Binocular Suppression Based Visual Masking Model for Stereo Image Watermarking

Yana Zhang[1,2], Kanza Khan[2], Lingling Lv[1], Pamela Cosman[2]

[1]Communication University of China, Beijing, China
[2]Department of Electrical and Computer Engineering, University of California, San Diego, USA
zynjenny@cuc.edu.cn, kanzakhan87@gmail.com, cucling@sina.com, pcosman@eng.ucsd.edu

Abstract: Digital watermarking algorithms for stereo image and video have been simple extensions of monoscopic watermarking methods to two views, and have not considered the aspects of the human visual system which are particular to stereo perception. One key aspect is binocular suppression, which says that high quality image data (for example with sharp edges) presented to one eye can suppress low quality information (for example, blurriness) which is simultaneously presented to the other eye, so that the combined subjective visual effect is one of high quality. We exploit this phenomenon by inserting watermark bits into different portions of the left and right image views, so that high quality in a region on one view suppresses the visibility of the watermark in the corresponding region of the other view. Tests with human observers confirm the visual superiority of the approach.

Keywords: Binocular suppression; Digital watermarking; Imperceptibility; Visual masking model

1 Introduction

Image and video copyright protection is an important area of content security. Overall, digital watermarking for 3D is not as mature as that for 2D media. Existing monocular watermarking methods are often extended directly to stereo images.

For 3D images, existing techniques are either view-based [1]-[8] or disparity-based [9]-[16]. Hwang et al. proposed to embed a watermark image into the right view [1]. Disparity information is extracted from both the left image and the watermarked right image using a matching algorithm. The left image and disparity information are transmitted to the receiver, where the watermarked right image is reconstructed from the received left image and disparity data through adaptive matching. A watermark image is finally extracted from this reconstructed right image, thus watermark extraction performance highly depends on the disparity-matching algorithm. In [2], a watermark pattern is warped for each view of the multi-view source video, and embedded to the texture maps of those views in the spatial domain. Niu et al. presented a visual sensitivity model based method for watermarking high definition stereo images in the DCT domain and focus on achieving the embedded watermark of stereo images in a more robust and invisible manner [3]. A stereo image watermarking method was presented based on the concept of smooth and non-smooth blocks [4], intra-relationship and inter-relationship [5]. In the model presented in [6], watermarks are embedded in left and right images of a video frame in the transform domain, which significantly improves the watermark robustness against resolution and coding attacks. Koz et al. embedded the watermark into every frame of multiple views by exploiting monoscopic spatial masking properties of the Human Visual System (HVS) [7]. Since similar blocks usually have the same change trend when two viewpoints are under the same attack, a watermark is embedded in similar areas of each view selected by global disparity [8].

The second kind of stereo image watermarking is disparity-based. Zhang et al. proposed that if embedding a watermark into disparity vectors, the quality of the base-layer reconstructed image would not be affected, and both watermark embedding and watermark extraction could be implemented in the compressed domain [9]. The watermarking scheme proposed in [10] relied on a depth map and used the quantization index modulation method to embed the watermark. Zhu et al. exploited the depth image to choose foreground blocks to embed the watermark since pixels in this kind of object are more probably to be preserved in the warping, therefore the watermark could also be preserved well. [11]. A proper embedding order plays an important role in watermarking Depth-Image-Based Rendering (DIBR) 3D images [12]. The blind watermarking scheme in [12] is robust against JPEG compression and noise adding attacks and could also tolerate large range variations of the depth image during rendering. In [13], the depth perceptual region of interest was defined according to the 3D depth sensation of the HVS. The fact that objects shift only horizontally when the center view is rendered from the left and right views led to a stereoscopic watermarking approach that moves the mean of a horizontal noise histogram [14]. This method is expected to survive 2D-3D conversion under various depth conditions. Reference [15] presented both view-based and disparity-based watermarking schemes to trade between transparency and robustness using hybrid methods (e.g. the IProtect method [16]).

Digital watermarking technology could also be used in stereo coding systems for compression. The disparity obtained from the stereo pair is embedded as a watermark into one image of the stereo pair, and only the

watermarked image is transmitted [17]. This disparity image is also used as a watermark and embedded into the left stereo image based on modifying singular value decomposition [18], or various transforms (Arnold transform [19], fractional Fourier transform [20], DCT [21]).

One of the most important performance criteria in digital image watermarking is imperceptibility. The embedded watermark should be invisible to normal observers. However, some of the 3D watermarking schemes listed above did not pay more attention on this aspect, while the others all tended to directly use HVS visual masking models for 2D images and video. According to binocular suppression (BS) theory, the HVS is insensitive to spatial errors that occur in only one view [22]. This explains the ability of the HVS to compensate for missing information and suppress errors that occur in a single view, while it obtains the necessary information from the other view. BS theory has spurred interest in asymmetric compression. For example, when one view of a video was coded at high quality (40dB) and the other view at lower quality (33dB), the resultant stereo video was indistinguishable from the case where both views of the video were coded at 40dB [23]. In this paper, we propose a visual masking model based on BS in order to enable better imperceptibility in stereo image watermarking.

The rest of the paper is organized as follows. In Section II, we introduce binocular suppression for a visual masking model. Section III provides experiments and results to test the effect of BS in a stereo watermarking system and Section IV concludes the paper.

2 Binocular suppression based visual masking

We begin with a monoscopic visual masking model. The JND (Just Noticeable Difference) value indicates the highest imperceptible threshold for an image, such that watermark insertion with that embedding intensity would be transparent. The basic characteristics of the HVS include frequency sensitivity, luminance masking, contrast masking and texture masking. A visual masking model was originated in [24] for use in compression. This was adapted to watermarking in [25]. We use the model from [25] in this letter with a small modification to the window size in equation (5) as we did in [26].

$\mathrm{JND}_l^{\theta}(i,j)$ denotes the JND value of the monoscopic image. After a Discrete Wavelet Transform (DWT), it is computed as the product of three terms depending on orientation θ, resolution level l, and location (i,j):

$$\mathrm{JND}_l^{\theta}(i,j) = 0.5 \cdot \mathrm{Frq}(l,\theta) \cdot \mathrm{Lum}(l,i,j) \cdot \mathrm{Tex}(l,i,j)^{0.2} \quad (1)$$

$\mathrm{Frq}(l,\theta)$ and $\mathrm{Lum}(l,i,j)$ denote frequency sensitivity and luminance masking of the monoscopic image, respectively:

$$Frq(l,\theta) = \begin{Bmatrix} \sqrt{2} & if\ \theta = HH \\ 1 & others \end{Bmatrix} \cdot \begin{Bmatrix} 1.00 & if\ l = 1 \\ 0.32 & if\ l = 2 \\ 0.16 & if\ l = 3 \end{Bmatrix} \quad (2)$$

$$\mathrm{Lum}(l,i,j) = \begin{cases} 2 - \mathrm{L}(l,i,j) & \text{if } \mathrm{L}(l,i,j) < 0.5 \\ 1 + \mathrm{L}(l,i,j) & \text{others} \end{cases} \quad (3)$$

$$\mathrm{L}(l,i,j) = \frac{1}{256} \mathrm{I}_l^{\mathrm{LL}}(i,j) \quad (4)$$

The third term $\mathrm{Tex}(l,i,j)$ measures the texture activity in the neighborhood of a coefficient:

$$\mathrm{Tex}(l,i,j) = \sum_{k=0}^{3-l} \frac{1}{16^k} \sum_{\theta}^{\mathrm{HL,LH,HH}} \sum_{x=-2}^{2} \sum_{y=-2}^{2} \left[\mathrm{I}_{k+1}^{\theta}\left(y + \frac{i}{2^k}, x + \frac{j}{2^k}\right) \right]^2 \times \mathrm{Var}\{\mathrm{I}_l^{\mathrm{LL}}(i,j)\} \quad (5)$$

Here $\mathrm{Var}\{\mathrm{I}_l^{\mathrm{LL}}(i,j)\}$ is the local variance of the sub-band in a neighborhood (a 5×5 window in our model) corresponding to the location (i,j).

This monoscopic visual masking model could be applied to just one image in a stereo pair, but this is undesirable because then one view is not copyright protected. The model could alternatively be applied independently to each view in the pair. However, the views are highly correlated and the monoscopic model applied independently will tend to choose many of the same locations in the two views for embedding watermark bits, potentially degrading quality. We aim to exploit BS using this model modulated by opposite checkerboards, which will allow high quality un-watermarked portions of one view to suppress the visible watermark in the corresponding regions of the other view. The black and white regions of a checkerboard pattern portray the watermark embedding areas for the left and right images, respectively.

In order to choose the size of blocks in the checkerboard, we consider spatial zones of binocular rivalry. These regions where one view dominates depend upon many factors such as spatial frequency and size [27, 28]. Observing stereo images on a 47″ LG 3D TV at a distance of about 74″ using the side-by-side rendering mode (the size of one view is 960×1080), we visually tested block sizes of 2×3, 4×9, 8×15, and 15×27 in the HL2 sub-band and found that 4×9 produced the best visual quality. Fig. 1(a) shows the original left and right views. Fig. 1(b) shows the JND values in the HL2 sub-band without BS. Both views have similar embedding positions. Fig. 2(c) shows the embedding positions in the HL2 sub-band with BS; here the JND values have been modulated by a checkerboard pattern. If the processing conditions were appropriate for BS to be effective, then this would mean that in each small spatial zone of binocular rivalry, the high quality in the un-watermarked view would suppress the watermark in the corresponding zone of the other view.

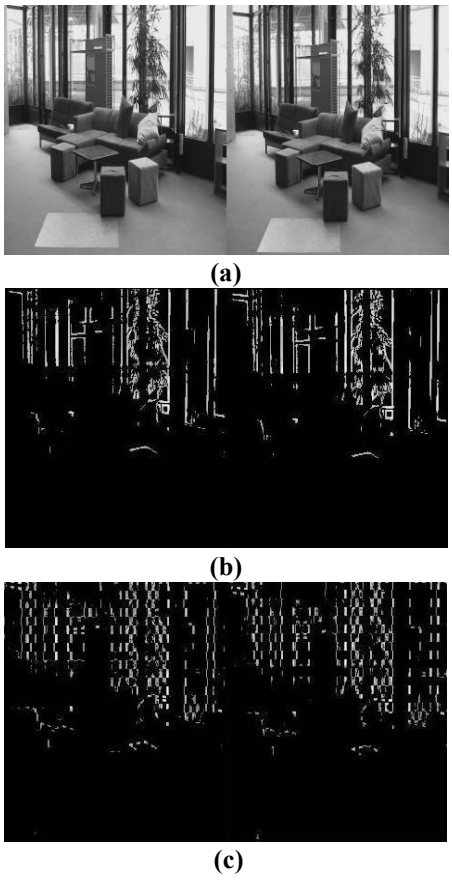

Figure 1 (a) original images **(b)** JND images without BS **(c)** JND images with BS

3 Experiments and results

Watermarking performance contains three aspects: watermark capacity (number of embedding bits), imperceptibility, and robustness (Bit Error Rate (BER) of the watermark bits extracted after attacks). The imperceptibility of two schemes (Scheme A is based on our proposed model and scheme B is based on the model in [25]) was compared under the same number of embedding bits and similar BER through subjective quality assessment.

3.1 Test platform

The effect that the proposed visual masking model has on imperceptibility was tested through the watermarking system of [26]. We refer to the watermark generator, the embedding strategy, and the extracting strategy in [26]. Both left and right view will be embedded with watermarks. The watermarking embedding strength is modulated by JND values from our visual masking model based on BS, as shown in Fig.2.

Ten original stereo images used for the experiment (shown in Fig. 3) have different characteristics and complexities [29]. They are all 8-bit grayscale, JPEG format, with size 1920 x 1080. The copyright image used for watermarking (shown in Fig. 4) is 64 x 64 bits.

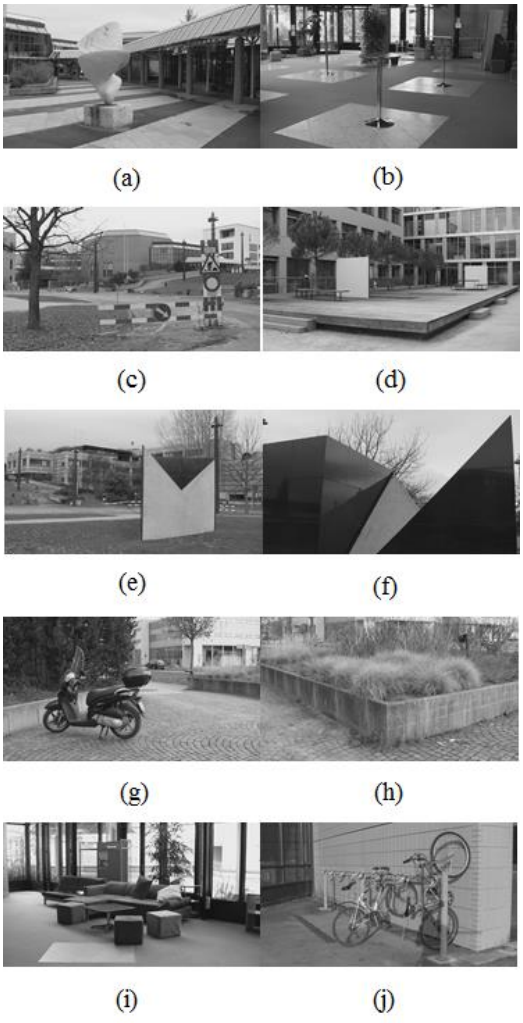

Figure 3 Original Stereo Images

Figure 4 Copyright Image

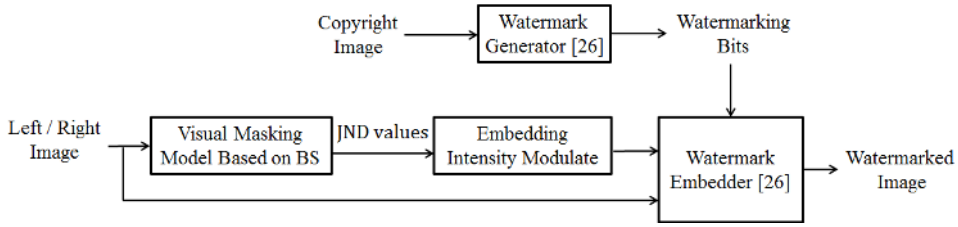

Figure 2 The digital watermarking testing platform

3.2 Parameter setting

The watermark embedding algorithm of [26] uses a modulating factor α which controls the embedding strength, and it controls the trade-off between perceptibility of the watermark and robustness to attack (The larger the strength is, the worse quality and the higher robustness it has). We adjust α so that the two schemes under comparison have similar robustness, and then we conduct a visual experiment to evaluate perceptibility. Also, α was set slightly high to deliberately increase visibility of the watermark. If watermarks were within the imperceptible range, subjects in the quality assessment experiment discussed later would not have detected any differences. When α_{BS}=3.5 and α_{NonBS}=3 for the model with BS and the model without BS, respectively, the average BER difference is 0.07%. Stirmark 4.0 was used for the robustness assessment, choosing JPEG compression and additive noise attacks, each with 10 different levels of attacking intensity. Therefore, 10 stereo images will have 200 BERs after attacks. In Fig. 5, the y-axis shows the BER difference (Scheme B – Scheme A) for an image subjected to an attack. The x-axis shows the index of the image/attack pairs, where they have been sorted in descending order of BER value. All BER differences are between -5% and +5%, and the mean and median values are positive, indicating scheme A has lower BER and is slightly more robust. As will be shown in the next section, scheme A is perceptually much better.

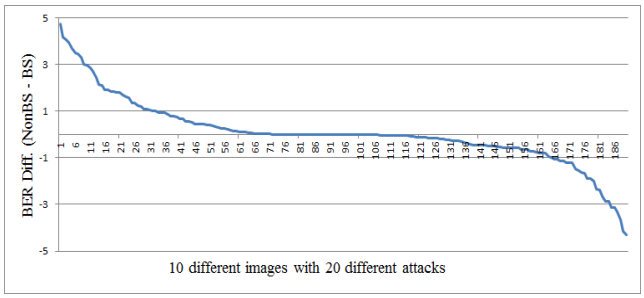

Figure 5 BER Differences for Left Image

3.3 Visual quality assessment

For subjective perception assessment, 17 observers rated the quality of watermarked image pairs created from the 10 images in Fig. 3. In each pair, one stereo image was watermarked with BS and the other was watermarked without BS. Pairs were shown in random order, and which scheme was shown first for each pair was also randomized separately for each subject. Visual quality grading was on a discrete scale from '-2' to '2'. Scores of '+2' or '+1' meant the image seen first was much better or slightly better quality, respectively, than the image seen second. Scores of '-2' and '-1' meant the reverse, and '0' meant there was no difference. The subject was allowed to toggle back and forth until they could reach a decision. If the BS-based model was observed second, these raw scores were converted to final data values by flipping the sign; otherwise, the raw score equals the final data value. Thus, positive data points indicate a preference for the BS-based model.

A total of 16 image pairs were inspected at a distance of roughly 74" from a 47" LG 3D-TV. Six of the 16 pairs were extras that served as checks on subjects' accuracy and consistency. Three of the extra pairs presented identically processed images, hence the subject should score '0' for such pairs. All participants responded properly with '0' scores to these pairs. The other three extra pairs were randomly repeated pairs in reversed order. If the responses to the repeated pairs matched the responses from the initial exposure to the stereo pair, then the subject was consistent. Only five scores from repeated pairs did not match the original scores assigned. Since the number of mismatched scores is small, the inconsistent data was not excluded from statistical analysis.

To compare two schemes, we had 221 visual evaluation scores (17 test subjects × 13 BS versus non-BS image pairs). The data values were distributed as follows: the scores of '-2', '-1', '0', '+1' and '+2' had 6, 10, 13, 96, and 96 occurrences, respectively. Fig. 6 shows the average score for each image pair on the y-axis. The x-axis indicates the watermarked image pair index—corresponding to indices of images in Fig. 3—sorted by average score. All 10 average scores are positive, suggesting images with BS are better quality.

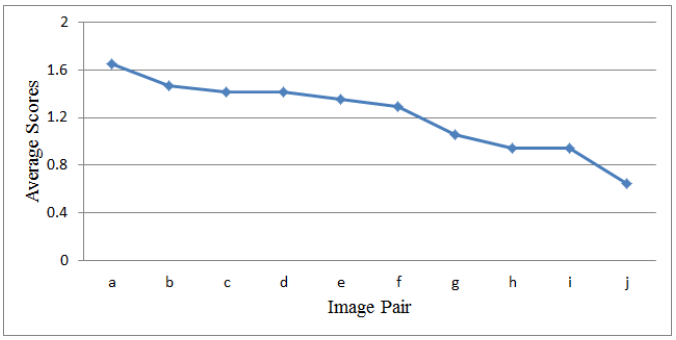

Figure 6 Average score for each image pair

Considering the null hypothesis to be that there is no difference in visual quality between the two watermarking schemes, we used the 221 data points in a z-test. With a p-value ≪ .001, we reject the null hypothesis and conclude that our results are significant at the 1% confidence level.

4 Conclusions

In this paper, we introduce binocular suppression theory into a visual masking model for a stereo image watermarking. This explicitly considers an important effect of the human visual system related to stereo perception. By inserting watermark data into left and right views in a checkerboard fashion, high quality in a region from one view suppresses the visibility of the watermark in the corresponding region of the other view. We conducted a subjective quality test in which the number of embedded watermark bits and the embedding strength were held constant between the two approaches.

All 17 subjects on average preferred the proposed watermarking approach, and it was also preferred on average for all 10 images presented. The results were statistically significant in supporting that images with the proposed visual masking model have higher perceptual quality.

Acknowledgements

Work on this paper was supported by National Science and Technology Program (2014BAH10F00), Communication University of China Research Program (3132014XNG1426), and the U.S. National Science Foundation under Grant CCF-1160832.

References

[1] D.C. Hwang, K.H. Bae, and E.S. Kim, "Stereo image watermarking scheme based on discrete wavelet transform and adaptive disparity estimation," Proc. SPIE 5208, Math. of Data/Image Coding, Compress., & Encrypt. VI, with Appl., pp. 196-205, Jan. 2004.

[2] E. Halici, and A. Aydin Alatan, "Watermarking for depth-image-based rendering." Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, 2009.

[3] Y. Niu, W. Souidene, and A. Beghdadi. "A visual sensitivity model based stereo image watermarking scheme." Visual Information Processing (EUVIP), 2011 3rd European Workshop on. IEEE, 2011.

[4] M. Yu, A. Wu, T. Luo, G. Jiang, W. Zhou, S. Fu, "A new stereo image watermarking method for 3D media." Procedia Engineering 29 (2012): 2399-2404.

[5] T. Luo, M. Yu, G. Jiang, A. Wu, F. Shao, and Z. Peng, "Novel DCT-based blind stereo image watermarking algorithm." Future Wireless Networks and Information Systems. Springer Berlin Heidelberg, 2012. 297-304.

[6] H.Y. Ding, Y.J. Zhou, Y.X. Yang, and R. Zhang, "Robust Blind Video Watermark Algorithm in Transform Domain Combining with 3D Video Correlation," J. of Multimedia, vol. 8, no. 2, pp. 161-167, Apr. 2013.

[7] A. Koz, C. Cigla, and A.A. Alatan, "Watermarking of free-view video," IEEE Trans. on Image Process., vol. 19, no. 7, pp. 1785-1797, 2010.

[8] C. Bai, M. Yu, G. Jiang, Z. Peng, and F. Shao "New HVS-Based Stereoscopic Image Watermarking Algorithm for 3D Media." Advances on Digital Television and Wireless Multimedia Communications. Springer Berlin Heidelberg, 2012. 500-507.

[9] Z. Zhang, Z. Zhu, and L. Xi, "Novel scheme for watermarking stereo video." International Journal of Nonlinear Science 3.1 (2007): 74-80.

[10] P. Campisi, "Object-oriented stereo-image digital watermarking." Journal of electronic imaging 17.4 (2008): 043024-043024.

[11] N. Zhu, G. Ding, and J. Wang. "A novel digital watermarking method for new viewpoint video based on depth map." Intelligent Systems Design and Applications, 2008. Eighth International Conference on. Vol. 2. IEEE, 2008.

[12] Y.H. Lin and J.L. Wu, "A digital blind watermarking for depth-image-based rendering 3D images," IEEE Trans. on Broadcast., vol. 57, no. 2, pp. 602-611, Jun. 2011.

[13] S.L. Fan, M. Yu, G.Y. Jiang, F. Shao, and Z.J. Peng, "A Digital Watermarking Algorithm Based on Region of Interest for 3D Image," 8th Intl. Conf. In Comput. Intellig. & Secur. (CIS), pp. 549-552, Nov. 2012.

[14] J.W. Lee, H.D. Kim, H. Y. Choi, S.H. Choi, and H.K. Lee, "Stereoscopic watermarking by horizontal noise mean shifting," Proc. SPIE 8303, Media Watermark., Secur., & Forens. 2012, pp. 830307, Feb. 2012.

[15] A. Chammem, M. Mitrea, and F. Prêteux, "DWT-based stereoscopic image watermarking." IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2011.

[16] M. Mitrea, F. Prêteux, J. Nunez, (for SFR and GET), "Procédé de tatouage d'une séquence video", French patent no. 05 54132 (29/12/2005), and EU patent no. 1804213 (04/07/2007).

[17] H. Chen and Y. Zhu, "An Effective Stereo Image Coding Method with Reversible Watermarking," J. of Comp. Info. Syst., vol. 8, no. 7, pp. 2761-2768, (2012).

[18] S. Kumar, and R. Balasubramanian, "An optimally robust digital watermarking algorithm for stereo image coding." Recent Advances in Multimedia Signal Processing and Communications. Springer Berlin Heidelberg, 2009. 467-493.

[19] S. Kumar, R. Balasubramanian, and T. Manoj. "Real coded genetic algorithm based stereo image watermarking." IJSDIA 1.1 (2009): 23-33.

[20] B. Gaurav, S. Kumar, R. Balasubramanian, and N. Sukavanam, "Stereo image coding via digital watermarking." Journal of Electronic Imaging 18.3 (2009): 033012-033012.

[21] S. Rawat, G. Gupta, R. Balasubramanian, and M.S. Rawat, "Digital watermarking based stereo image coding," Contemp. Comput., vol. 94, pp. 435-445, Aug. 2010.

[22] W.J.M. Levelt, On Binocular Rivalry, vol. 2, The Hague: Mouton, 1968.

[23] G. Saygili, C.G. Gurler, and A.M. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," IEEE Trans. on Broadcast., vol. 57, no. 2, pp. 593-601, Jun. 2011.

[24] A. S. Lewis, and G. Knowles. "Image compression using the 2-D wavelet transform." Image Processing, IEEE Transactions on 1.2 (1992): 244-250.

[25] M. Barni, F. Bartolini, and A. Piva. "Improved wavelet-based watermarking through pixel-wise masking." IEEE Transactions on Image Processing, 10.5 (2001): 783-791.

[26] Y. Zhang, C. Yang, Q. Zhang, and P. Cosman, "Primal Sketch Based Adaptive Perceptual JND Model for Digital Watermarking," 48th Annual Conf. on Info. Sci. and Syst. (CISS), pp. 1-6, Mar. 2014.

[27] R. Blake, R.P. O'Shea, and T.J. Mueller, "Spatial zones of binocular rivalry in central and peripheral vision," Visual Neurosc., vol. 8, no. 5, pp. 469-478, 1992.

[28] I.P. Howard, Seeing In Depth, Vol. 1, Univ. of Toronto Press, pp. 297, 2002.

[29] 3D Image quality Assessment by Multimedia Signal Processing Group, EPFL, http://mmspg.epfl.ch/3diqa.