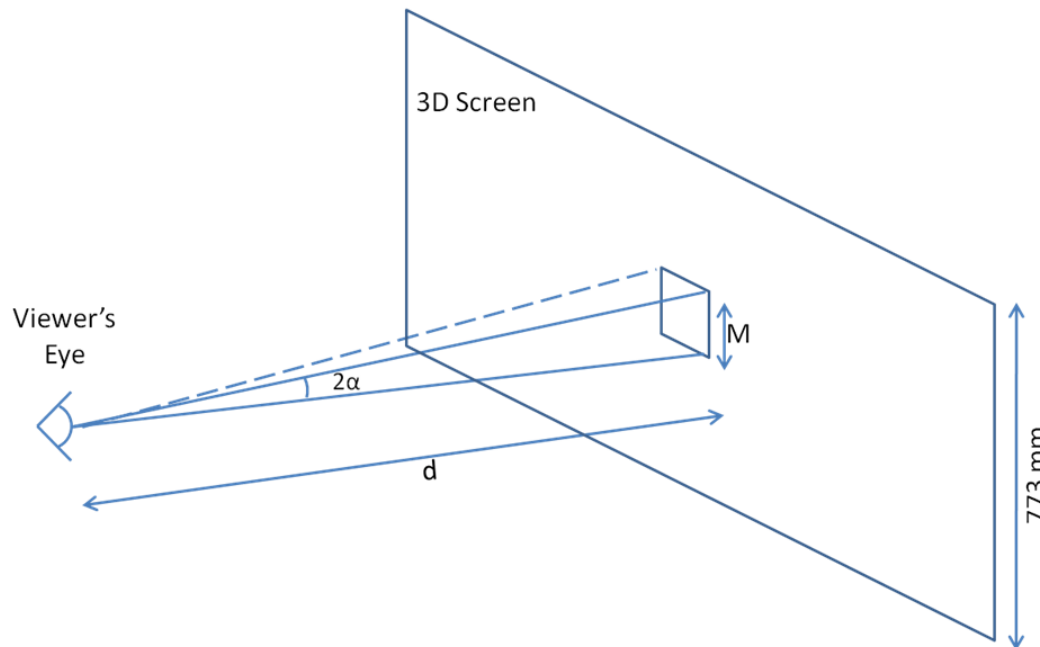
$$Q_{D'} = VIF(D, D')^\beta \sum_{i=1}^N \frac{\sigma_{d_i}^2}{N \cdot \max(\sigma_{d_j}^2 \mid j = 1, 2, \dots, N)}$$

$$\sigma_{d_i}^2 = \frac{1}{64 \times 64 - 1} \sum_{k,l=1}^{64} (M_{d_i} - R_{k,l})^2$$

Quality of Depth Map:

- local disparity variance is calculated over a block size area that can be fully projected onto the eye fovea

$$\sigma_{d_i}^2 = \frac{1}{64 \times 64 - 1} \sum_{k,l=1}^{64} (M_{d_i} - R_{k,l})^2$$



Determining Weighting Constants:

■ Subjective tests: training

$$HV3D = w_1 Q_{R'} + w_1 Q_{L'} + w_2 Q_{R'L'} + w_3 Q_{D'} \leftarrow \text{Combined quality components}$$

$$HV3D = w_1 q_1 + w_2 q_2 + w_3 q_3 + w_4 q_4 \leftarrow \text{Factorizing for weighting constants}$$

$$\min_{w_i, i=1,2,3,4} \{ \|HV3D - MOS\|^2 \}$$

Optimization problem

$$HV3D = QW$$

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & q_{1,3} & q_{1,4} \\ q_{2,1} & q_{2,2} & q_{2,3} & q_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ q_{20,1} & q_{20,2} & q_{20,3} & q_{20,4} \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$$

$$W = (Q^T Q)^{-1} Q^T MOS \leftarrow \text{Least Mean Square solution}$$

w ₁	w ₂	w ₃	w ₄
0.14	0.1208	0.05	0.1353

HV3D metric:

■ Maximum of HV3D:

$$HV3D_{\max} = 2w_1 + 4w_4 + w_2 + w_3 \cdot \sum_{i=1}^N \frac{\sigma_{d_i}^2}{N \cdot \max(\sigma_{d_j}^2 \mid j = 1, 2, \dots, N)}$$

■ Normalizing HV3D:

$$\begin{aligned} HV3D = & [w_1 VIF(Y_R, Y_{R'}) + w_4 VIF(U_R, U_{R'}) + w_4 VIF(V_R, V_{R'}) \\ & + w_1 VIF(Y_L, Y_{L'}) + w_4 VIF(U_L, U_{L'}) + w_4 VIF(V_L, V_{L'}) \\ & + w_2 VIF(D, D')^\beta \cdot \sum_{i=1}^N \frac{SSIM(IDCT(XC_i), IDCT(XC'_i))}{N} \\ & + w_3 VIF(D, D')^\beta \cdot \sum_{i=1}^N \frac{\sigma_{d_i}^2}{N \cdot \max(\sigma_{d_i}^2 \mid i = 1, 2, \dots, N)}] / HV3D_{\max} \end{aligned}$$

2-View Case Scenario: Performance Evaluation

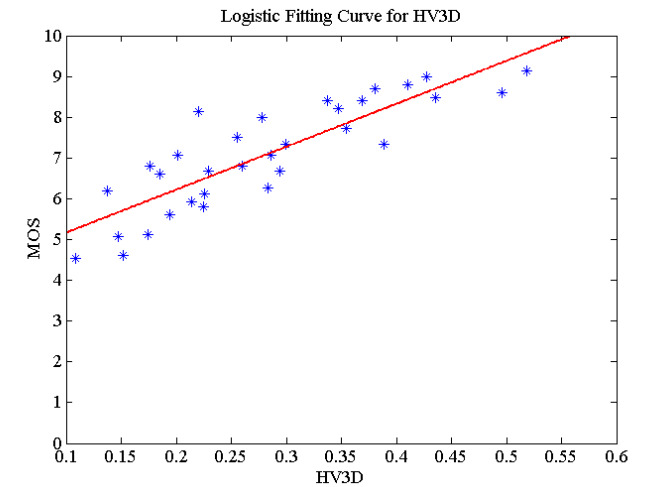
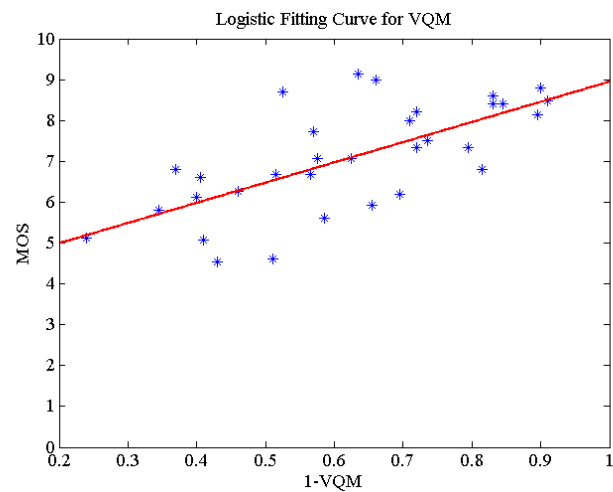
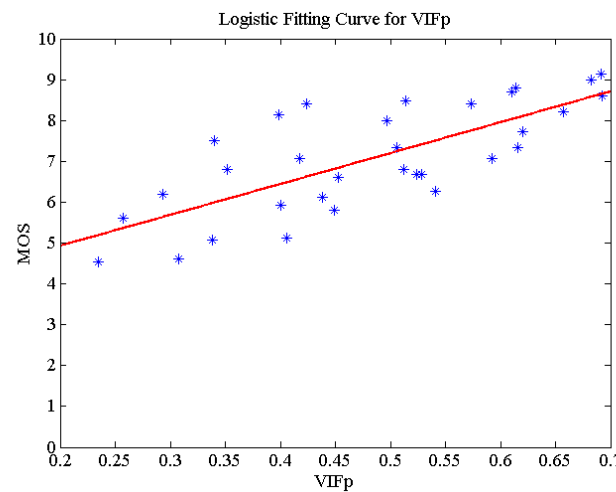
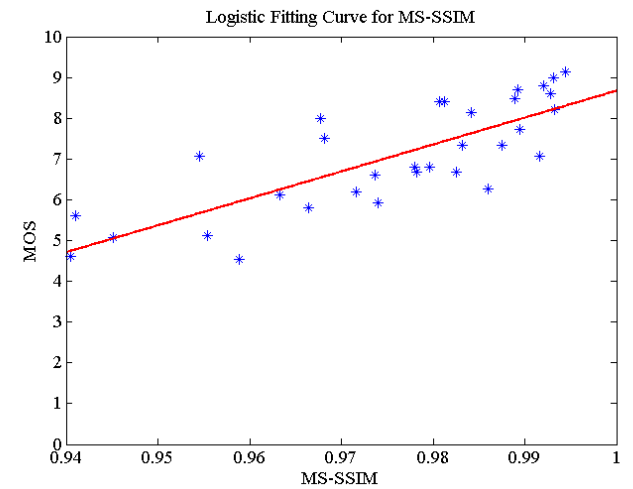
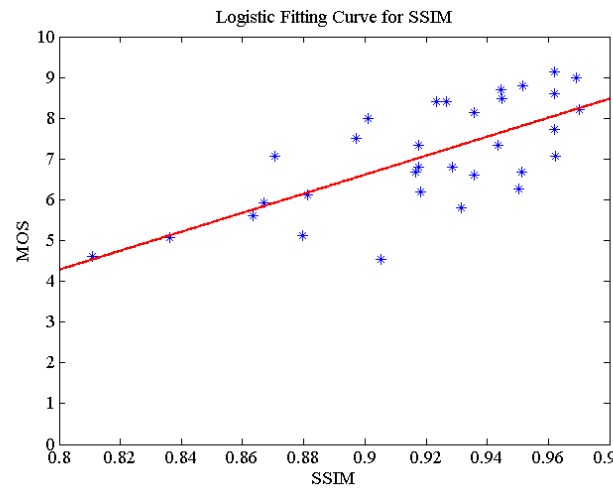
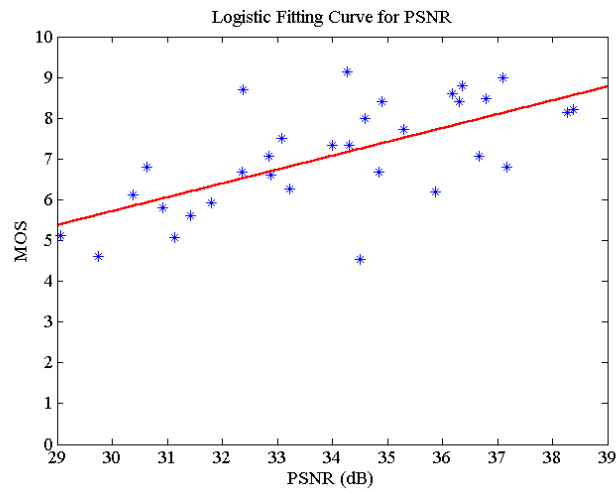
- Video sequences and subjective tests according to:

Table 1: Input views and stereo pair for 2-view test scenario

Seq. ID	Test Sequence	Test Class	Input views	View to synthesize	Stereo pair
S01	Poznan_Hall2	A	7-6	6.5	6.5-6
S02	Poznan_Street		4-3	3.5	3.5-3
S03	Undo_Dancer		2-5	3	3-5
S04	GT_Fly		5-2	4	4-2
S05	Kendo	C	3-5	4	4-5
S06	Balloons		3-5	4	4-5
S07	Lovebird1		6-8	7	7-8
S08	Newspaper		4-6	5	5-6

Correlation with Subjective Tests

■ Fitting curves:



Correlation with Subjective Tests

- Correlation coefficients between the results of different quality metrics and subjective test:

Metric	Spearman rank correlation coefficient (SCC)	Pearson linear correlation coefficient (PCC)
PSNR	0.6357	0.6554
SSIM	0.6709	0.7034
VQM	0.6845	0.6805
VIFp	0.7188	0.7475
MS-SSIM	0.8033	0.7916
HV3D	0.8646	0.8566