LOCALLY LINEAR EMBEDDING-BASED PREDICTION FOR 3D HOLOSCOPIC IMAGE CODING USING HEVC

Luís F. R. Lucas[1,5]*, Caroline Conti*[1,2]*, Paulo Nunes*[1,2]*, Luís Ducla Soares*[1,2]*, Nuno M. M. Rodrigues*[1,3]*, Carla L. Pagliari*[4]*, Eduardo A. B. da Silva*[5]*, Sérgio M. M. de Faria*[1,3]

[1]Instituto de Telecomunicações; [2]ISCTE, Inst. Univ. de Lisboa; [3]ESTG, Inst. Pol. de Leiria, Portugal; [4]DEE, Inst. Militar de Engenharia; [5]PEE/COPPE/DEL/Poli, Univ. Federal do Rio de Janeiro, Brazil;
e-mails: luis.lucas,eduardo@smt.ufrj.br, caroline.conti,lds,paulo.nunes@lx.it.pt, carla@ime.eb.br, nuno.rodrigues,sergio.faria@co.it.pt

ABSTRACT

Holoscopic imaging is a prospective acquisition and display solution for providing true 3D content and fatigue-free 3D visualization. However, efficient coding schemes for this particular type of content are needed to enable proper storage and delivery of the large amount of data involved in these systems. Therefore, this paper proposes an alternative HEVC-based coding scheme for efficient representation of holoscopic images. In this scheme, some directional intra prediction modes of the HEVC are replaced by a more efficient prediction framework based on locally linear embedding techniques. Experimental results show the advantage of the proposed prediction for 3D holoscopic image coding, compared to the reference HEVC standard as well as previously presented approaches in this field.

Index Terms— 3D holoscopic image coding, image prediction, locally linear embedding, HEVC

1. INTRODUCTION

Three-dimensional (3D) holoscopic imaging, also known as integral and plenoptic imaging, is a prospective acquisition and display solution for providing true 3D content and enabling a more natural and fatigue-free 3D visualization. Furthermore, several camera setups have been seeking to improve the performance of this technology for 3D content acquisition [1] and various advantages of employing a 3D holoscopic system have been identified, for instance: i) a single aperture camera is needed for 3D content acquisition; and ii) new post-production degrees of freedom are allowed, such as controlling the plane of focus and the perspective angle.

However, a critical factor for introducing this technology into the consumer electronics market is to provide 3D holoscopic content with convenient viewing resolution. To fit the high definition (HD) requirements, acquisition and display devices with much higher resolution are needed [2]. Consequently, coding tools for this particular content are desirable for efficient storage and delivery of such large amount of data.

Although the most recent video coding standard, High Efficiency Video Coding (HEVC) [3], is able to fulfill the current performance requirements for HD and ultra-HD video coding, previous authors' works [2, 4] have shown that further improvements are still possible for 3D holoscopic content. Notably, by integrating an efficient prediction scheme to exploit the inherent correlations of this content.

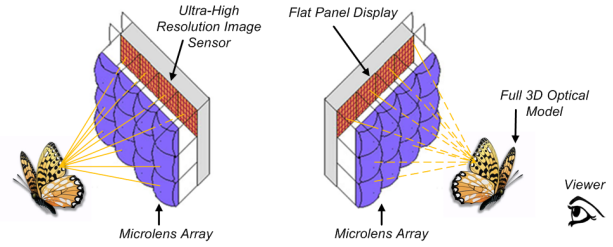
In this context, a novel prediction framework for 3D holoscopic image coding in HEVC is proposed, by using an alternative method based on locally linear embedding (LLE) [5]. Previous research work on intra prediction based on neighbour embedding methods, like LLE, has shown to improve the prediction results of the H.264/AVC standard, for most natural images. The proposed prediction framework combines the LLE-based method with planar, DC and a reduced set of directional prediction modes. Experimental results demonstrate that the proposed prediction framework highly improves the rate-distortion performance when compared to the reference HEVC prediction model.

The remainder of this paper is organized as follows: Section 2 describes the structure of the 3D holoscopic content and presents a brief review on 3D holoscopic image coding solutions. Section 3 describes the LLE-based prediction method as suggested in [5]. The proposed prediction framework for HEVC is described in Section 4. Section 5 presents the experimental results, and Section 6 concludes the paper.

2. 3D HOLOSCOPIC IMAGE REPRESENTATION AND CODING

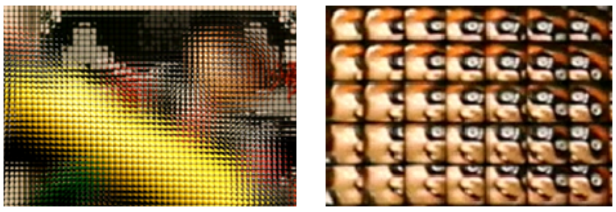
3D holoscopic imaging is a type of light field technique since it allows recording light intensity and direction information of a captured 3D object. In this section, a brief description of this type of content is provided as well as a brief review on some relevant 3D holoscopic image coding solutions.

This work was funded by FCT - "Fundação para a Ciência e Tecnologia", Portugal, under PEst-OE/EEI/LA0008/2013 project, SFRH/BD/79553/2011 and SFRH/BD/79480/2011 grants, and by CAPES/Pro-Defesa, Brazil, under grant 23038.009094/2013-83.



(a) (b)

Fig. 1. A 3D holoscopic imaging system: (a) acquisition side; (b) display side.



(a) (b)

Fig. 2. Holoscopic image captured with a 250 μm pitch microlens array: (a) full image with resolution of 1920×1088; (b) enlargement of 196×140 pixels showing the micro-images.

2.1. 3D Holoscopic Content Structure

The principle of 3D holoscopic imaging was firstly proposed in [6] and referred to as integral photography. Basically, a 3D holoscopic system comprises a regularly spaced array of small spherical micro-lenses, known as a "fly's eye" lens array, which can be used in both acquisition and display of the 3D holoscopic content, as shown in Fig. 1.

In the acquisition side (Fig. 1a), each micro-lens works as an individual small low resolution camera conveying a particular perspective of the 3D object at slightly different angles. As a result, the planar intensity distribution representing a 3D holoscopic image comprises a two-dimensional (2D) array of micro-images, as illustrated in Fig. 2.

Due to the small angular disparity between neighbouring micro-lenses, the recorded 3D holoscopic image presents a significant spatial redundancy between neighbouring micro-images. Section 5 shows that the proposed coding scheme based on LLE intra prediction exploits this non-local spatial correlation and is able to improve the coding efficiency.

2.2. Previous Work on 3D Holoscopic Image Coding

Several 3D holoscopic image coding schemes have been proposed in the literature trying to exploit the particular structure of this type of content. In [7], a hybrid compression scheme was presented, which combined two-dimensional Discrete Wavelet Transform (2D-DWT) and two dimensional discrete cosine transform (2D-DCT). In this scheme, the 2D-DWT was applied to each individual micro-image and, then, the 2D-DCT was applied on sets of 2D-DWT coefficients from neighbouring micro-images. The resulting 2D-DCT coefficients were then quantized and entropy coded. Similarly in [8], the authors proposed to decompose the 3D holoscopic image into several viewpoint images by extracting one pixel (from each micro-image) which had the same relative position inside the micro-image. Hence, a 3D-DWT was applied to a stack of these viewpoint images. Thus, the lower frequency bands were assembled and compressed using a forward two-levels 2D-DWT followed by arithmetic encoding. In [9], the authors presented a comparison between the performance of JPEG2000 Part 10 and H.264/AVC standards for encoding sets of micro-images and viewpoint images. These sets of micro-images (and also viewpoint images) were put together along the third dimension (or temporal dimension) to be then coded using the JPEG2000 Part 10 and H.264/AVC.

Since the 3D holoscopic image is actually a 2D image (comprising 3D information), it is still possible to encode a 3D holoscopic image entirety (without decomposing it into sets of micro-images or viewpoint images) with a regular 2D content encoder. In this sense, in the author's previous work [2], a self-similarity (SS) compensated prediction was introduced into the HEVC encoder to efficiently handle 3D holoscopic images. Similar to motion compensation, the SS compensation used as reference the previous coded and reconstructed area of the current frame itself. Then, it tried to compensate the aforementioned non-local spatial correlation between neighboring micro-images. Given the resemblance of this coding scheme (in [2]) and the proposed LLE-based prediction scheme, a comparison between both will also be carried out in Section 5.

3. INTRA IMAGE PREDICTION BASED ON LOCALLY LINEAR EMBEDDING

Nonlinear dimensionality reduction methods have been recently investigated for intra image prediction. In [5], two intra prediction modes based on the locally linear embedding (LLE) and nonnegative matrix factorization (NMF) methods are presented and evaluated in the H.264/AVC framework. The main idea of these prediction methods is to estimate the coding block by using a linear combination of k-nearest neighbour (k-NN) patches, determined in the causal coded and reconstructed region of the image. While LLE-based approach uses a constraint which forces the sum of the weighting coefficients to be equal to 1, NMF imposes a linear combination using nonnegative weights.

At its core, LLE [10] is a constrained optimization algorithm which maps the high dimensional nonlinear data into a coordinate system of lower dimensionality. LLE tries to preserve the local linear structure of the high dimensional data in the low dimensional representation. To characterize the local linear geometry, each data point is approximated as a linear combination of its nearest-neighbours. The sum to one con-

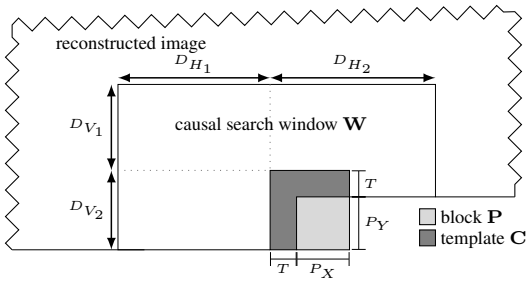


Fig. 3. Search window used by LLE-based prediction.

straint on the weighting coefficients, forces the approximation of each data point to lie in the linear subspace spanned by its nearest neighbours.

In the LLE-based prediction method [5], only the template $\mathbf{C}$ (see Fig. 3) is approximated as a linear combination of its k-NN templates. A causal search window $\mathbf{W}$, where the k-NN templates are searched, is defined. The estimated linear coefficients are then used to estimate the unknown pixels of the block $\mathbf{P}$ being predicted, by combining the co-located pixels of the block associated to each NN template.

Consider the N-pixel region $\mathbf{S}$, the union of the unknown N_p-pixel block $\mathbf{P}$ being predicted, and the known N_c-pixel template $\mathbf{C}$. Let vector $\vec{b}$ be composed by the N pixel values of region S, stacked in a column (assuming the zero value for unknown samples of block $\mathbf{P}$). Also, let an $N \times M$ matrix $\mathbf{A}$, represent a basis dictionary, by stacking all the patches similar to region $\mathbf{S}$, which exist in the causal search window $\mathbf{W}$, as shown in Fig. 3.

The dictionary matrix $\mathbf{A}$ (and vector $\vec{b}$), can be separated into two vertically concatenated sub-matrices $\mathbf{A_c}$ and $\mathbf{A_p}$ (and two vectors $\vec{b}_c$ and $\vec{b}_p$), corresponding to the pixels in the spatial location of template $\mathbf{C}$ and predicting block $\mathbf{P}$, respectively. The k-NN method used in LLE can be viewed as a sparsity constraint which chooses the k closest patches (represented by the columns of $\mathbf{A}_c$), to the template $\mathbf{C}$ (vector $\vec{b}_c$), in terms of Euclidean distance.

Denote $\mathbf{A}_c^k$ as the submatrix of $\mathbf{A}_c$ that contains the selected k-NN patches. The LLE-based prediction can be presented as the following optimization problem:

$$\min_{\vec{\mathrm{x}}_k} \|\vec{b}_c - \mathbf{A}_c^k \vec{\mathrm{x}}_k\|_2^2 \quad \text{subject to} \quad \sum_m \vec{\mathrm{x}}_{k_m} = 1 \qquad (1)$$

where $\vec{\mathrm{x}}_k$ represents the k optimal linear coefficients. By using the covariance matrix, $\mathbf{D}_k$, computed for the k-NN templates in $\mathbf{A}_c^k$ in reference to $\vec{b}_c$, the weighting coefficients can be obtained by solving the linear system $\mathbf{D}_k \vec{\mathrm{x}}_k = \vec{1}$ (where $\vec{1}$ is the column vector of ones), and rescaling the weights to sum to one.

By varying the value of k, different sparsity constraints can be tested. The value of k which results in the optimal solution $\vec{\mathrm{x}}_{opt}$, producing the smallest approximation error in the block $\mathbf{P}$, is chosen and explicitly signalled to the decoder. Note that the template-matching (TM) algorithm, commonly used for intra and inter image/video prediction applications [5], can be viewed as a particular case of the LLE-based method, in which the sparsity constraint is $k = 1$.

Given the optimal weighting coefficients derived from the template $\mathbf{C}$ and causal window $\mathbf{W}$, the samples of the prediction block are then computed by $\vec{b}_p = \mathbf{A}_p \vec{\mathrm{x}}_{opt}$.

4. PROPOSED PREDICTION FRAMEWORK FOR 3D HOLOSCOPIC CODING USING HEVC

As shown in Section 2, due to the micro-lens array used in the 3D holoscopic imaging system, 3D holoscopic images present a significant spatial redundancy between micro-images in a neighborhood along the 2D array. This kind of non-local spatial correlation has been motivating the research on alternative coding methods, such as the LLE-based intra prediction, which estimates the current coding block samples using a linear combination of causal patches with a low signalling overhead. Therefore, it is expected that the proposed prediction framework based on LLE may improve the prediction of 3D holoscopic content, compared to the conventional intra prediction framework of current 2D coding standards.

Due to the performance advantage of the current state-of-the-art in video coding technology, HEVC, compared to the previous standards, it was considered as the basis for incorporating the proposed LLE-based prediction framework. Furthermore, the new HEVC coding tools designed specifically for high resolution images, *e.g.* larger and more flexible blocks sizes (up to 64×64), are also more appropriate for coding the 3D holoscopic content.

The HEVC intra prediction is based on 33 directional modes plus the planar and DC modes, as shown in Fig. 4. For chrominance, only the planar, vertical, horizontal, DC and the corresponding luminance mode can be used. The basic unit for compression, the coding unit (CU) is a square block with up to 64×64 pixels, which can be recursively split into four smaller CUs with size of down to 8×8. When intra prediction is used, each CU contains one equally sized prediction unit (PU), except in the smaller CU level, in which the PU can be further segmented into four smaller blocks of size 4×4. This PU carries information about the partitioning of the CU and the signalling for the chosen prediction mode.

In this proposal, the LLE-based method was implemented in the HEVC prediction framework by replacing some intra prediction modes. To avoid the explicit signalling of the optimal number of NN templates, given by k_{opt}, eight directional prediction modes of the HEVC were replaced (represented in Fig. 4 by the dashed lines and bold numbers). These replaced modes are uniformly spaced not to prejudice any direction in particular. The values $k = 1, \ldots, 8$, are tested for the k-NN method used in the LLE-based prediction. The correspondence of these values with the intra prediction mode is given by $k = (mode + 1)/4$.

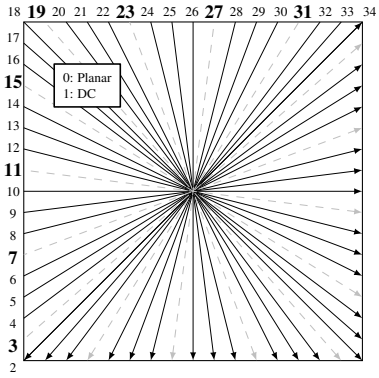
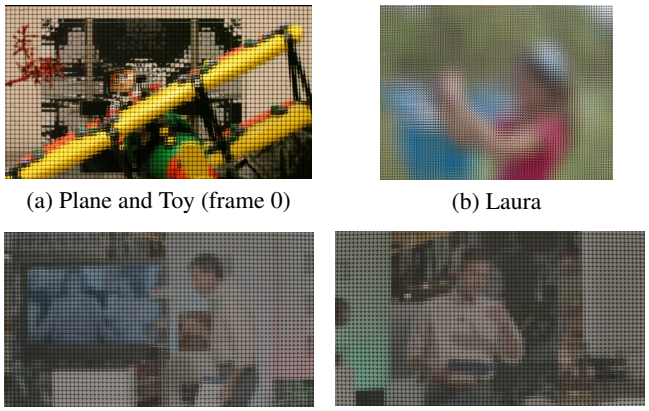


Fig. 4. Set of 35 prediction modes used in the HEVC algorithm. The directional modes represented by dashed lines were replaced by the proposed LLE-based mode.



(a) Plane and Toy (frame 0) (b) Laura

(c) Demichelis Cut (frame 0) (d) Demichelis Spark (frame 0)

Fig. 5. 3D holoscopic images used in the proposed test set.

By using this approach, there is no need for further signalling and the HEVC bitstream structure is not modified. This simplifies the inclusion of this novel prediction method into HEVC framework. Hence, the prediction samples generated by the selected directional modes are replaced by the output produced by the LLE-based mode using k-NN with $k = 1, \ldots, 8$, as proposed in [5]. Notice that k-NN templates and linear coefficients are derived in the encoder, as well as in the decoder, as the reconstructed samples of the causal search window are available in both sides. To include enough samples to exploit the redundant information across several micro-images for most 3D holoscopic images, the following dimensions were defined for the search window: $D_{V_1} = 128$, $D_{V_2} = P_Y + T$, $D_{H_1} = 128$ and $D_{H_2} = 64 + P_X + T$ (see Fig. 3). The template thickness was set to $T = 4$.

Finally, the optimal block partition and prediction modes are chosen using the usual HEVC rate-distortion optimization (RDO) procedure. The predictive coding of the mode flags from the neighbour blocks has been kept from the HEVC algorithm, although it had been originally optimized for the directional prediction modes. The proposed mode is also available for chrominance prediction.

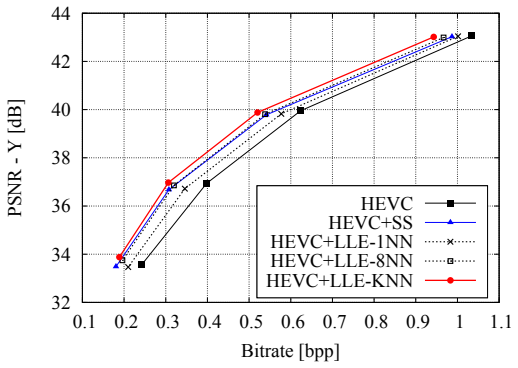
5. EXPERIMENTAL RESULTS

The performance of the proposed LLE-based prediction framework for 3D holoscopic image coding in HEVC (referred to as *HEVC+LLE-KNN*) was evaluated against the original HEVC standard (referred to as *HEVC*) as well as against the proposed SS compensated prediction in [2] (referred to as *HEVC+SS*). The reference software version HM-13.0 was used as the benchmark and also as the basis to implement the proposed coding scheme. Moreover, two additional tests were performed using particular cases of the LLE-based prediction in which the k-NN method uses a fixed k: *HEVC+LLE-8NN* refers to HEVC enhanced with LLE-based mode using 8-NN; while *HEVC+LLE-1NN* uses only 1-NN (which is equivalent to the template matching algorithm). For these additional experiments only one intra directional mode was replaced.

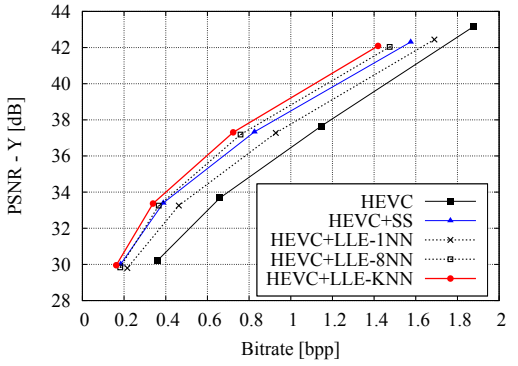
Four 3D holoscopic test images were used for these tests, as shown in Fig. 5: *Plane and Toy* (1920×1088), *Demichelis Cut* (2880×1620), *Demichelis Spark* (2880×1620) and *Laura* (7240×5432). This test set was encoded according to the common test conditions defined in [11], using the "Intra, main" configuration, and using four QPs (22, 27, 32 and 37).

As can be seen in Fig. 6, the proposed coding scheme for 3D holoscopic images (*HEVC+LLE-KNN*) is clearly more advantageous, outperforming all the other tested scenarios. The Bjontegaard Delta (BD) PSNR (and BD-rate) results of the proposed approach over *HEVC* are, respectively: 1.2 dB (-18.5%) for *Plane and Toy*, 3 dB (-37.9%) for *Laura*, 1.4 dB (-37.4%) for *Demichelis Cut* and 1.4 dB (-38.5%) for *Demichelis Spark*. It can be also observed that, in spite of the gains of *HEVC+SS* over *HEVC*, the proposed *HEVC+LLE-KNN* is able to consistently achieve an advantage over *HEVC+SS* presenting BD-PSNR (and BD-rate) results of 0.28 dB (-4.9%) for *Plane and Toy*, 0.67 dB (-11.3%) for *Laura*, 0.52 dB (-15.3%) for Demichelis Cut, and 0.43 dB (-13.5%) for Demichelis Spark. Regarding computational complexity, the template matching step is expensive. It increases encoding times by one or two orders of magnitude when compared to to HEVC.

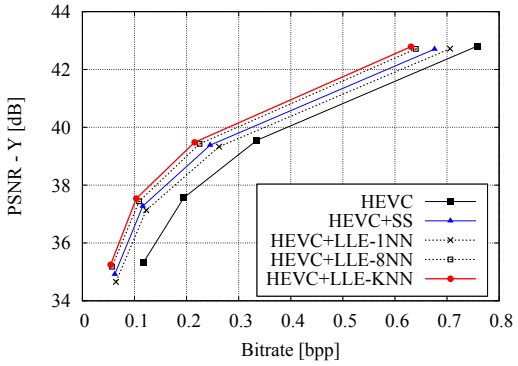
When only one mode is replaced and a fixed number of NN is considered, significant gains are still achieved over *HEVC*. However, in the case of *HEVC+LLE-1NN*, or template matching, these gains are inferior to the other approaches, as well as the explicit block matching algorithm implemented by *HEVC+SS*. In contrast, *HEVC+LLE-8NN*, based on 8-NN templates, presents similar results to *HEVC+SS* for Plane and Toy, and provides some coding gains for the other images. The optimal method is the main proposal (*HEVC+LLE-KNN*) in which eight directional modes are replaced by LLE-based mode using a different number of NN templates.



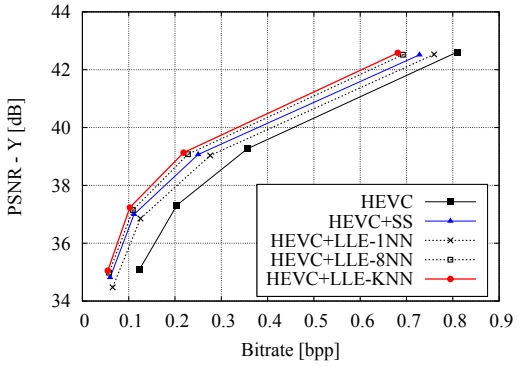
(a) Plane and Toy (frame 0)



(b) Laura



(c) Demichelis Cut (frame 0)



(d) Demichelis Spark (frame 0)

Fig. 6. Rate-distortion results for the proposed test set.

6. FINAL REMARKS

This paper proposes a novel coding scheme for 3D holoscopic image coding based on locally linear embedding (LLE) method for intra prediction. Experimental results demonstrate that the proposed scheme presents rate-distortion gains over HEVC and self-similarity (SS) method, by exploiting spatial redundancies presented in this content. Future work will include the research on an improved prediction framework, by combining the proposed approach with the explicit block matching algorithm provided by the SS method.

REFERENCES

[1] A. Aggoun, E. Tsekleves, M.R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L.D. Soares, "Immersive 3D holoscopic video system," *MultiMedia, IEEE*, vol. 20, no. 1, pp. 28–37, Jan 2013.

[2] C. Conti, P. Nunes, and L.D. Soares, "New HEVC prediction modes for 3D holoscopic video coding," *ICIP*, pp. 1325–1328, Sept 2012.

[3] ITU-T and ISO/IEC JTC 1/SC 29 (MPEG), *High efficiency video coding*, Recommendation ITU-T H.265 and ISO/IEC 23008-2, 2013.

[4] A. Agooun, O.A. Fatah, J.C. Fernandez, C. Conti, P. Nunes, and L. Ducla Soares, "Acquisition, processing and coding of 3D holoscopic content for immersive video systems," *3DTV-CON*, pp. 1–4, Oct 2013.

[5] M. Turkan and C. Guillemot, "Image prediction based on neighbor-embedding methods," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1885–1898, 2012.

[6] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *Journal de Physique Thorique et Applique*, vol. 7, no. 1, pp. 821–825, Nov 1908.

[7] E. Elharar, A. Stern, O. Hadar, and B. Javidi, "A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform," *Display Technology, Journal of*, vol. 3, no. 3, pp. 321–325, Sept 2007.

[8] A. Aggoun, "Compression of 3D integral images using 3D wavelet transform," *Display Technology, Journal of*, vol. 7, no. 11, pp. 586–592, Nov 2011.

[9] R. Olsson, "Empirical rate-distortion analysis of JPEG 2000 3D and H.264/AVC coded integral imaging based 3D-images," *3DTV-CON*, pp. 113–116, May 2008.

[10] Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *SCIENCE*, vol. 290, pp. 2323–2326, 2000.

[11] Frank Bossen, "Common HM test conditions and software reference configurations," *Document JCTVC-L1100*, 2013.