VISUAL ENTROPY: A NEW FRAMEWORK FOR QUANTIFYING VISUAL INFORMATION BASED ON HUMAN PERCEPTION

Sewoong Ahn, Kwanghyun Lee and Sanghoon Lee

Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea, 120-749.

ABSTRACT

In recent years, how to quantify visualizations of an object and surface displayed in 3D space is now more prominent with a rapid increase in the demand for three-dimensional (3D) content. In order to measure the content information in terms of human visual perception, it is necessary to quantify the visual information in accordance with the human visual system. In this paper, we propose a framework for expressing visual information in bits termed visual entropy based on information theory. The visual entropy of 2D content (2DVE) is composed of texture entropy on the 2D surface and depth entropy based on the monocular cue. In addition to 2DVE, the visual entropy of 3D content (3DVE) includes the depth entropy based on the binocular cue. A series of simulations are conducted to demonstrate the effectiveness of visual entropy, including a performance trade-off between 2D and 3D visualizations measured according to the bitrate.

Index Terms— visual information, 2D and 3D visual entropies, texture entropy, depth entropy, information theory

1. INTRODUCTION

Recent successful technological developments of three-dimensional (3D) displays and image processing have led to an explosive increase in consumers' demands for 3D content. Advanced techniques for reconstruction of virtual 3D space provide viewers with a more realistic viewing experience and deliver vivid sensations of movement and depth. However, a number of important issues on the quantification of content information in terms of human visual perception have not yet been fully investigated. Since the solution to these problems could be used to predict the quality of the experience or compare the difference in the quantity of visual information between 2D and 3D, it is essential to quantify visual information for evaluation of visual content and displays.

The concept of entropy has been applied to quantify content information [1]. In information theory, the entropy is defined as a measure of the uncertainty in a random variable represented by a certain probability density function (PDF), which enables the calculation of the expected value of information in the data. Therefore, a region which contains various color and texture components has high amount of information. However, this conventional approach does not account for the HVS or perception of content information. In this paper, we define a novel visual entropy factor to quantify the attainable information in accordance with human visual perception. To accurately quantify visual information, we investigate the visual sensitivity while reflecting the HVS and statistics of 2D scenes based on information theory.

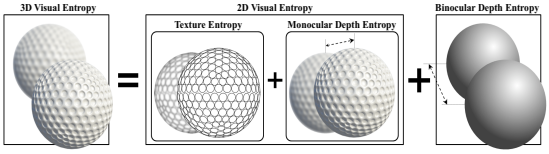

Fig. 1. Conceptual illustration of visual entropy

2. FRAMEWORK OF VISUAL ENTROPY

With respect to 3D perception, The human brain has evolved to compute saliency in real time over the entire visual field [2][3]. Visual information then reaches the human fovea, and then the optic nerve carries this transduced electrical signal to the brain. By interpreting this signal, viewers perceive the texture distribution and subsequently determine the position of objects in terms of depth. They then recognize the 3D distribution of the texture information, i.e., the 3D structure, based on the depth. Therefore, based on the process of human perception in 3D scenes, we can conclude that the visual entropy consists of texture and depth entropy.

The perception of depth information arises from a variety of visual cues [4] including the following: 1) monocular cues that are represented in two dimensions and can be obtained by only one eye.2) binocular cues that are based on the sensory information in three dimensions from both eyes. Defocus blur, a phenomenon in which the area within the depth of field appears sharp and the other areas are blurry, can be used to extract monocular depth cues. When viewing 3D content, humans can perceive depth from binocular cues by rotating their eyes. In order to achieve comfortable 3D viewing, accommodation and vergence exhibit dynamic properties. The process produces the binocular depth cues, which can be obtained by measuring the statistics of a disparity map. Therefore, depth

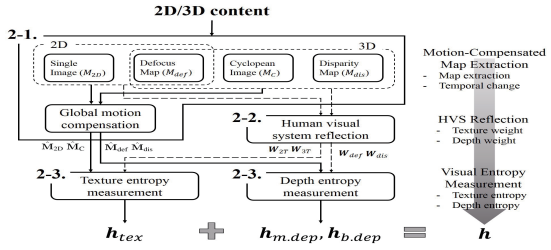

Fig. 2. Block diagram for measuring 2D and 3D visual entropies.

entropy consists of monocular and binocular depth entropy.

The proposed measurement is called 2D visual entropy (2DVE) and 3D visual entropy (3DVE). 2DVE quantifies the texture distribution on a 2D surface and the defocus blur. 3DVE consists of 3DVE information and binocular depth entropy based on the disparity. The schematic representation of visual entropy is shown in Fig. 1. The 2DVE of a 2D golf ball is determined by texture (dimple patterns on the golf ball) and defocus blur. The 3DVE of a 3D golf ball is the sum of the 2DVE and depth entropy based on binocular disparity. According to information theory, the definition of entropy h of a discrete random variable X with possible values $\{x_1, x_2, \cdots, x_n\}$ and $p(X)$ is given by

$$h(X) = -\sum_i p(x_i) \log_2 p(x_i). \quad (1)$$

Here, the entropy of 2D video contents follows the form of (1), but the human visual properties are included for obtaining the pixel value x_i. The process of measuring 2DVE and 3DVE is shown in Fig.2.

2.1. Motion-Compensated Map Extraction

Contents used for map extraction can be divided in to 2D and 3D case. In the 2D content case, the input image ($\mathbf{M}_{2D}$) itself is the texture map, and the defocus map ($\mathbf{M}_{def}$) is used as a depth map and $\mathbf{M}_{def}$ is obtained by the defocus map estimation in [5]. In the 3D content case, the cyclopean image ($\mathbf{M}_C$) is obtained by linear combination of the left and right images in [6]. The defocus map ($\mathbf{M}_{def}$) and the disparity map ($\mathbf{M}_{dis}$) are applied to obtain the depth entropy for reflecting the monocular and binocular cues of 3D depth perception. $\mathbf{M}_{dis}$ was obtained from the left and right frames using the depth estimation reference software [7].

Even though the regions contain complex texture and depth components, the total visual entropy should be significantly reduced when there are high correlations among frames. Here, we reflect such temporal changes in the final maps by measuring the temporal differences in the maps ($\mathbf{M}_{2D}$, $\mathbf{M}_C$, $\mathbf{M}_{def}$, and $\mathbf{M}_{dis}$) between the current and previous frames. Let $G_x(t)$ and $G_y(t)$ be the global motion along the x- and y-axis at the t^{th} frame, respectively, and these are estimated by using the dominant motion detection method in [8]. The maps are then subtracted from the previous frame that is compensated by the estimated global motion ($\hat{\mathbf{M}}(\mathbf{x}, \mathbf{y}, t) = \mathbf{M}(\mathbf{x}, \mathbf{y}, t) - \mathbf{M}(\mathbf{x} - G_x(t), \mathbf{y} - G_y(t), t-1)$). Thus, the motion compensated versions $\hat{\mathbf{M}}_{2D}$, $\hat{\mathbf{M}}_C$, $\hat{\mathbf{M}}_{def}$, and $\hat{\mathbf{M}}_{dis}$ are obtained from $\mathbf{M}_{2D}$, $\mathbf{M}_C$, $\mathbf{M}_{def}$, and $\mathbf{M}_{dis}$, respectively.

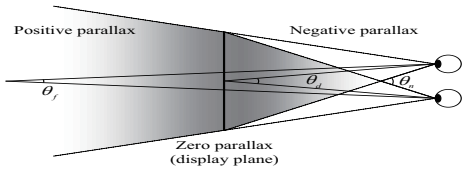

Fig. 3. Angles from the eyes to the farthest point, the disparity plane, and the nearest point, respectively.

2.2. Human Visual System Reflection : Perceptual weights

After computing map extraction, the perceptual weights ($\mathbf{w}_{2T}$, $\mathbf{w}_{3T}$, $\mathbf{w}_{def}$, $\mathbf{w}_{dis}$)are calculated for reflecting the HVS to measure visual entropy. In case of the texture entropy, saliency detection algorithm [8] is used to detect salient regions in both 2D and 3D videos. After that, texture weights ($\mathbf{w}_{2T}$, $\mathbf{w}_{3T}$) are derived by using texture map ($\mathbf{M}_{2D}$, $\mathbf{M}_C$) in Section 2.1, respectively. When measuring texture weights, we have to consider foveation and fusion to consider perceptual properties of human. For 2D contents, the monocular foveation (fov_m) in [9] is included. For 3D contents, the binocular foveation (fov_b) and fusion (fus) from [10] are considered. Finally, the texture weights are modeled as:

$$\begin{cases} \mathbf{w}_{2T} = fov_m(\mathbf{M}_{2D}), & \text{if 2D content} \\ \mathbf{w}_{3T} = fus\big(fov_b(\mathbf{M}_C)\big), & \text{if 3D content} \end{cases} \quad (2)$$

In case of the depth entropy, depth weights ($\mathbf{w}_{def}$, $\mathbf{w}_{dis}$) are derived by using depth map ($\mathbf{M}_{def}$, $\mathbf{M}_{dis}$). For defocus weight $\mathbf{w}_{def}$, it is obtained by using a thin lens model. The defocus blur σ is used to characterize the volume of depth D using an equation in [11]. Thus, the depth D can be expressed in normalized form by the defocus blur σ. Depending on the change in D, σ is the range from the minimum $\sigma_{\min}$ to the maximum $\sigma_{\max}$. The defocus weight $\mathbf{w}_{def}$ ($\ni \forall w_{def}(x,y)$) is obtained from the defocus map $\mathbf{M}_{def}$ ($\ni \forall M_{def}(x,y)$) according to the normalization by using the range between $\sigma_{\min}$ and $\sigma_{\max}$ such that

$$w_{def}(x,y) = \frac{M_{def}(x,y) - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}} = 1 - \frac{FM_{def}(x,y)}{r(V_0 - F)} \quad (3)$$

For disparity weight $\mathbf{w}_{dis}$ ($\ni \forall w_{dis}(x,y)$), it is geometri-

cally determined based on the visual angle in Fig. 3, where

$$w_{\text{dis}}(x,y) = \begin{cases} \frac{\delta(M_{\text{dis}}(x,y))}{\theta_d - \theta_n}, & \text{Crossed disparity} \\ \frac{\delta(M_{\text{dis}}(x,y))}{\theta_d}, & \text{Uncrossed disparity} \end{cases} \quad (4)$$

$M_{\text{dis}}(x,y) \in \mathbf{M}_{\text{dis}}$ is a disparity element at (x,y), $\delta(\cdot)$ is the function of angular disparity, and θ_n, θ_d, and θ_f denote the angles between the eyes and the proximal point, display plane, and distal point in Fig. 3. For example, if $\delta(M_{\text{dis}}(x,y)) = \theta_d - \theta_n$ (nearest point) or θ_d (farthest point), then $w_{\text{dis}} = 1$.

2.3. Visual Entropy Measurement

As shown in Fig. 1, the 2DVE (h_{2D}) is the sum of the texture and monocular depth entropies ($h_{\text{2D}} = h_{\text{tex}} + h_{\text{m.dep}}$), and the 3DVE ($h_{\text{3D}}$) consists of the 2DVE and the binocular depth entropy ($h_{\text{3D}} = h_{\text{2D}} + h_{\text{b.dep}}$) where h_{tex}, $h_{\text{m.dep}}$, and $h_{\text{b.dep}}$ are the texture, monocular depth, and binocular depth entropies.

The texture entropy h_{tex} represents the amount of texture information on the surface of the video. Here, $h_{\text{tex}}(= \mathbf{E}_{\text{T}}\left(\hat{\mathbf{M}}, \mathbf{w}\right))$ is defined as the function according to the maps ($\hat{\mathbf{M}} \in \left\{\hat{\mathbf{M}}_{\text{2D}}, \hat{\mathbf{M}}_{\text{C}}\right\}$) and the weights ($\mathbf{w} \in \{\mathbf{w}_{\text{2T}}, \mathbf{w}_{\text{3T}}\}$). Fig. 4 shows the detailed procedure for calculation $\mathbf{E}_{\text{T}}$. In 2D and 3D videos, the single image $\hat{\mathbf{M}}_{\text{2D}}$ and cyclopean image $\hat{\mathbf{M}}_{\text{C}}$ are applied to the texture entropy measurement by using the texture weights $\mathbf{w}_{\text{2T}}$ and $\mathbf{w}_{\text{3T}}$, respectively. When measuring the texture entropy, the single image or cyclopean image is transformed by discrete cosine transform (DCT) in a unit $N \times N$ block [Fig. 4(a)]. The histogram is then constructed by accumulating the $(u,v)^{\text{th}}$ DCT coefficients in the transformed blocks, where $u, v = 1, 2, \cdots, N$ [Fig. 4(b) and (c)]. At this point, using $\mathbf{w}$, a coordinate transformation including foveation- and fusion-based visual sensitivity is applied. Let $\mathbf{p}_{u,v}$ be the probability mass function (PMF) of the histogram of the $(u,v)^{\text{th}}$ DCT coefficients. Let $\mathbf{B}_{\text{T}}$ be the set of bins of the histogram w.r.t. the wavelet coefficients, and $p_{u,v}(j)$ be the PMF of the j^{th} bin ($\mathbf{p}_{u,v} = \{p_{u,v}(j) | \forall j \in \mathbf{B}_{\text{T}}\}$). Using the PMF $\mathbf{p}_{u,v}$, the amount of visual entropy at the $(u,v)^{\text{th}}$ DCT coefficient can be obtained by (1) ($h_{u,v} = -\sum_{j=-\infty}^{\infty} p_{u,v}(j) \log_2 p_{u,v}(j)$). Moreover, human visual sensitivity varies depending on the frequency band. Thus, the weighted sum of $h_{u,v}$ according to the contrast sensitivity function (CSF) is given as follows:

$$h_{\text{tex}} = \frac{\sum_{u,v} CSF(u,v) \cdot h_{u,v}}{\sum_{u,v} CSF(u,v)}, \quad (5)$$

where $CSF(u,v)$ is the contrast sensitivity of the $(u,v)^{\text{th}}$ DCT coefficient [12]. Unlike texture information, humans tend to perceive depth by mapping it into 3D space. Thus, depth entropy should be determined by the 3D distribution of its defocus or disparity map ($\mathbf{M}_{\text{def}}$, $\mathbf{M}_{\text{dis}}$). Here, the 3D depth entropy $h_{\text{x.dep}}(= \mathbf{E}_{\text{D}}\left(\hat{\mathbf{M}}, \mathbf{w}\right))$ is defined as the function according to the maps ($\hat{\mathbf{M}} \in \left\{\hat{\mathbf{M}}_{\text{def}}, \hat{\mathbf{M}}_{\text{dis}}\right\}$) and the weights ($\mathbf{w} \in \{\mathbf{w}_{\text{def}}, \mathbf{w}_{\text{dis}}\}$), where x($\in$ m,b) determines whether it is monocular depth entropy ($h_{\text{m.dep}}$) or binocular depth entropy ($h_{\text{m.dep}}$). The process of $\mathbf{E}_{\text{D}}$ is explained as follows.

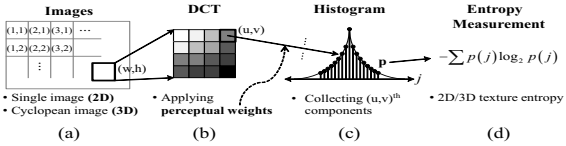

Fig. 4. Process of texture entropy measurement for the single and cyclopean images.

At first, the 3D space is firstly divided into 3D blocks in the quantized levels. The x, y, and z axes in the 3D space are divided into k, respectively. In our experiment, we set k=8. Then, the proportion of the depth elements in each 3D block at each level becoms the PMF of the block. Let $\mathbf{p}_{\text{D}}$ ($\mathbf{p}_{\text{D}} = \left\{p_{\text{D/k}}(x,y,z) = \frac{\mathbf{N}(x,y,z)}{N_{\text{h}} N_{\text{w}}} | 1 \le x,y,z \le k\right\}$) be the PMF with respect to the 3D elements, where $\mathbf{N}(x,y,z)$ and $N_{\text{h}} N_{\text{w}}$ are the number of depth elements at a specific depth amplitude in the 3D block (x,y,z) and the total number of depth elements in the 3D space, respectively. By applying $\mathbf{p}_{\text{D}}$ to the definition of the entropy as in (1), the depth entropy h_{D} can be obtained ($h_{\text{D}} = -\sum_{x,y,z} p_{\text{D}}(x,y,z) \log_2 (p_{\text{D/k}}(x,y,z))$).

The depth weight $\mathbf{w}$ ($\mathbf{w}_{\text{def}}$ or $\mathbf{w}_{\text{dis}}$) in Section 2.2 is then applied to the entropy measurement. Therefore, the depth entropy is given by

$$h_{\text{x.dep}} = -\sum_{x,y,z} w(x,y,z) * p_{\text{D}}(x,y,z) \log_2 p_{\text{D}}(x,y,z), \text{x} \in \{\text{m,b}\}. \quad (6)$$

To quantify total visual entropy of 2D and 3D video, Let $\mathbf{T}$ and $\mathbf{T}^{\text{C}}$ be the sets of the intra and inter frames, respectively. For the intra frame, the visual entropy is determined as the sum of the texture and depth entropies of the frame without motion compensation.

$$h_{\text{2D}}(t) = h_{\text{tex}}(t) + h_{\text{m.dep}}(t), \quad (7)$$

$$h_{\text{3D}}(t) = h_{\text{tex}}(t) + h_{\text{m.dep}}(t) + h_{\text{b.dep}}(t) \quad (8)$$

where t is the frame index. For the inter frame, additional perceived visual entropies of the t^{th} frame based on those of the $(t-1)^{\text{th}}$ frame are also obtained in the same way ((7) and (8)). Since they are obtained using motion compensation, they are denoted by $\Delta h_{\text{2D}}(t)$ and $\Delta h_{\text{3D}}(t)$. Since the maps applied to the texture and depth entropy measurements are the global motion-compensated versions, the total visual entropies of 2DVE $\bar{h}_{\text{2D}}$ and of 3DVE $\bar{h}_{\text{3D}}$ for the total frames are given by

$$\bar{h}_{\text{2D}} = \sum_{t \in \mathbf{T}} h_{\text{2D}}(t) + \sum_{t \in \mathbf{T}^{\text{C}}} \Delta h_{\text{2D}}(t), \quad (9)$$

$$\bar{h}_{\text{3D}} = \sum_{t \in \mathbf{T}} h_{\text{3D}}(t) + \sum_{t \in \mathbf{T}^{\text{C}}} \Delta h_{\text{3D}}(t). \quad (10)$$

Table 1. Average of each element of visual entropy for the test sequences in [13]

sequence	in/out	camera	$h_{\text{tex}}(2D)$	$h_{\text{tex}}(3D)$	$h_{\text{m.dep}}$	$h_{\text{b.dep}}$
Crosswalk2	out	static	2.64	2.70	2.17	2.29
Flower1	out	handheld	3.82	3.55	2.34	2.71
Library2	in	static	2.08	1.68	1.92	0.64
Library3	in	handheld	2.67	2.32	1.84	0.70
Library5	in	moving	2.79	2.65	1.89	0.63
Library7	in	static	2.48	2.23	1.46	0.65
Marathon1	out	static	1.97	2.10	2.11	1.02
Market1	in	handheld	2.46	2.40	2.10	0.48
Metro2	in	static	2.31	2.30	1.87	1.06
Metro3	in	moving	2.40	2.34	1.90	1.25
Restaurant1	in	handheld	1.97	1.96	1.72	0.62
Street2	out	moving	2.74	2.56	2.30	2.88
University1	out	moving	3.19	3.00	2.17	1.28

Table 2. Average values of Table 1 with respect to the shooting condition

	$h_{\text{tex}}(2D)$	$h_{\text{tex}}(3D)$	$\alpha h_{\text{m.dep}}$	$\beta h_{\text{b.dep}}$	h_{2D}	h_{3D}
outdoor	2.84	2.77	2.20	1.46	5.03	6.43
indoor	2.36	2.20	1.84	0.71	4.19	4.75
moving	2.97	2.87	2.11	1.30	5.08	6.27
handheld	2.73	2.56	2.00	1.13	4.73	5.68
static	2.21	2.14	1.98	0.95	4.19	5.07

3. SIMULATION RESULTS

To evaluate visual entropy, the 3D test sequences in [13] were employed. Table 1 lists the average values of each sequence. Using (7), the 2DVE h_{2D} was obtained as the sum of h_{tex} and $h_{\text{m.dep}}$. Using (8), the 3DVE h_{3D} was also obtained as the sum of h_{tex}, $h_{\text{m.dep}}$, and $h_{\text{b.dep}}$. As illustrated in the Table 1, h_{3D} is additionally increased by $h_{\text{b.dep}}$ relative to h_{2D} for the sequences. This indicates that the difference between h_{3D} and h_{2D} remains high when the content has high absolute disparity or complex disparity distribution according to (4) and (6). To understand the relationship between content and entropy, we divided the results into outdoor/indoor and moving/handheld/static scenes, as shown in Table 2. The indoor natural scenes generally had a simple distribution in terms of frequency and depth compared to the outdoor scenes. Thus, the outdoor scenes had higher visual entropy across all entropy components than the indoor ones did due to the widely distributed frequency and depth. For this reason, the difference between 2DVE h_{2D} and 3DVE h_{3D} of the outdoor scene is larger than that of the indoor scene. In addition, visual entropy is affected by the camera and whether it is moving, handheld, or stationary. Camera motion causes background ego motions for which the amount of new information is relatively large. Thus, the average visual entropy of the sequences collected by a moving or handheld camera is higher than that of the static scene.

As an application of the visual entropy, the trade-off between 2D and 3D video contents is investigated according to the compression ratio. If the same bitrate is applied to the 2D and 3D video contents, the 3D video is more severely distorted because of the left and right views. Thus, in the low bitrate case, it is expected that h_{tex} of 3D will be lower than that of 2D. On the other hand, for 3D content, $h_{\text{b.dep}}$ depends on the binocular depth, which is not present in 2D data. Thus,

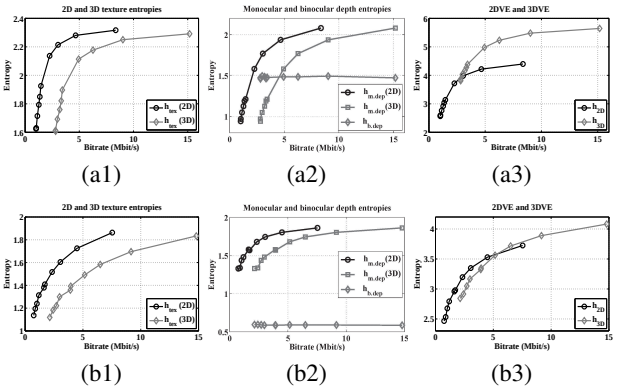

Fig. 5. Visual entropy plotted against bitrate. (1)-column: 2D and 3D texture entropies, (2)-column: monocular and binocular depth entropies, and (3)-column: 2DVE and 3DVE for (a)-row: "Crosswalk2" and (b)-row: "Library3".

the existence of a cross-point in terms of visual entropy is expected with respect to changes in bitrate. In the simulation, we verified the trade-off between the variously encoded 2D and 3D video contents by measuring 2DVE and 3DVE.

As shown in Figs. 5 (1)-column, $h_{\text{tex}}(2D)$ and $h_{\text{tex}}(3D)$ are very similar to each other, but the bitrate of the 3D video is almost twice that of 2D. Thus, $h_{\text{tex}}(2D)$ is larger than $h_{\text{tex}}(3D)$ in terms of bitrate. When the bitrate is increased, the details of high-frequency components can be captured. At this point, the regions of high frequency are the same as the regions of high defocus weight. Therefore, the higher the bitrate, the higher $h_{\text{m.dep}}$ can be shown. These results are represented over all of the sequences. On the contrary, $h_{\text{b.dep}}$ remains consistent regardless of the bitrate. This is the same as the results in [14][15]. Videos are not compressed to a level that disrupts the global structure of the image. Thus, compression artifacts do not affect the global correspondence, and hence, the depth perception is not affected by compression [14][15]. As the bitrate is lower than the cross-point, the visual information will be more readily transmitted by 2D, and vice versa.

4. CONCLUSION

When servicing video contents, it is essential to quantify the visual information in the perceived visual space. A human perceives visual information based on surface textures and depths. Based on this, we proposed a new framework for measuring visual information of 2D and 3D contents in terms of visual entropy. In order to demonstrate the usefulness of 3DVE, we demonstrated the performance trade-off between 2D and 3D entropies according to the bitrate. Moreover, using the 2DVE and 3DVE, we plan to extend this work to encompass visual quality assessment as well as visual information management between 2D and 3D videos.

5. REFERENCES

[1] Ihara, Shunsuke (1993). Information theory for continuous systems. World Scientific. p. 2. ISBN 978-981-02-0985-8.

[2] H. J. Seo and P. Milanfar "Static and space-time visual saliency detection by self-resemblance," *J. Vis.*, vol. 9, no. 12, p. 15, 2009.

[3] L. Itti, C. Koch, and E. Niebur "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[4] I. P. Howard and B. J. Rogers, "Seeing in depth," *Depth perception*, vol. 2, Toronto: I Porteous.

[5] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognit.*, vol. 44, no. 9, pp. 1852-1858, Sep. 2011.

[6] M. J. Chen, C. C. Su, D. L. Kwon, L. K. Cormack, and A. C. Bovik, "Full-Reference Quality Assessment of Stereopairs Accounting for Rivalry," *Asilomar Conf. on Signals, Syst. and Computers*, Nov. 2012

[7] K.-J. Oh, S. Yea, and Y.-S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3D video," in *Proc. Picture Coding Symp.*, May 2009, pp. 1.4.

[8] H. Kim, S. Lee and A. C. Bovik, "Saliency Measurement on Stereoscopic Videos," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1476-1490, Apr. 2014.

[9] S. Lee, M. S. Pattichis and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977-992, Jul. 2001.

[10] K. Lee, A. K. Moorthy, S. Lee and A. C. Bovik, "3D Visual Activity Assessment based on Natural Scene Statistics," *IEEE Trans. Image Processing*, vol. 23, no. 1, pp. 450-465, Jan. 2014.

[11] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 4, pp. 523-531, Jul. 1987.

[12] B. Chitprasert and K. Rao, "Human visual weighted progressive image transmission," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 1040-1044, Jul. 1990.

[13] (2008). *IEEE Standards Association Stereoscopic Database* [Online]. Available: http://grouper.ieee.org/groups/3dhf/

[14] P. Seuntiens, L. Meesters and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Trans. Appl. Perception*, vol. 3, no. 2, pp. 95-109, Apr. 2006.

[15] V. D. Silva, H. K. Arachchi, E. Ekmekcioglu and A. Kondoz, "Toward an impairment metric for stereoscopic video: a full-reference video quality metric to assess compressed stereoscopic video," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3392-3404, Sep. 2013.