

SHARPNESS MISMATCH DETECTION IN STEREOSCOPIC CONTENT WITH 360-DEGREE CAPABILITY

Simone Croci, Sebastian Knorr, Aljosa Smolic

V-SENSE Project, School of Computer Science and Statistics, Trinity College Dublin

ABSTRACT

This paper presents a novel sharpness mismatch detection method for stereoscopic images based on the comparison of edge width histograms of the left and right view. The new method is evaluated on the LIVE 3D Phase II and Ningbo 3D Phase I datasets and compared with two state-of-the-art methods. Experimental results show that the new method highly correlates with user scores of subjective tests and that it outperforms the current state-of-the-art. We then extend the method to stereoscopic omnidirectional images by partitioning the images into patches using a spherical Voronoi diagram. Furthermore, we integrate visual attention data into the detection process in order to weight sharpness mismatch according to the likelihood of its appearance in the viewport of the end-user’s virtual reality device. For obtaining visual attention data, we performed a subjective experiment with 17 test subjects and 96 stereoscopic omnidirectional images. The entire dataset including the viewport trajectory data and resulting visual attention maps are publicly available with this paper.

Index Terms— 3D quality assessment, sharpness mismatch detection, 360 video, saliency, virtual reality

1. INTRODUCTION

Stereoscopic 3D (S3D) is a popular instrument to increase the level of immersion in film entertainment and virtual reality (VR) by providing the viewer two different views of a 3D scene for the left and right eye. Asymmetries between the left and right view of a stereoscopic image, however, lead to the so called binocular rivalry [1], which can cause visual discomfort and degrade the quality of experience (QoE) [2].

The motivation of this paper is to automatically assess the quality of stereoscopic images in the presence of binocular rivalry, in particular sharpness mismatch (SM), and to properly highlight such mismatches in order to reduce efforts and time within the post-production process. Therefore, we introduce a novel histogram-based SM detection method (HSMD),

which is one of the main contributions of this paper. We compare the performance of HSMD against two state-of-the-art methods [3, 4] on two datasets: LIVE 3D Phase II [5] and Ningbo 3D Phase I [6], and prove that HSMD models subjectively perceived quality more accurately than the current state-of-the-art in SM detection.

Furthermore, we extend HSMD to omnidirectional images (ODIs) by extracting patches from the spherical representation using the spherical Voronoi diagram as proposed in [7]. An important aspect of ODIs is that typically only a portion of the images, the so called viewport, is displayed when viewed with a head-mounted display (HMD). Thus, the QoE depends on the viewing direction of the end-user and thus the visual attention. We incorporate saliency in the detection of SM in order to take visual attention into account, and to weight SM depending on the likelihood of its appearance in the viewport of the end-users. In the following, we refer to spherical HMSD (S-HSMD) as the extension of HSMD to ODIs.

Finally, for the computation of the visual attention maps, we performed a subjective test with 17 participants in order to collect viewport trajectory data of 96 stereoscopic ODIs. The images including the viewport trajectory data and the computed visual attention maps are made publicly available (see [8]), which is the second main contribution of this paper. To our knowledge, no datasets with stereoscopic omnidirectional images are available currently.

2. RELATED WORKS

Sharpness mismatch detection. Over the recent years, binocular rivalry issues and conflicts of depth cues have been investigated in detail for traditional S3D content, e.g. for cinema screens [9] and 3D-TV [2, 10], and more recently for omnidirectional content for HMDs [11].

In [12], a no-reference quality metric for stereoscopic images was proposed which models the binocular quality perception of the human visual system in the context of blurriness and blockiness. Various artifact detection methods, including two methods for the detection of sharpness mismatch, were introduced in [13] and [14]. Both methods rely on dense disparity estimation, and analyze either high-frequency differences between both views [13], or differences of edges us-

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776.

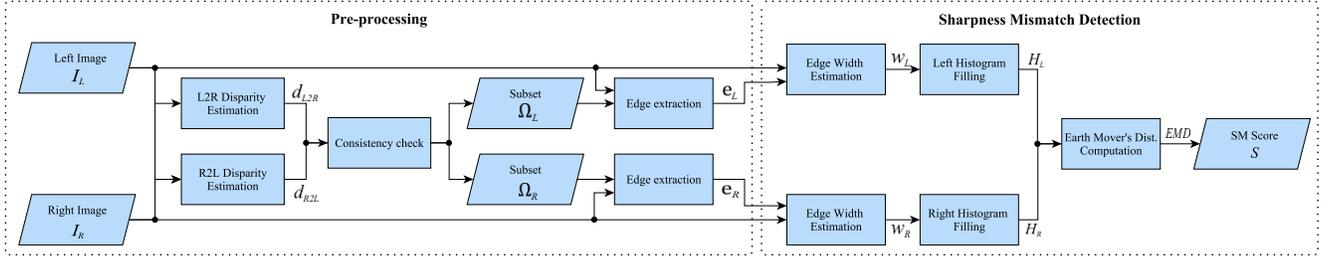


Fig. 1: Overview of the processing steps of HSMD

ing a gradient-based method [14]. Liu et al. [4] presented an automatic no-reference approach to measure the probability of sharpness mismatch (PSM). They demonstrated that the proposed metric outperforms the state-of-the-art metrics for sharpness mismatch. This approach was extended to ODIs in Croci et al. [7] by extracting Voronoi patches and taking visual attention into account.

Visual attention. Visual attention modelling and saliency prediction in traditional visual content are ongoing research topics for decades [15]. Visual attention for 360 content, however, is a relatively new research area with only a few publications in the last decade. The authors of [16] presented a spherical saliency model to compute saliency maps by fusing together static and motion features.

More recently, a testbed suitable for the recording of viewport trajectories of omnidirectional content was introduced in [17]. The authors of [18] introduced a dataset of head movements of users watching 360-videos. Finally, in [19], the authors proposed a new method to transform the collected viewport data into saliency maps.

3. PROPOSED METHOD

3.1. Histogram-based sharpness mismatch detection

Our HSMD approach consists of a pre-processing and an actual sharpness mismatch detection step as illustrated in Figure 1. In the pre-processing step, the disparity maps d_{L2R} from the left to the right view and d_{R2L} in the other direction, are estimated using the Semi-Global Block Matching approach [20]. Since the disparity estimation can be noisy and inaccurate, we apply a consistency check as introduced in [7], which results in the subsets of pixels $\Omega_L \subseteq I_L$ and $\Omega_R \subseteq I_R$ in the left I_L and right I_R image, where each pixel $(x, y) \in \Omega_L$ has a valid correspondence $(x', y) \in \Omega_R$ with $x' = x - d_{L2R}(x, y)$.

Then, edge pixels $e_L \in \Omega_L$ and $e_R \in \Omega_R$ are extracted in both images using the Canny edge detector [21]. For each edge pixel, the edge width and contrast are estimated using the method described in [22]. Based on the edge pixels, two 2D histograms $H_L(c_i, w_j)$ and $H_R(c_i, w_j)$ with edge contrast bins c_i and edge width bins w_j are filled for the left and right view, respectively. Finally, the SM score is ob-

tained by computing the distance between the two histograms. First, 1D edge width histograms $H_L^i(w_j) = H_L(c_i, w_j)$ and $H_R^i(w_j) = H_R(c_i, w_j)$ are extracted from the original 2D histograms for each edge contrast bin c_i . In order to obtain the SM score independent of the amount of edge pixels, i.e. the total area of the histograms, we normalize the 1D edge width histograms with:

$$\hat{H}_{[L,R]}^i = \frac{H_{[L,R]}^i}{A^i}, \quad (1)$$

where $A^i = \max(A_L^i, A_R^i)$, and A_L^i and A_R^i are the areas of the left and right histograms H_L^i and H_R^i : $A_{[L,R]}^i = \sum_{\forall j} H_{[L,R]}^i(w_j)$.

A well-established metric to measure differences between two histograms H_0 and H_1 is the Earth Mover's distance $EMD(H_0, H_1)$ [23]. More precisely, EMD computes the flow f_{ij} which represents the amount that is transferred from bin i in H_0 to bin j in H_1 . Formally, $EMD(H_0, H_1)$ is defined as follows:

$$EMD(H_0, H_1) = \min_{\{f_{ij}\}} \left(\sum_{\forall i} \sum_{\forall j} f_{ij} d_{ij} \right) + \alpha \left| \sum_{\forall i} H_0(i) - \sum_{\forall j} H_1(j) \right|, \quad (2)$$

subject to the following constraints:

$$f_{ij} \geq 0, \quad \sum_{\forall j} f_{ij} \leq H_0(i), \quad \sum_{\forall i} f_{ij} \leq H_1(j), \quad (3)$$

$$\sum_{\forall i} \sum_{\forall j} f_{ij} = \min \left(\sum_{\forall i} H_0(i), \sum_{\forall j} H_1(j) \right), \quad (4)$$

where α is a user-defined parameter, and d_{ij} is the distance between the bins i and j . We define $d_{ij} = |i - j|/N$ with N equal to the number of bins.

The final SM score S is then obtained by summing the 1D histogram distances for each contrast bin c_i weighted by the number of edge pixels as follows:

$$S = \sum_{\forall i} EMD(\hat{H}_L^i, \hat{H}_R^i) \nu_i, \quad (5)$$

with

$$\nu_i = \frac{A_L^i + A_R^i}{\sum_{\forall k} (A_L^k + A_R^k)}. \quad (6)$$

3.2. Extension of HSMD to omnidirectional images

To extend HSMD to ODIs, we apply the method proposed by Croci et al. [7]. First patches are extracted from the ODI using a spherical Voronoi diagram, the subsets of corresponding pixels between the views of each patch are estimated, and then each patch is processed independently by HSMD. Next, the visual attention map is estimated and used in order to weight each patch and compute the global SM score of the ODI as follows:

$$S_{global} = \frac{\sum_{\forall i} g(\Psi_i) S_i}{\sum_{\forall i} g(\Psi_i)}, \quad (7)$$

where Ψ_i is the average pixel saliency inside the patch i , and g is a weighting function that controls the influence of the saliency and can be freely chosen, e.g. an identity function as used in the evaluations in Section 4.2. On the other side, the total amount of patches with SM can be computed with:

$$\sum_{\forall i} \mathbb{1}_{g(\Psi_i)S_i \geq \rho}, \quad (8)$$

where $\mathbb{1}_{g(\Psi_i)S_i \geq \rho}$ is an indicator function, which is equal to one if the inequality $g(\Psi_i)S_i \geq \rho$ is met, and zero otherwise. ρ is a user-defined threshold set to 0.2 for the experiments in Section 4.2. We refer to Croci et al. [7] for a more detailed description of the derivation of these equations.

4. EVALUATIONS

4.1. Evaluation of HSMD

In order to evaluate the performance of the proposed method for traditional S3D content, we compared HSMD against the methods introduced in Narvekar et al. [3] (*Cumulative Probability of Blur Detection*, CPBD) and Liu et al. [4] (*Probability of Sharpness Mismatch*, PSM).

The performance comparison was evaluated based on two datasets: LIVE 3D Phase II [5] and Ningbo 3D Phase I [6]. These two datasets were obtained by introducing different degrees of distortions to some artifact-free stereoscopic reference images. For each image, the datasets provide a subjective difference mean opinion score (DMOS) which was obtained through subjective experiments.

For the comparison of HSMD against the state-of-the-art methods CPBD [3] and PSM [4], we evaluated the correlation between the subjectively obtained DMOS and the SM scores of all three methods by fitting a logistic function to transform the SM scores to DMOS. A well-suited logistic function was proposed by the Video Quality Expert Group in [24] and is defined by

$$DMOS_p(S) = \frac{\beta_1 - \beta_2}{1 + e^{-\frac{S - \beta_3}{\|\beta_4\|}}} + \beta_2, \quad (9)$$

where $DMOS_p$ is the predicted DMOS of the SM score S .

	LCC	SROCC	RMSE	MAE	OR
CPBD [3]	0.8359	0.587	2.054	1.739	-
PSM [4]	0.8542	0.5426	1.945	1.599	-
HSMD	0.8708	0.6296	1.839	1.455	-

Table 1: LIVE 3D Phase II dataset.

	LCC	SROCC	RMSE	MAE	OR
CPBD [3]	0.7069	0.4307	5.091	4.192	0.02
PSM [4]	0.9276	0.7572	2.913	2.094	0
HSMD	0.9548	0.8205	2.152	1.563	0

Table 2: Ningbo 3D Phase I dataset.

To compare the methods, the following performance metrics were applied in order to evaluate how well the logistic function predicts the subjective DMOS: Pearson’s Linear Correlation Coefficient (LCC), Spearman’s Rank Ordered Correlation Coefficient (SROCC), Root Mean Squared Prediction Error (RMSE), Mean Absolute Prediction Error (MAE), and Outlier Ratio (OR). LCC and SROCC measure the prediction accuracy and the monotonicity, respectively. The larger these two metrics are, the more accurate and monotonic the prediction is. For RMSE, MAE, and OR, the smaller the metric, the better the performance of the prediction is. Note that the LIVE 3D Phase II dataset doesn’t provide the standard deviation of the DMOS, which is necessary to compute the OR.

Tables 1 and 2 show the performance metrics for the two datasets. The best values are marked in bold. As can be seen, HSMD outperforms CPBD and PSM for all metrics.

A weakness of HSMD and PSM compared to CPBD is the need of disparity maps. For this reason, geometrical misalignment may negatively influence the analysis.

4.2. Evaluation of S-HSMD

In order to evaluate S-HSMD, we created a dataset of ODIs, and performed a subjective experiment to obtain visual attention data. The ODIs, the viewport trajectory data, and the visual attention maps are available in the supplemental material (see [8]).

Dataset. The dataset consists of 96 stereoscopic ODIs collected from different public sources. The resolution of the ODIs ranges from 1920x960 to 4640x2320 pixels per view. In order to have a large variety, the dataset has the following characteristics: 32 indoor scenes, 51 landscape scenes, 48 scenes containing humans, 47 ODIs with both pole caps covered, 19 ODIs with only the top pole cap covered and 30 ODIs without pole caps, 90 ODIs were captured in native 3D while 6 were post-converted to 3D.

The dataset was captured with a range of different 360°-rigs. These are: Google Odyssey (7), Jaunt rig prototype I and II (7), Panocam POD 3D (9), VUZE VR (5), Nokia OZO (4), customized rig by INVR (3), customized rig by Jumpgate (4), Omnicam 3D (1), customized rig with Mobius cameras (1), unknown 3D rigs (49) and post-converted (6).

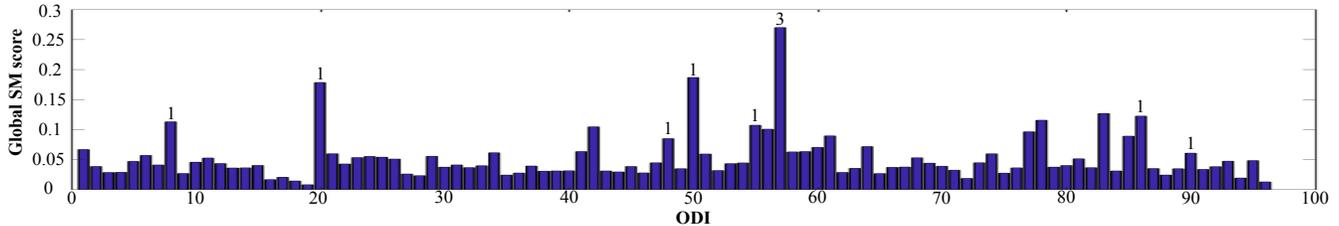


Fig. 2: SM detection in 96 ODIs: global SM scores S_{global} and number of patches with SM.

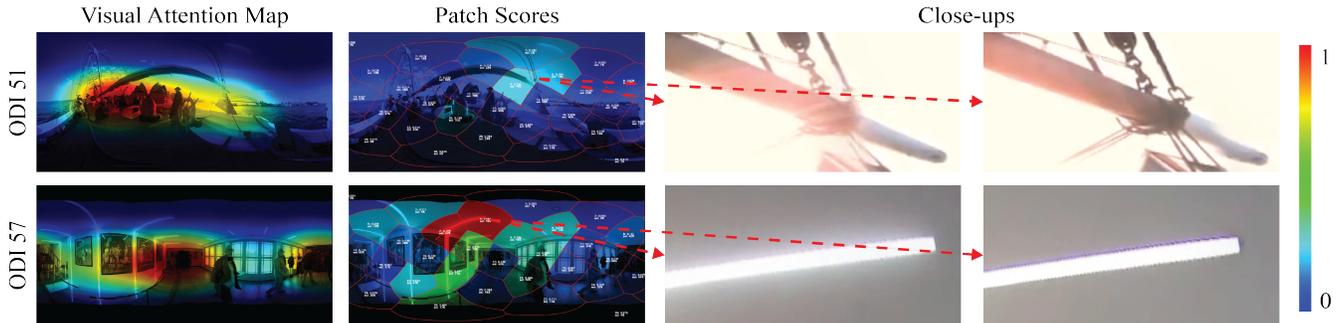


Fig. 3: Examples of ODIs with detected sharpness mismatch. Left to right: visual attention maps (red: high visual attention, blue: low visual attention), overlaid Voronoi patches (red: high SM score, blue: low SM score), close-ups of left and right view.

Subjective test. To compute the visual attention maps we used the method described in Croci et al. [7], and we organized a subjective test with 17 subjects (3 females and 14 males) between 20 and 56 years and with normal stereo vision. The HMD used in the experiment is the Oculus Rift DK2. While the subjects were looking at the images we recorded the viewport center locations on each of the ODIs at the frequency of 75 Hz, assuming that the center point of the viewport corresponds to the visual target location of the user. The test was divided into a training and a test session. During the training session the subjects got familiar with the experiment, while a demo image was displayed. During the test session the 96 ODIs of the dataset were displayed in random order. Each image was displayed for 15 seconds, and according to [19], the data captured during the first second was discarded as it adds trivial information on the starting viewing direction.

Performance evaluation. For each ODI of the dataset, we computed the number of patches with SM (Eq. 8) and the global score S_{global} (Eq. 7), as shown in Figure 2. ODI 57, an indoor scene without pole caps and captured with the Panocam POD 3D, has the largest global score (0.2702) and also the highest amount of patches with detected SM (3). In ODI 42 and 83, despite a relatively high global SM score, no patches with SM have been detected. Here, a large number of patches has high SM scores, but they are still below the defined threshold for SM detection.

For each ODI we visualized the patch scores using the jet colormap (see supplemental material [8]). Figure 3 shows exemplary the visual attention maps and the visualization of the patch scores for ODI 51 (Nokia Ozo) and ODI 57 (Panocam POD 3D). As illustrated in the close-ups in Figure 3, SM

was correctly detected in ODI 57. During the evaluation, we discovered that our approach also detects asymmetric distortions like glares, stitching/blending artifacts, and contamination. ODI 51 shows exemplary the detection of asymmetric glares.

5. CONCLUSION

We proposed a new histogram-based method for SM detection in S3D content (HSMD), compared the performance against two state-of-the-art methods (CPBD, PSM) on two datasets, and proved that HSMD models subjectively perceived quality more accurately than the current state-of-the-art in SM detection. Then, we extended HSMD to omnidirectional images (S-HSMD) and integrated saliency to weight SM according to the visual attention of end-users. The performance of S-HSMD was evaluated on a dataset of 96 ODIs, and the results show that S-HSMD detects SM and related asymmetric distortions like asymmetric glares quite well. In order to obtain visual attention data of the dataset with 96 ODIs, we performed a subjective test with 17 participants. The ODIs including the viewport trajectory data and the visual attention maps are publicly available with this paper (see [8]). To our knowledge, it is the first time that a dataset with stereoscopic omnidirectional content has been provided.

6. REFERENCES

- [1] S. Knorr, K. Ide, M. Kunter, and T. Sikora, “The avoidance of visual discomfort and basic rules for producing “good 3D” pictures,” *SMPTE Motion Imaging Journal*, vol. 121, no. 7, pp. 72–79, 2012.

- [2] M. Lambooi, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 30201, 2009.
- [3] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *International Workshop on Quality of Multimedia Experience*, 2009, vol. 20, pp. 87–91.
- [4] M. Liu, K. Müller, and A. Raake, "Efficient no-reference metric for sharpness mismatch artifact between stereoscopic views," *Journal of Visual Communication and Image Representation*, vol. 39, pp. 132–141, 2016.
- [5] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, Sept 2013.
- [6] X. Wang, M. Yu, Y. Yang, and G. Jiang, "Research on subjective stereoscopic image quality assessment," in *IS&T/SPIE Electronic Imaging: Multimedia Content Access: Algorithms and Systems III*, 2009, vol. 7255, pp. 7255 – 7255 – 10.
- [7] S. Croci, S. Knorr, and A. Smolic, "Saliency-based sharpness mismatch detection for stereoscopic omnidirectional images," in *European Conference on Visual Media Production*, London, UK, 2017.
- [8] S. Croci, S. Knorr, and A. Smolic, "Supplemental material (911MB)," <https://drive.google.com/open?id=1Y0nBXRCaQpWF-TxVRJVol2mdf2D066pn>.
- [9] D. Vatolin, A. Bokov, M. Erofeev, and V. Napadovsky, "Trends in S3D-movie quality evaluated on 105 films using 10 metrics," in *IS&T/SPIE Electronic Imaging: Stereoscopic Displays and Applications XXVII*, 2016, vol. 2016, pp. 1–10.
- [10] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *Journal of Vision*, vol. 11, no. 8, pp. 1–29, 2011.
- [11] S. Knorr, S. Croci, and A. Smolic, "A modular scheme for artifact detection in stereoscopic omnidirectional images," in *Irish Machine Vision and Image Processing Conference*, 2017.
- [12] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 591–602, 2014.
- [13] D. Vatolin and A. Bokov, "Sharpness mismatch and 6 other stereoscopic artifacts measured on 10 Chinese S3D movies," in *IS&T/SPIE Electronic Imaging: Stereoscopic Displays and Applications XXVIII*, 2017, pp. 137–144.
- [14] A. Bokov, D. Vatolin, A. Zachesov, A. Belous, and M. Erofeev, "Automatic detection of artifacts in converted S3D video," in *IS&T/SPIE Electronic Imaging: Stereoscopic Displays and Applications XXV*, 2014, vol. 9011, pp. 901112–901114.
- [15] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [16] I. Bogdanova, A. Bur, H. Hgli, and P.-A. Farine, "Dynamic visual attention on the sphere," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 100 – 110, 2010.
- [17] E. Upenik, M. Reábek, and T. Ebrahimi, "A testbed for subjective evaluation of omnidirectional visual content," in *Picture Coding Symposium*, 2016.
- [18] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in *ACM Multimedia Systems Conference*, 2017, pp. 199–204.
- [19] A. De Abreu, C. Ozcinar, and A. Smolic, "Look around you: Saliency maps for omnidirectional images in VR applications," in *International Conference on Quality of Multimedia Experience*, 2017, pp. 1–6.
- [20] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2005, vol. 2, pp. 807–814 vol. 2.
- [21] J. F. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [22] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, pp. 163–172, 2004.
- [23] O. Pele and M. Werman, "Fast and robust Earth Mover's Distances," in *IEEE International Conference on Computer Vision*, Sept 2009, pp. 460–467.
- [24] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Tech. Rep., ITU, COM 9-80-E, Geneva, Switzerland, 2000.