Informational Aesthetic Measure for 3D Stereoscopic Imaging

Balázs Teréki[1], Pirkko Oittinen[2], and László Szirmay-Kalos[1]

[1] Department of Control Engineering and Information Technology,
Budapest University of Technology, Hungary
szirmay@iit.bme.hu
[2] Department of Media Technology,
Aalto University School of Science, Finland
pirkko.oittinen@aalto.fi

Abstract. Informational aesthetic measures can quantify the quality of images in terms of information content. Using such measures we can automatically select the "better" solution from two alternatives, or maximizing the aesthetic measure, the "best" option can be found. Such approaches can replace lengthy user interaction sequences tuning camera or light source parameters and automatically propose good camera and lighting settings. In stereoscopic imaging one of the goals is to provide depth perception, so the quality of such images can be characterized by the information of depth obtained by human observers. In this paper, we propose an algorithm to measure the depth information of a stereo image pair, which is based on the entropy of the disparity map, and also takes into account perceptual metrics. With the proposed measures, the parameters of a stereoscopic virtual camera can be optimized, including, for example, the interocular distance.

1 Introduction

Having a virtual scene, we can obtain images by setting studio objects like the camera and light sources differently, which are better or poorer in characterizing the scene. To find good studio object settings that lead to meaningful images, users usually initiate a long manual search based on trial and error. This long procedure can be replaced or helped by automatic algorithms that search for good rendering parameters. To support the search process, the quality of the images should be characterized by numeric values, and we should aim at maximizing these values. Measures expressing the quality of images are called *aesthetic measures* [RFS08]. Generally, aesthetics is related to order and complexity. One option to mathematically interpret such terms is to exploit the tools provided by *information theory*, which associate aesthetics (beauty) with objectively computable information content.

The application of information theory is justified by the observation that the rendering process can be interpreted as a channel communicating the information of the virtual world in form of images (Figure 1). An image is meaningful if it

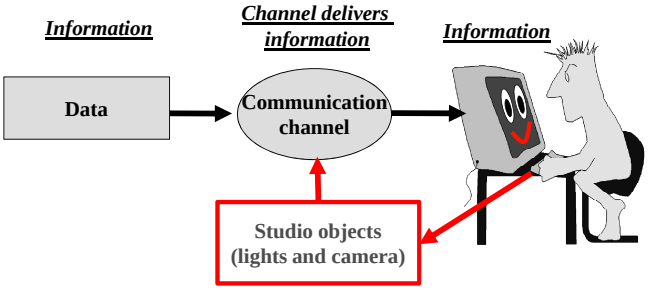


Fig. 1. Information theoretic model of rendering: information of the original data goes through the information channel of the rendering process and is transformed to the image. A meaningful rendering process is expected to provide high information content, i.e. the entropy of the image must be high, and to maximize the correlation between the original data and the information that can be extracted from the image, i.e. the mutual information between the data and the image must be significant.

provides enough information and this information is strongly related to the data stored in the virtual world. In information theory, the amount of information is usually measured by the *entropy* and the strength of the relation between the data and its interpretation by the *mutual information*. In general, the model has focus data that should be communicated and data that can be controlled in order to improve the communication of the focus data. For example, the geometry and material properties of the objects may belong to the set of focus data while light sources and the camera can be set freely. In other scenarios, only the geometry or the density is important, and even the material properties can be subjects of user control. In abstract data visualization, the communication channel can be defined in a very flexible way.

In computer graphics, information theory has been used to describe complexity of 3D scenes [FDABS99], to find the best views to render an object for triangle meshes [VFSH01], volumes, and molecular models. Other applications include the optimal setting of light sources [Gum02], registration of volumes, expressive shading [RBFS10], and automatic definition of transfer functions [RBB*11]. For a review, refer to [SFR*09].

In stereo imaging two images are rendered for the two eyes, which can reproduce *vergence*, i.e. the phenomenon that the same point is visible in two different directions from the two eyes. The vergence depends on the *interocular distance* and the distance of the point of interest. The vergence can be characterized by the angle difference in which a point is visible from the two eyes, or, expressing it from the point of view of the objects, the translation between the two projections of the same point, which is called the *disparity*. Disparity values of points visible in the pixels constitute the *disparity map*.

Significant research efforts have been devoted to the examination of human depth perception and also to the added problems of 3D displays that can reproduce just a limited set of depth cues. Due to the limitation of retinal disparity,

when we focus on nearby objects, the depth perception of distant parts of the scene is lost. In stereo 3D displays, the vergence is reproduced, but other cues like accommodation or motion parallax are missing, which may cause discomfort or *3D fatigue* for the viewer [HGAB08]. To solve these problems, Lang et al. [LHW*10] introduced the concept of non-linear transformation of the disparity map, which belongs to the wider category of image or video tone mapping [RWPD06] and retargeting [SS09], but works on the disparity maps. Niu et al. [NLFJ12] examined the cropping and scaling of stereoscopic image pairs to avoid problems like the removal of a salient object from one of the images destroys its depth perception.

In this paper, we combine two lines of research for 3D stereo imaging: information theory based studio object setting and the perceptual issues of depth perception. First, we survey the most related previous work focusing on the application of entropy to the definition of aesthetics, then the new measure for the aesthetics of stereo images is defined and its computation presented.

2 Entropy based aesthetic measures

What a good camera or light setting means might be arguable, and there are several definitions and approaches focusing on different parameters, like the number of visible faces, the projected area, number of degenerate faces, saliency i.e. points that significantly differ from their environment, etc. However, it is intuitive that the purpose of tuning the camera and the lights is to help the user understand more of the scene, i.e. to increase the information arrived at the observer from the scene through the information channel associated with the rendering process.

The mathematical description of the information channel is given by the information theory. The information associated with a random variable X is described by its *Shannon entropy*:

$$H(X) = -\sum_x p(x) \log_2 p(x) \tag{1}$$

where $p(x) = P(x = X)$ is the probability density of random variable X. This measure shows how much uncertainty is eliminated when the experiment is executed to get a value of the random variable. Alternatively, $H(X)$ is the *expected surprise* since $H(X) = E[-\log_2 p(x)]$ and $-\log_2 p(x)$ is indeed the surprise factor which is large when $X = x$ has a low probability and small when it has a high probability.

In order to apply this measure for images, a correspondence must be found between the important parameters of images and the definition of random variables. It means that the perception of the deterministic image should be interpreted as a random process. For example, when quality focuses on the number and projected area of visible faces, the random variable can be defined as index i of a visible face and its probability as the ratio of its visible area A_i and the area of the screen S, which corresponds to the random model that the eye looks

toward a point that is uniformly distributed on the camera window. With this model, the entropy of the random variable i is:

$$H = -\sum_i \frac{A_i}{S} \log_2 \frac{A_i}{S}. \tag{2}$$

This definition is called *viewpoint entropy* [VFSH01]. Viewpoint entropy introduced this way has several problems. For example, it is based on the tessellation of surfaces, i.e. when every face is decomposed to smaller faces the image might not change at all, but the entropy H increases as stated by the Jensen inequality since the entropy is a concave function of face area A_i. From another point of view, this measure does not take into account perceptual properties. This problem is solved in [VS03] by re-interpreting A_i as the total area of the pixels that are perceptually different from other pixels in the neighborhood. In our work, we also use this approach. We propose a similar measure to characterize the quality of a stereo camera setting. As the objective of stereo imaging is to provide the illusion of depth, our model is based on the *disparity map* instead of the rendered image.

3 Depth information in a stereo image

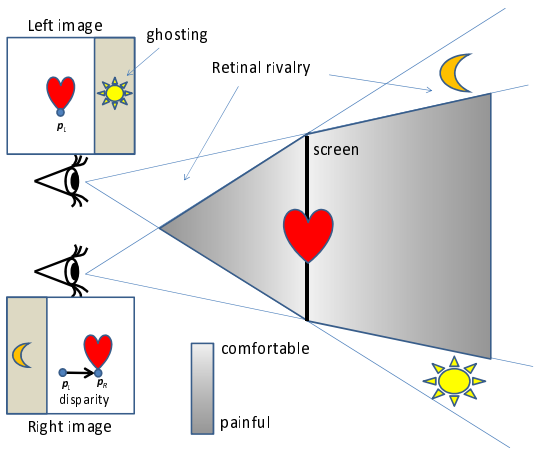


Fig. 2. Stereo imaging and the resulting image pair. Those objects are in the comfort zone where the accommodation cue caused by the distance of the display does not contradict to the vergence cue caused by the binocular disparity. Objects showing up in only one image cause ghosting and have no correct depth perception.

In stereo imaging two pictures are rendered for the two eyes, which have slight differences, called *binocular disparity*, similarly to real-world images captured by

the two eyes that have different positions. These differences are larger when the objects are closer, providing information to the brain to calculate depth in the scene (Figure 2). The difference of the two images can be characterized by the *disparity map* that stores the translation vector in each pixel, that expresses the difference of the locations of this point on the two images. The disparity map is the potentially available maximum information about the depth, which can only be partially utilized by the brain since the human vision system's ability is limited in identifying corresponding point pairs in the two images. Correspondence can be easily established at object corners, where the point has a distinct color that is significantly different from its environment, or generally at salient parts, but is not possible inside faces where the color is constant or slowly changing. Moreover, the human visual system has finite resolution in identifying the differences in depth and has a comfort zone where two points are believed to have a correspondence.

We propose an aesthetic measure that expresses the information content of the disparity map that can be utilized by the human visual system. The requirements for such a measure are the followings:

- It should be zero if two independent images are shown that do not depict the same part scene. Objects in the ghosting region should not have any contribution.
- It should increase when the scene has higher dynamic depth range, i.e. when more objects or objects parts of different perceived depths show up.
- It is expected to be higher when the object depths are more evenly distributed in the perceivable range.
- It should decrease when the disparities leave the comfort zone of the human eye [LHW*10].
- It should be identical for two different settings if the human eye has the same sensation, i.e. it must not be sensitive to the disparity of points for which stereo correspondence cannot be established, or to disparity differences that are too small to be noticeable.

The proposed measure is based on the entropy and the transformation of the information involved in the disparity map. The computation can be regarded as a sequence of the following steps (Figure 3):

1. *Disparity map computation*: We render the scene twice setting the camera according to the two eye positions. Reading back the depth buffer associated with the first image, for each pixel, surface points whose color are in the first image are unprojected and their 3D world positions are reconstructed. Projecting these points with the transformation associated with the second camera, we compare the depth of this projection to the depth stored in the depth buffer of the second camera. If the two depth values are similar, then the point is visible from both cameras, thus a stereo point pair is found. The image space translation is stored in the pixels of the disparity map. Pixels corresponding to points that are not visible in both images and thus cannot be associated with a disparity value are flagged as *invalid*.

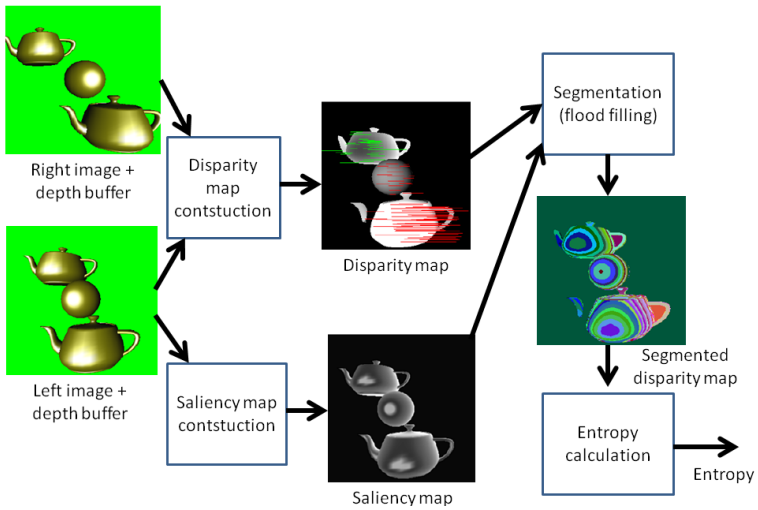


Fig. 3. Steps of the calculation of the aesthetic measure.

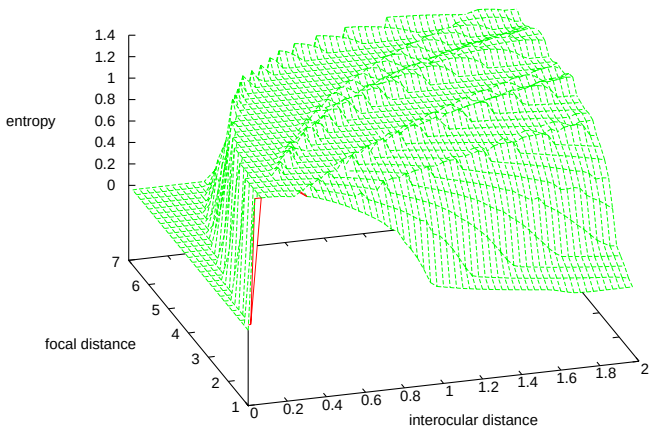


Fig. 4. Entropy with respect to the interocular and focal distances for the Teapot scene of Figure 3.

2. *Saliency map computation*: The human visual system may not be able to match corresponding points if their colors are not characteristically different from their environment. To incorporate this into our model, we also compute a saliency map from the color image. Saliency expresses how different the color of a point is from its neighborhood. Saliency map can be constructed using the difference of Gaussians to express the change of color and intensity and the Sobel filter to express the orientation dependence [LDC06]. We used the method described in [AHES09] that expresses the saliency as

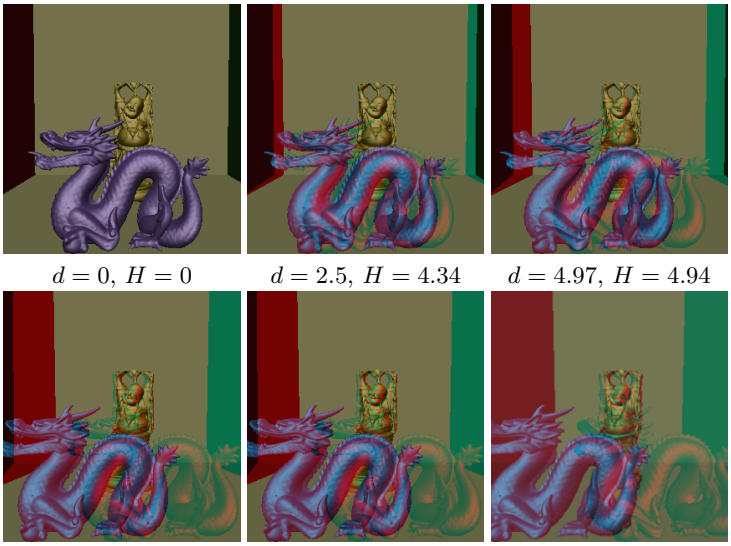
$$S(x, y) = |I_\nu - I_{\omega_{hc}}(x, y)|$$

where I_ν is the mean color, $I_{\omega_{hc}}(x, y)$ is the corresponding image pixel color in the Gaussian blurred version of the original image to eliminate fine texture details and noise. The norm is the L_2 norm in *Lab color space*. The norm of the difference is used since we want only the magnitude of the differences. There is no downsampling, we obtain a full resolution saliency map.
3. *Perception based segmentation of the disparity map*: Using a *flood-fill* type algorithm, the disparity map is partitioned into a finite number of connected regions in a way that each region is a collection of pixels that are perceived by the human eye in the same depth. Without considering perceptional phenomena, partitioning would start with the quantization of the disparity map, then the flood-fill algorithm would identify the connected regions of the quantized map. To improve this approach, we should incorporate the features of human perception into the region filling process and classify two neighboring pixels similar if they cannot be distinguished by the human eye. We consider the ability of matching only salient points and that matching is more effective when different cues are not contradicting, i.e. when the object is in the comfort zone where accommodation and vergence are consistent. Current stereo displays force the eye to accommodate to the display distance, which is a different cue than vergence if the objects are far from this, i.e. when objects are out of the comfort zone. These far away objects cannot be fused by our visual system due to too large retinal disparities. The comfort zone corresponds to small absolute disparity values, thus the tolerance of classifying two pixels as similar is made proportional to an increasing function of the absolute disparity value, taking also the sign into account. The reason of considering the sign is that the human eye has a larger comfort zone for objects that appear in front of the screen than for objects that show up behind the screen. For example, on a 30 foot 2048 pixel resolution cinema screen, the comfort zone is $+30$ (appears behind screen) and -100 (appears in front of screen) screen [LHW*10]. To take both saliency and accommodation into account, we make the tolerance inversely proportional to the saliency. This way, image regions of low saliency and of uncomfortable depth are not decomposed by the segmentation algorithm since the human eye cannot perceive significant depth differences. Note that we are looking for *connected regions* formed by adjacent pixels since the information content of two smaller objects being in the same distance is higher than that of a single larger object.

4. *Entropy computation*: We use equation 2 to compute the information content in the discrepancy map with the following substitutions: A_i is the screen space area (i.e. the number of pixels) of a connected region after segmentation and S is the total screen space area.

4 Results

The discussed method has been implemented in OpenGL environment and tested with different scenes. In these experiments we examined the effect of the interocular distance on the aesthetics of stereo imaging, but it would be also straightforward to modify other parameters, like the light source positions that affect the result via the saliency map, or the distance of the screen, which would modify the ghosting regions and also the comfort zone of the scene.



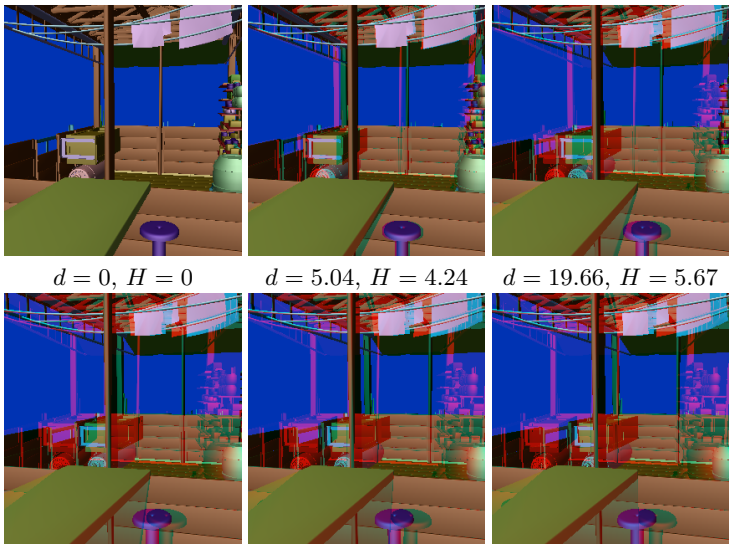
$d = 0,\ H = 0$ $d = 2.5,\ H = 4.34$ $d = 4.97,\ H = 4.94$

$d = 10.87,\ H = 5.54$ $d = 13.07,\ H = 4.88$ $d = 17.04,\ H = 4.82$

Fig. 5. Cornell box with a dragon and a Buddha: Stereo images with interocular distances d and entropy values H.

Figure 5 shows a Cornell box like scene where a dragon is placed close to the eye position while the Happy Buddha is in the background. Interactively modifying the interocular distance, we computed the entropy of the segmented disparity map. When the interocular distance is zero, then the disparity map is constant, so there is only one class in its segmentation, which results in zero entropy. Increasing the interocular distance, the dynamic range of the disparity

map gets larger, which in itself would result in more segmented classes and thus would increase the information. However, ghosting, i.e. objects lacking stereo pairs show up, which reduce information content, and disparity values that are far from zero are quantized more radically, which results in further reduction. The different effects establish an optimum, and we observe an information increase when the interocular distance grows from zero, then after reaching its maximum it starts to decrease. The maximum can be used as an optimal setting.



$d = 0$, $H = 0$ $d = 5.04$, $H = 4.24$ $d = 19.66$, $H = 5.67$

$d = 32.53$, $H = 6.30$ $d = 37.94$, $H = 6.12$ $d = 43.80$, $H = 5.76$

Fig. 6. Tree house: Stereo images with interocular distances d and entropy values H.

Figure 6 shows a Tree house scene where a lot of objects can be found near to the edge of the view frustum. These objects quickly disappear from one of the images causing ghosting and a rapid information loss.

5 Conclusions

This paper proposed the application of information theoretic measures for automatically finding stereo imaging parameters that provide aesthetic 3D images. Based on the recognition, that stereo imaging should be responsible for depth perception, we build our model on the information available in the disparity map. However, to consider also perceptional phenomena, like the ability of the human eye to match point pairs and the contradiction of accommodation according to the screen distance and vergence caused by objects at a different distance, we

non-linearly transformed and quantized the disparity map before the entropy is calculated. The visual validity of the model will be explored in our future work.

Acknowledgments

This work is supported by TÁMOP-4.2.2.B-10/1-2010-0009 and OTKA K-104476.

References

[AHES09] Achanta R., Hemami S., Estrada F., Süsstrunk S.: Frequency-tuned salient region detection. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2009).

[FDABS99] Feixas M., Del Acebo E., Bekaert P., Sbert M.: An information theory framework for the analysis of scene complexity. *Computer Graphics Forum 18*, 3 (1999), 95–106.

[Gum02] Gumhold S.: Maximum entropy light source placement. In *IEEE Visualization* (2002), pp. 275 –282.

[HGAB08] Hoffman D., Girshick A., Akeley K., Banks M.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3 (2008).

[LDC06] Longhurst P., Debattista K., Chalmers A.: A GPU based saliency map for high-fidelity selective rendering. In AFRIGRAPH '06, (2006), pp. 21–29.

[LHW*10] Lang M., Hornung A., Wang O., Poulakos S., Smolic A., Gross M.: Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph. 29*, 3 (2010), 10.

[NLFJ12] Niu Y., Liu F., Feng W.-H., Jin H.: Aesthetics-based stereoscopic photo cropping for heterogeneous displays. *IEEE Trans Multimedia 14*, 3 (2012), 783–796.

[RBB*11] Ruiz M., Bardera A., Boada I., Viola I., Feixas M., Sbert M.: Automatic transfer functions based on informational divergence. *IEEE Trans on Visualization and Computer Graphics 17*, 12 (2011), 1932 –1941.

[RBFS10] Ruiz M., Boada I., Feixas M., Sbert M.: Viewpoint information channel for illustrative volume rendering. *Comput. Graph. 34*, 4 (2010), 351–360.

[RFS08] Rigau J., Feixas M., Sbert M.: Informational aesthetics measures. *IEEE Computer Graphics and Applications, 28*, 2 (2008), 24 –34.

[RWPD06] Reinhard E., Ward G., Pattanaik S., Debevec P.: *High Dynamic Range Imaging.* Morgan Kaufmann, 2006.

[SFR*09] Sbert M., Feixas M., Rigau J., Viola I., Chover M.: *Information Theory Tools for Computer Graphics.* Morgan & Claypool Publishers, 2009.

[SS09] Shamir A., Sorkine O.: Visual media retargeting. In *ACM SIGGRAPH ASIA 2009 Courses* (2009), pp. 11:1–11:13.

[VFSH01] Vázquez P., Feixas M., Sbert M., Heidrich W.: Viewpoint selection using viewpoint entropy. In *Proceedings of the Vision Modeling and Visualization Conference* (2001), vol. 1010, p. 01.

[VS03] Vázquez P., Sbert M.: Perception-based illumination information measurement and light source placement. *Computational Science and Its Applications – ICCSA* (2003), 988–988.