Evaluation of 3D-Video Compression for Automotive Stereo Vision Systems

Julian Forster, Daimler AG, Ulm, julian.forster@daimler.com
Roland Schweiger, Daimler AG, Ulm, roland.schweiger@daimler.com
Heiko Neumann, Institute of Neural Information Processing, University of Ulm, heiko.neumann@uni-ulm.de
Albrecht Rothermel, Institute of Microelectronics, University of Ulm, albrecht.rothermel@uni-ulm.de

Abstract

This paper is an evaluation of the distribution of quantized coefficients resulting from the disparity of an automotive stereo vision system. The system captures the scene with two cameras, computes the disparity with Semi-Global Matching and encodes the left view and the disparity for transmission. Real world and synthetic video sequences were used to evaluate the coefficient distributions of the system under normal and challenging weather conditions. The results show, that the quantized disparity coefficients in frequency space have consistently lower entropy compared to the coefficients of the video scenes. Therefore, it is advantageous for the system to compress the disparity instead of one of the two video streams.

1. Introduction

1.1. Motivation

Driver assistance systems depend on a variety of sensors, interconnected with electronic control units (ECUs). Camera-based systems, at the state of the art, use uncompressed proprietary interfaces (e.g. LVDS[1]). In order to get a settled and agreed interface definition, the International Organization for Standardization (ISO) initiated the standardization process of the video communication interface for cameras in road vehicles [1]. An implementation of this kind of interface is based on automotive Ethernet, providing a data rate of 100 Mbit per second. Using uncompressed video streams, the maximum possible resolution of a stereo vision system with 12 bits per pixel and 30 frames per second, would be 360 × 270 pixel. Today, cameras for driver assistance systems have at least a megapixel resolution. In this case, compression of the video data is unavoidable.

In our previous work [2], we proposed a stereo vision system with video and disparity compression, based on a H.264/AVC encoder. The system captures the scene with two cameras (see Fig. 1). A stereo algorithm, in this case Semi-Global Matching [3], computes the disparity image. The disparity and the left view of the stereo system are then compressed and transmitted over Ethernet to the ECU.

With this information the ECU can calculate a 3D-reconstruction of the scene to realize various automotive functions. For example, the car distance Z to an obstacle can be estimated. This can be done with a certain accuracy ΔZ, that depends on the accuracy of the stereo algorithm and the distortion of the video compression method.

[1] low-voltage differential signaling

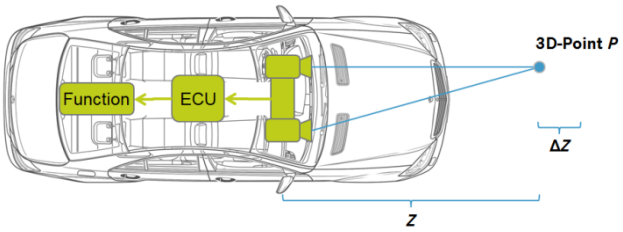


Figure 1 Stereo Vision System: The system captures the scene with two cameras, computes the disparity and encodes the left view and the disparity for transmission to the ECU. With the disparity information an automotive function can calculate the distance Z to an obstacle.

We already examined the peak signal to noise ratio of the left video and the mean disparity error of the system under different rate allocations between the video and disparity streams (see [2]). In [4] we evaluated the impact of compression before stereo matching and compared the results with the computation of Semi-Global Matching on the raw video. Here, we investigate the impact of compressing the resulting disparity map from matching high-resolution input features instead of compressing the two video streams before stereo matching. In order to achieve a quantitative measure we calculate the quantized coefficients of the discrete cosine transform (DCT) of the video and disparity representations and compare the resulting characteristics.

1.2. Related Work

In [5] the distribution of DCT-coefficients in the field of image compression is examined and an approximation of the AC-coefficients with Laplace distributions is proposed. The work of [6] introduces the generalized Gaussian distribution to model the DCT-coefficients more accurate than with Laplace distributions.

The theoretical work of [7] derives a mathematical relation between the distribution of the variance in the image data and a Laplace distribution of the DCT-coefficients. In [8] the development of rate and distortion estimations based on Laplace and Cauchy distributions for a H.264 video encoder is shown.

1.3. Structure

In Section 2 the method used for the evaluation of the coefficients and the necessary mathematical relations are described. The data used for evaluation is shown in Section 3. Finally, Section 4 presents the results and the conclusion is drawn in Section 5.

2. Method

The H.264/AVC standard is based on the integer transform [9], which is derived from the discrete cosine transform. To compare the characteristics of video and disparity compression, we analyze the distributions of the DCT-coefficients and then calculate their entropy for a quantitative evaluation. The discrete cosine transform is defined as (see [10]):

$$Y'_{u,v} = \frac{1}{4}C(u)C(v)\sum_{i=0}^{7}\sum_{j=0}^{7}(Y_{i,j} - 2^{b-1})\cos\frac{(2i+1)u\pi}{16}\cos\frac{(2j+1)v\pi}{16}, \quad u,v = 0,\dots,7, \tag{1}$$

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{u, v} = 0 \\ 1, & \text{u, v} \neq 0 \end{cases},$$

where $[Y_{i,j}]_{i,j=0,\dots,7}, Y_{i,j} \in [0, 2^{b-1}] \cap \mathbb{N}_0$ is an 8×8 luminance macroblock of the video with bit depth b. The transform coefficients $Y'_{u,v} \in \mathbb{R}$ are then uniformly quantized with:

$$Z'_{u,v} = \left\lfloor \frac{Y'_{u,v}}{Q} + \frac{1}{2} \right\rfloor, u,v = 0,\dots,7, \tag{2}$$

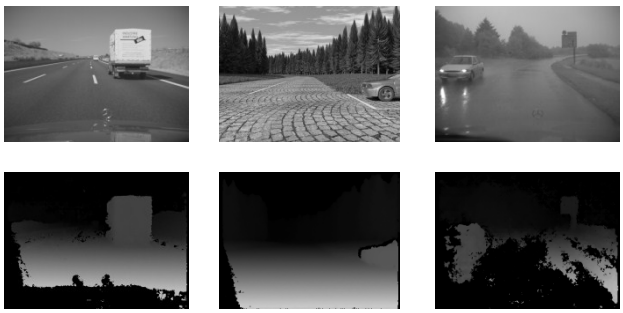
where $Q > 0$ is the quantization parameter.
The entropy of each coefficient [8] is then defined as:

$$H_{u,v} = -\sum_{i=-\infty}^{\infty} P_{u,v}(i)\log_2 P_{u,v}(i), \tag{3}$$

where $P_{u,v}(i) \in [0,1]$ is the probability, that the coefficient $Y'_{u,v}$ is quantized to $Z'_{u,v} = i$.

3. Evaluation Data

Monochrome stereo sequences with a quantization of 12 bit per pixel were used for evaluation (see Fig. 2). The *Day*- and *Rain*-sequences are real world scenes and the *Enpeda*-sequence is a synthetic scene from [11]. The *Rain*-sequences were chosen to also evaluate the system under challenging weather conditions.



(a) (b) (c)

Figure 2 Example images (left view) with corresponding disparity maps of the evaluation sequences: (a) *Day*-, (b) *Enpeda*-, (c) *Rain*-sequence.

4. Results

In Fig. 3 the normalized histograms of four quantized DCT-coefficients $Z'_{1,1}$, $Z'_{3,3}$, $Z'_{5,5}$, $Z'_{7,7}$ of the *Day*-sequence are shown with a quantization parameter $Q = 1$, which is the lowest possible quantization step. The empirical distribution of each coefficient is based on $1.28 \cdot 10^6$ macroblocks and is fitted with a generalized Gaussian distribution [12]:

$$f_{u,v}(x) = \frac{\beta_{u,v}}{2\alpha_{u,v}\Gamma\left(\frac{1}{\beta_{u,v}}\right)} \mathrm{e}^{-\left(\frac{|x-\mu_{u,v}|}{\alpha_{u,v}}\right)^{\beta}}, \tag{4}$$

where $\Gamma(\cdot)$ denotes the mathematical gamma function, $\mu_{u,v}$ is the mean, $\alpha_{u,v}$ is the scale and $\beta_{u,v}$ is the shape parameter.
The scales $\alpha_{u,v}$ of the distributions decrease with increasing frequency. Table 1 lists the scale parameters of the fitted generalized Gaussian distributions. For comparison the distributions of the quantized DCT-coefficients calculated on the disparity of the *Day*-sequence are shown in Fig. 4. In contrast to the textured video, the disparity only consists of gradients separated by sharp discontinuities. The coefficient distributions of the disparity therefore have a smaller scale parameter compared to the distributions calculated on the video.

Table 1 Scale parameters $\alpha_{u,v}$ of the fitted generalized Gaussian distributions (*Day*-sequence)

	$\alpha_{1,1}$	$\alpha_{3,3}$	$\alpha_{5,5}$	$\alpha_{7,7}$
Video	19.3	18.9	10.7	2.5
Disparity	0.80	0.29	0.003	0.004

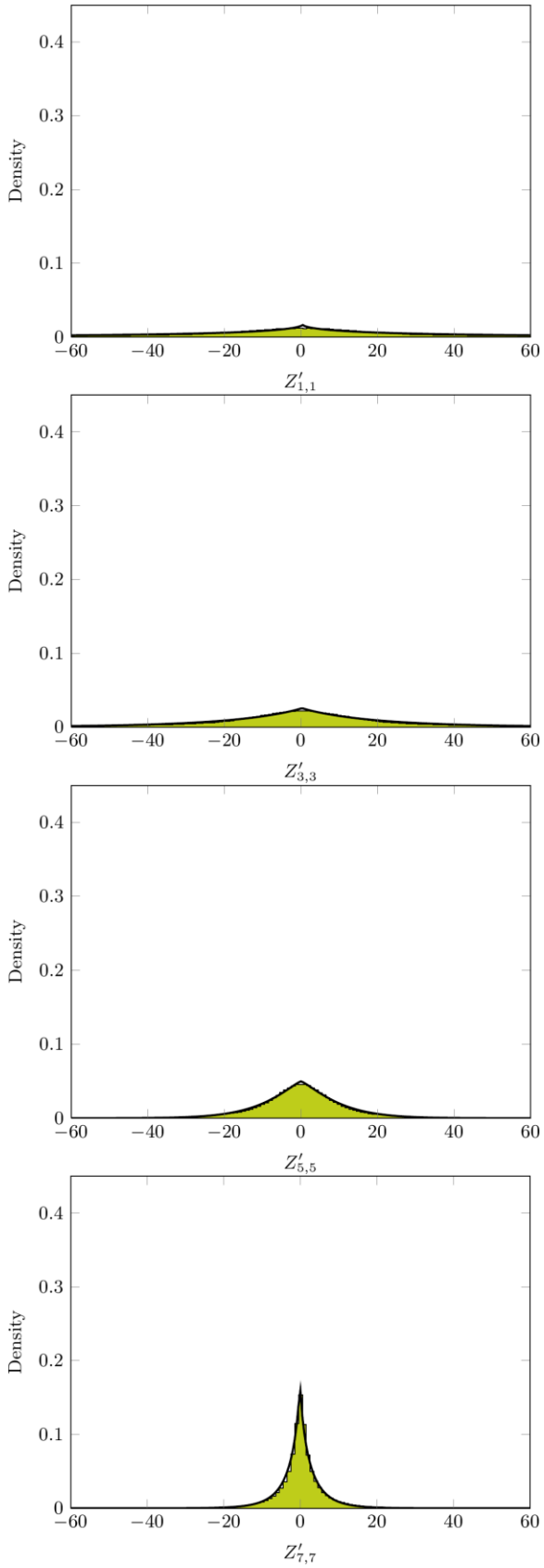


Figure 3 Normalized histograms of four quantized DCT-coefficients $Z'_{u,v}$ on the diagonal of the transformed 8×8 block (*Day*-sequence): The generalized Gaussian distribution is plotted as a black line. The ordinate shows the probability density.

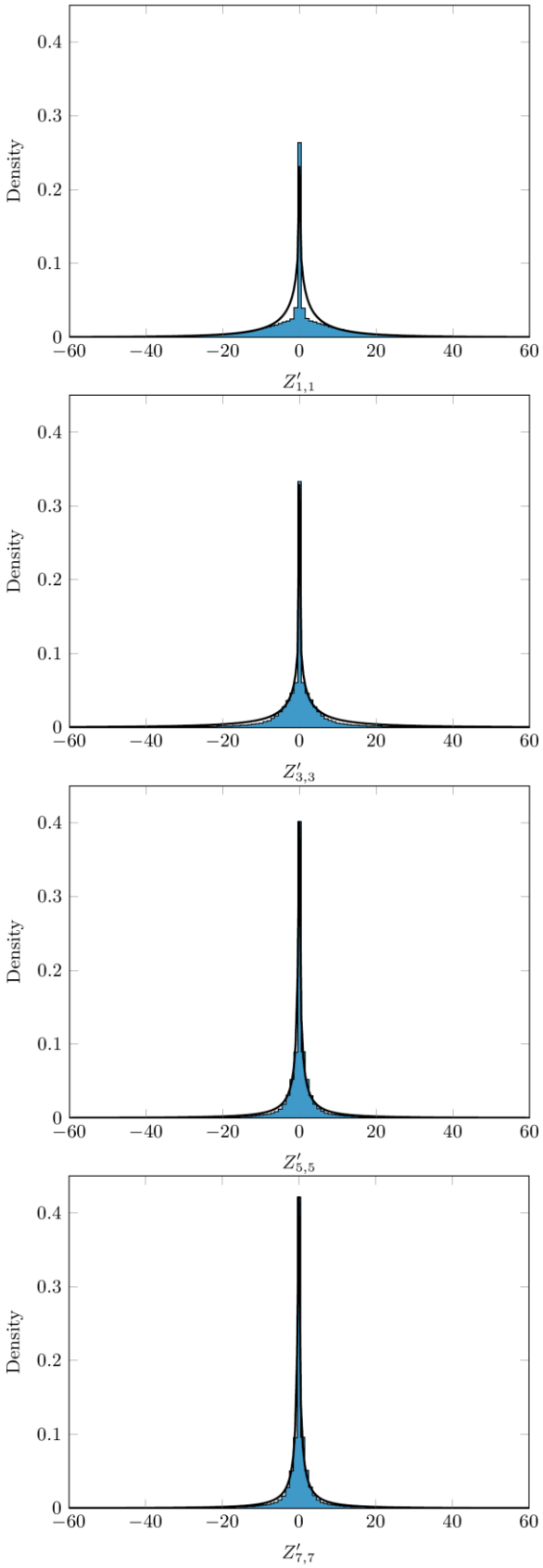


Figure 4 Normalized histograms of four quantized DCT-coefficients $Z'_{u,v}$ on the diagonal of the transformed 8×8 block (Disparity of the *Day*-sequence): The generalized Gaussian distribution is plotted as a black line. The ordinate shows the probability density.

Table 2 Entropy $H_{u,v}$ of the DCT-coefficients (*Day*-sequence)

$v\backslash u$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**14.0**	9.5	8.4	7.9	7.3	6.8	6.3	6.0
1	10.3	**8.8**	8.2	7.6	7.1	6.6	6.1	5.7
2	9.3	8.5	**7.9**	7.4	6.9	6.4	6.0	5.6
3	8.7	8.0	7.6	**7.1**	6.7	6.2	5.8	5.4
4	8.1	7.6	7.2	6.8	**6.4**	6.0	5.6	5.2
5	7.5	7.0	6.7	6.4	6.1	**5.7**	5.3	5.0
6	6.9	6.5	6.3	6.0	5.7	5.4	**5.1**	4.8
7	6.3	6.0	5.8	5.6	5.4	5.1	4.8	**4.5**

Table 3 Entropy $H_{u,v}$ of the DCT-coefficients (Disparity *Day*-sequence)

$v\backslash u$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**10.5**	6.7	5.9	5.5	4.5	4.7	4.7	4.8
1	7.5	**6.1**	5.7	5.2	4.3	4.6	4.5	4.5
2	6.2	5.7	**5.4**	5.1	4.3	4.5	4.4	4.4
3	5.9	5.3	5.2	**5.0**	4.2	4.4	4.3	4.3
4	4.9	4.5	4.4	4.3	**4.0**	3.9	3.8	3.8
5	5.2	4.8	4.6	4.5	3.9	**4.2**	4.1	4.1
6	4.9	4.6	4.5	4.4	3.8	4.1	**4.0**	4.0
7	4.9	4.5	4.4	4.3	3.8	4.0	4.0	**4.0**

Table 4 Entropy $H_{u,v}$ of the DCT-coefficients (*Enpeda*-sequence)

$v\backslash u$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**14.4**	11.5	10.7	10.1	9.6	9.2	8.9	8.6
1	12.0	**10.9**	10.4	9.9	9.5	9.1	8.8	8.5
2	11.3	10.6	**10.1**	9.7	9.3	9.0	8.7	8.4
3	10.8	10.2	9.9	**9.5**	9.2	8.9	8.6	8.3
4	10.4	9.9	9.7	9.3	**9.0**	8.7	8.5	8.3
5	10.1	9.7	9.5	9.2	8.9	**8.6**	8.4	8.2
6	9.8	9.5	9.3	9.0	8.8	8.5	**8.3**	8.1
7	9.6	9.3	9.1	8.9	8.7	8.5	8.3	**8.1**

In Table 2 the entropy of each DCT-coefficient is shown for the left video stream of the *Day*-sequence. The entropy of the 8×8 block is approximately symmetrical to the main diagonal. Beginning with the highest entropy $H_{0,0} = 14$ bit of the DC-coefficient, the entropy is decreasing with increasing frequency until the lowest entropy $H_{7,7} = 4.5$ bit is reached. This is the effect used by lossy video compression to reduce data by quantization of the high frequency coefficients. In Table 3 the entropy values of the coefficients computed on the disparity of the *Day*-sequence are shown for comparison. In general, all entropy values are smaller compared to Table 2 and the mean entropy is 6.8 bit per pixel for the left video in contrast to 4.8 bit per pixel for the disparity. Most of the bits can be saved at the low frequencies. The results for the *Enpeda*-sequence are shown in Table 4 and 5. The mean entropy is 9.4 bit per pixel for the left video and 3.7 bit per pixel for the disparity.

Table 5 Entropy $H_{u,v}$ of the DCT-coefficients (Disparity *Enpeda*-sequence)

$v\backslash u$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**11.1**	5.9	4.9	4.4	3.5	3.6	3.5	3.7
1	6.9	**5.1**	4.6	4.1	3.3	3.4	3.3	3.5
2	5.2	4.6	**4.3**	4.0	3.2	3.2	3.2	3.3
3	5.1	4.2	4.0	**3.8**	3.1	3.1	3.1	3.1
4	4.3	3.5	3.3	3.2	**2.9**	2.7	2.7	2.7
5	4.3	3.6	3.5	3.3	2.8	**2.9**	2.9	2.8
6	3.9	3.5	3.3	3.2	2.7	2.9	**2.8**	2.8
7	4.1	3.4	3.2	3.1	2.7	2.8	2.7	**2.8**

Table 6 Entropy $H_{u,v}$ of the DCT-coefficients (*Rain*-sequence)

$v\backslash u$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**14.1**	9.0	7.4	6.8	6.2	5.8	5.4	5.3
1	9.5	**7.8**	7.0	6.4	6.0	5.6	5.2	5.0
2	8.0	7.3	**6.7**	6.3	5.8	5.5	5.2	4.9
3	7.4	6.8	6.4	**6.0**	5.6	5.3	5.0	4.8
4	6.7	6.3	6.0	5.7	**5.4**	5.1	4.9	4.6
5	6.2	5.9	5.7	5.4	5.2	**4.9**	4.7	4.4
6	5.7	5.5	5.3	5.1	4.9	4.7	**4.5**	4.2
7	5.3	5.1	5.0	4.9	4.7	4.5	4.3	**4.0**

Table 7 Entropy $H_{u,v}$ of the DCT-coefficients (Disparity *Rain*-sequence)

$v\backslash u$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**8.0**	6.5	5.7	5.4	4.3	4.8	4.7	4.7
1	6.6	**6.0**	5.6	5.1	4.1	4.7	4.5	4.5
2	5.8	5.6	**5.2**	5.0	4.1	4.6	4.4	4.3
3	5.4	5.1	5.0	**4.9**	4.0	4.5	4.3	4.2
4	4.3	4.1	4.1	4.0	**3.9**	3.8	3.8	3.7
5	4.8	4.6	4.5	4.5	3.8	**4.3**	4.2	4.1
6	4.7	4.5	4.4	4.3	3.7	4.2	**4.1**	4.0
7	4.7	4.4	4.3	4.2	3.7	4.1	4.0	**4.0**

In Table 6 and 7 the results of the *Rain*-sequence are shown. Here, the mean entropy is 5.8 bit per pixel for the video and 4.6 bit per pixel for the disparity. The entropy of the left video stream of the *Rain*-sequence is generally lower compared to the other sequences. This is the result of lower contrast and blurring effects caused by rain. On the left video of this scene the compression algorithm therefore can reach higher compression ratios with smaller quantization parameters. In contrast, the entropy of the coefficients computed on the disparity is comparable to the results of the *Day*-sequence. The *Rain*-sequence is challenging for the stereo matcher, resulting in a less smooth disparity map with more discontinuities. Consequently the entropy gain between video and disparity is smaller for the challenging weather conditions. However, still more than 1 bit per pixel can be saved.

5. Conclusion

We evaluated the distribution of quantized coefficients resulting from the disparity of an automotive stereo vision system with video and disparity compression, based on a H.264/AVC encoder. As the integer transform of the H.264/AVC standard is derived from the discrete cosine transform, we used an entropy measure based on the DCT-coefficients. Real world and synthetic sequences were used to evaluate the coefficient distributions of the system under normal and challenging weather conditions. For an efficient compression, the disparity maps of the stereo matching should be as smooth as possible. As a stereo algorithm based on regularization with a smoothness constraint, Semi-Global Matching computes disparity maps, that are particularly suitable for video compression. The results show, that the quantized disparity coefficients in frequency space have consistently lower entropy compared to the coefficients of the video scenes. Therefore, it is advantageous for the system to compress the disparity instead of one of the two video streams, in order to achieve a higher compression ratio.

6. References

[1] ISO 17215, Road vehicles – Video communication interface for cameras (VCIC), 2012.

[2] J. Forster, R. Wagner, J. Wuenschmann, R. Schweiger, A. Terzis, A. Rothermel, Video and Disparity Compression for Automotive Stereo Vision Algorithms, In Proc. of ICCE-Berlin 2012 the 2nd IEEE International Conference on Consumer Electronics, Berlin, Germany, September 3–5, 2012.

[3] H. Hirschmueller, Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, In Proc. of CVPR 2005, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 807–814, San Diego, CA, USA, June 20–26, 2005.

[4] J. Forster, X. Jiang, A. Terzis, A. Rothermel, Evaluation of Compression Algorithms for Automotive Stereo Matching, In Proc. of IV 2012 the IEEE Intelligent Vehicles Symposium, pp. 1017–1022, Alcalá de Henares, Spain, June 3–7, 2012.

[5] R. Reininger, J. Gibson, Distributions of the Two-Dimensional DCT Coefficients for Images, IEEE Transactions on Commun. Vol. COM-31, pp. 835–839, June, 1983.

[6] F. Mueller, Distribution Shape of Two-Dimensional DCT Coefficients of Natural Images, Electronic Letters Vol. 29, No. 22, pp. 1935–1936, October 28, 1993.

[7] E. Lam, J. Goodman, A Mathematical Analysis of the DCT Coefficient Distribution for Images, IEEE Transactions on Image Processing, Vol. 9, No. 10, pp. 1661–1666, October, 2000.

[8] Y. Altunbasak, N. Kamaci, An Analysis of the DCT Coefficient Distribution With the H.264 Video Coder, In Proc. of ICASPP 2004 the Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 3, pp. iii-177–iii-180, Montreal, Quebec, Canada, May 17–21, 2004.

[9] I. Richardson, The H.264 Advanced Video Compression Standard, John Wiley and Sons, Ltd, 2nd edition, 2010. ISBN 978-0-470-51692-8

[10] G. K. Wallace, The JPEG Still Picture Compression Standard, IEEE Transactions on Consumer Electronics, 38(1), pp. xviii–xxxiv, February, 1992.

[11] T. Vaudrey, C. Rabe, R. Klette, J. Milburn, Differences Between Stereo and Motion Behaviour on Synthetic and Real-World Stereo Sequences, In Proc. of IVCNZ 2008 the 23rd Int. Conf. on Image and Vision Computing, pp. 1–6, New Zealand, 2008.

[12] S. Nadarajah, A Generalized Normal Distribution, Journal of Applied Statistics, 32(7), pp. 685–694, September, 2005.