Multidim Syst Sign Process
DOI 10.1007/s11045-016-0417-0

Recurrent pattern matching based stereo image coding using linear predictors

Luís F. R. Lucas[1,2] · **Nuno M. M. Rodrigues**[2,3] · **Carla L. Pagliari**[4] ·
Eduardo A. B. da Silva[1] · **Sérgio M. M. de Faria**[2,3]

Received: 29 July 2015 / Revised: 5 January 2016 / Accepted: 21 April 2016

Abstract The Multidimensional Multiscale Parser (MMP) is a pattern-matching-based generic image encoding solution which has been investigated earlier for the compression of stereo images with successful results. While first MMP-based proposals for stereo image coding employed dictionary-based techniques for disparity compensation, posterior developments have demonstrated the advantage of using predictive methods. In this paper, we focus on recent investigations on the use of predictive methods in the MMP algorithm and propose a new prediction framework for efficient stereo image coding. This framework comprises an advanced intra directional prediction model and a new linear predictive scheme for efficient disparity compensation. The linear prediction model is the main novelty of this work, combining adaptive linear models estimated by least-squares algorithm with fixed linear models provided by the block-matching algorithm. The performance of the proposed intra prediction and disparity compensation methods when applied in an MMP encoder has been evaluated experimentally. Comparisons with the current stereo image coding standards showed that the proposed MMP algorithm significantly outperforms the Stereo High Profile of H.264/AVC standard. In addition, it presents a competitive performance relative to the MV-HEVC standard. These results also suggest that current stereo image coding standards may benefit from the proposed linear prediction scheme for disparity compensation, as an extension to the omnipresent block-matching solution.

This project was funded by FCT—"Fundação para a Ciência e Tecnologia", Portugal, under the Grant SFRH/BD/79553/2011. This work was partially financed by CAPES/Pro-Defesa under Grant Number 23038.009094/2013-83.

✉ Luís F. R. Lucas
luis.lucas@smt.ufrj.br

[1] PEE/COPPE/DEL/Poli, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-972, Brazil

[2] Instituto de Telecomunicações, 2411-901 Leiria, Portugal

[3] ESTG - Instituto Politécnico de Leiria, 2411-901 Leiria, Portugal

[4] DEE, Instituto Militar de Engenharia, Rio de Janeiro 22290-270, Brazil

Keywords Stereo image coding · Pattern matching · Disparity compensation · Linear prediction · Least-squares prediction · Block matching algorithm

1 Introduction

The increasing demand for 3D multimedia contents in the last years has motivated the development and proliferation of new 3D systems and applications. The most common technology is based on the stereo-view video format (Vetro 2010), which is traditionally represented by a pair of texture views. The multiview video format is a more advanced 3D technology which has received an increased attention, to a large extent due to the recent developments in autostereoscopic displays. Despite the better viewing experience provided by multiview systems, they present some drawbacks mostly related with the large amount of information generated by the multiple views. Geometry-based multiview representation approaches, that transmit fewer texture video views and the associated depth maps have been proposed (Muller et al. 2011). The underlying idea is to generate the missing texture views in the decoder side by means of a synthesis procedure that uses depth map information.

The trivial approach for 3D video coding, known as simulcast, compresses independently each texture (and depth map) view using existing still image and single view video coding standards. Due to its simplicity, this scheme requires the smallest computational effort. However, it does not exploit the existing redundancies between views, which limits its compression performance. In order to improve the coding efficiency of these signals, more advanced coding solutions have been proposed to exploit the correlations between texture (or depth map) views, namely through the use of disparity compensated prediction techniques (Merkle et al. 2007; Chen et al. 2008; Mueller et al. 2013).

Disparity compensation (DC) is an important tool in the design of modern stereo and multiview image and video coding algorithms. Block-based disparity compensation approaches have been early adopted (Dinstein et al. 1988; Perkins 1992; Woo and Ortega 1999), mainly due to the block-based structure of most image and video encoders. This is also the case of the current state-of-the-art 3D extensions of video coding standards, such as the MVC (Multiview Video Coding) extension of H.264/AVC (Chen et al. 2008; Merkle et al. 2007; ITU-T and ISO/IEC JTC1 2010), and the MV-HEVC and 3D-HEVC standards (Multiview and 3D extensions of the High Efficiency Video Coding) (Tech et al. 2013; Stankowski et al. 2012; Mueller et al. 2013; ITU-T and ISO/IEC JTC 1/SC 29 (MPEG) 2013). The main purpose of the 3D extensions is to enable inter-view prediction to increase the coding performance and better compress depth maps in the video plus depth format. Disparity compensation is performed similarly to temporal motion-compensation, which uses the block-matching (BM) algorithm (Kaup and Fecker 2006). The main difference between them is related to the reference frames used in the estimation process, that belong to multiple views at the same time instances.

Block-matching DC schemes can use both fixed or variable block sizes. When using fixed block size, only the information about the estimated disparity is transmitted for each block. On the other hand, the use of variable block sizes provides an enhanced estimation of the disparity, because blocks containing multiple objects, with distinct depths, can be partitioned into smaller sub-blocks. The drawback of this approach is the additional overhead required to encode the block's dimension and position. The quadtree block partitioning scheme is a common approach which only uses square-sized blocks. In Sethuraman et al. (1995) and Accame et al. (1995), quadtree segmentation was combined with a multiresolution decomposition approach in order to reduce false block matches as well as the computational load. Another successful solution, which is adopted in the current video coding standards, H.264/AVC and HEVC (ITU-T and ISO/IEC JTC1 2010; ITU-T and ISO/IEC JTC 1/SC 29 (MPEG) 2013), considers a larger set of possible block sizes, including rectangular dimensions.

Despite the success of the block-matching disparity compensation schemes, they may present some issues under certain circumstances. For instance, they often fail to compensate mismatched areas given by occlusions or deformations, such as perspective distortions. In order to tackle these issues, some proposals distinguish the occlusion areas in the disparity compensated residue so that they can be independently encoded (Frajka and Zeger 2002). Others use a Markov Random Field (MRF) model to compute a smoother disparity map (Woo and Ortega 1996; Ellinas and Sangriotis 2006) or an overlapped block DC to reduce the blocking effect in the disparity map (W. Woo and Ortega 2000). A method to better predict the mismatching effects was proposed in Seo et al. (2000), using a per-block least-squares-based 2D filtering over the reference image. Its drawback is the large bitrate required to transmit the estimated filter coefficients for each block. Other stereo image coding paradigms can be found in literature, such as the one in Palaz et al. (2011), which proposes a joint sparse approximation framework based on a dictionary of geometric functions learned on a database of stereo images. However, this method does not have a competitive rate-distortion performance when compared to the state-of-the-art MV-HEVC standard.

In this paper, we present an alternative algorithm to encode 3D images, specifically stereo pairs. The proposed approach uses an efficient intra prediction scheme and a novel DC framework, which is an extension to our previous developed work on linear prediction (Lucas et al. 2011a). The linear predictors can be implicitly derived using a least-squares-based approach, or explicitly signalled based on BM algorithms. This solution generalizes the proposed implicit and explicit DC methods in a new framework that is more efficient, due to its ability to exploit several linear predictors with different characteristics. As for residue coding, a pattern-matching-based algorithm, known as Multidimensional Multiscale Parser (MMP) is used.

The MMP algorithm has been successfully presented in literature for still image coding (Carvalho et al. 2002; Rodrigues et al. 2008; Francisco et al. 2010), as well as for stereo image coding (Duarte et al. 2005; Lucas et al. 2011a, b). In a stereo image coding scenario, the MMP intra coding techniques are commonly used to independently encode the reference image (usually assigned to the left image), while DC methods that exploit the redundancy between views are employed to encode the right image of the stereo pair. Previous MMP-based proposals for stereo image coding used dictionary-design techniques (Duarte et al. 2005). However, predictive-based methods (Lucas et al. 2011a, b) have achieved a superior rate-distortion performance in recent years. This fact motivates further development and improvement of existing predictive-based DC methods in MMP algorithm for an efficient compression of stereo images.

Experimental results show that the proposed MMP-based stereo image encoder present a state-of-the-art rate-distortion performance, competitive to the one of transform-based MV-HEVC standard. These results also demonstrate the superiority of proposed algorithm over previous MMP-based approaches for stereo image coding. It is important to notice that, although the proposed DC scheme has been evaluated in the context of the MMP algorithm, it can be applied to other stereo image encoders based either on pattern-matching or transform-coding paradigms.

This paper is organized as follows: Sect. 2 briefly reviews the original MMP algorithm for intra image coding, as well as the existing stereo image coding extensions of MMP. The proposed improvements to the original intra prediction methods of MMP algorithm are presented in Sect. 3, while the novel proposed linear prediction scheme for DC is described in Sect. 4. Section 5 presents a discussion of the experimental results, and Sect. 6 concludes the paper.

2 Multidimensional Multiscale Parser—MMP

In this section the original intra-based MMP algorithm for still image coding (Rodrigues et al. 2008; Francisco et al. 2010) is briefly described. Furthermore, the existing stereo variants of MMP algorithm, based on dictionary-design (Duarte et al. 2005) and linear prediction methods are presented (Lucas et al. 2011a, b).

2.1 The MMP-intra algorithm for image compression

The MMP algorithm for intra image coding combines dictionary-based coding with an efficient intra prediction framework. The main idea of MMP is to approximate the prediction residue by using elements from a dictionary that uses multiple scales. By reusing the previously encoded patterns of the image, MMP is able to learn image features and better encode redundant information.

MMP starts by dividing the input image into 16×16 non-overlapping blocks which are sequentially encoded. Each block may be recursively partitioned according to a flexible segmentation rule (Francisco et al. 2008). Each partitioning occurs in either the vertical or horizontal direction producing two equally sized sub-blocks. By applying this rule down to 1×1 block size, a total of 25 scales are defined by all the possible combinations: $2^m \times 2^n$, for $m, n = 0, \ldots, 4$.

Predictive coding was introduced in the MMP encoder in order to improve its encoding performance, especially for smooth images (Rodrigues et al. 2008). MMP uses a hierarchical prediction framework based on ten prediction modes, that are tested on each sub-block. These modes include the MFV (most frequent value) (Rodrigues et al. 2005), eight directional modes inspired on the ones used in H.264/AVC encoder and the intra LSP (least-squares prediction) mode (Graziosi et al. 2009). The prediction step is applied on sub-blocks obtained through the MMP flexible partitioning scheme, having dimensions from 16×16 to 4×4.

The residue generated by each prediction mode is optimized using the MMP paradigm. In this process, a given residue patch is recursively segmented down to 1×1 scale and each block is approximated according to a rate-distortion cost $J = D + \lambda R$, where λ is a *Lagrangian multiplier*, R is the rate required for the representation and D is the squared-distortion (SSE) generated by that approximation. Then, in order to generate the optimal segmentation tree, the Lagrangian cost of each segmentation option is evaluated at each node of the fully expanded tree, by scanning from the bottom to the top. Whenever the cost of the parent block is inferior to the sum of the costs of the child sub-blocks, the associated tree node is pruned.

The optimal MMP block segmentation is represented by a binary segmentation tree, that contains all the information required to generate the block approximation. The corresponding bitstream is constructed by scanning the tree from top to bottom, coding all nodes and leaves, using a context adaptive arithmetic coder. Two possible flags can signal each node, depending on whether the segmentation is horizontal or vertical. The leaves are signalled using a specific flag, that is followed by the index of the dictionary pattern that approximates the block on that leaf. The decoder is able to replicate the coding decisions and thus to use the same dictionary, which is updated using the same process as performed in the encoder, without requiring any side information.

The adaptation of the MMP dictionary is a key factor to its coding performance. MMP learns the image features by incorporating previously encoded patterns into an adaptive dictionary. During the encoding and decoding processes, the patterns used to represent each block are concatenated and added to the dictionary. Since it uses multiscale patterns, MMP organizes the dictionary according to the elements' scales. When an element is added to the dictionary, expanded and contracted versions of that element are computed and inserted into the dictionary at corresponding scales. This procedure ensures that the new block will be available to encode future blocks, irrespective of their dimensions. As the MMP dictionary is updated after each coding block, it rapidly learns the image's features.

2.2 Stereo image coding using MMP

The first MMP-based proposal for stereo image coding was presented in Duarte et al. (2005). This approach uses dictionary design methods to improve the MMP performance for the compression of the dependent image (typically the right image).

The algorithm encodes a row of blocks (ROB) at each time, alternating between the reference image and the dependent image of the stereo pair. When encoding the right image, besides the MMP standard dictionary, the algorithm uses an additional dictionary that comprises the codewords obtained by sliding a window over the coded ROB of the reference left image. These codewords are referred to as displaced elements. The use of a dictionary of displaced patterns is comparable with the the block-matching disparity estimation, since the disparity compensated blocks are available in the dictionary.

Despite its interesting methodology, the MMP algorithm presented in Duarte et al. (2005) shows a rate-distortion performance well below the one of current standards, such as MVC or MV-HEVC. This can be partially justified by the fact that the MMP version described in Duarte et al. (2005) did not include intra prediction methods, presenting a lower performance for the compression of both the left (reference) and right (dependent) images. Furthermore, the dictionary-design-based DC techniques did not provide the same performance as predictive techniques for the compression of the dependent image.

The use of predictive methods for DC in MMP algorithm has been investigated using the template-matching (TM) algorithm (Lucas et al. 2011b) and using the least-squares prediction (LSP) to linearly predict the disparity (Lucas et al. 2011a). It was shown that LSP is able to provide more complex disparity representations, by linearly combining several samples from the left and right images. The use of causal samples of the right image is appropriated in the presence of disoccluded samples, which are only visible in right image. Linear prediction additionally allows to compensate luminance variations, usually caused by miss-calibration or cameras receiving light from different directions.

The principle of LSP algorithm (Li and Orchard 2001) for generic image prediction is to filter a set of samples, belonging to the causal reconstructed neighbourhood of the current block. In LSP, the prediction $\hat{X}(\mathbf{n})$ of sample $X(\mathbf{n})$ is given by:

$$\hat{X}(\mathbf{n}) = \sum_{i=1}^{N} a_i X(\mathbf{n} - \mathbf{s}(i)), \tag{1}$$

where $\mathbf{n}$ is the position to be predicted using N neighbouring samples at positions $\mathbf{n} - \mathbf{s}(i)$, the $\mathbf{s}(i)$ gives the relative positions of the filter support in the causal data, and a_i are the filter coefficients. In order to avoid the transmission of the filter coefficients, they are locally optimized in a least-squares sense, using a causal reconstructed region of the image, denominated *training window* (TW). The performance of LSP methods relies on the assumption that the causal TW is somewhat correlated with the unknown block, and the estimated filter coefficients in the causal TW provide reasonable prediction results for the unknown block for most cases.

In order to use the LSP method for stereo image prediction, the linear filter support should be designed in such a way that the most correlated reconstructed samples from the left image are used for linear prediction. In Lucas et al. (2011a), the portion of the filter support belonging to the left image is positioned based on the average disparity of the region, estimated by the TM algorithm. The TM algorithm has been also investigated for DC as an independent method in MMP algorithm (Lucas et al. 2011b). TM principle is similar to BM algorithm, predicting the unknown block using a displaced block of the left image. The main difference is that TM implicitly derives the disparity vector of the block, by using the block's template, commonly formed in the block neighbourhood. This procedure avoids the transmission of the estimated vector to the decoder, since the disparity estimation procedure using the neighbouring template can be replicated at the decoder side.

The template area used by TM algorithm in Lucas et al. (2011a) corresponds to the samples of the TW used by LSP algorithm (i.e. the neighbouring samples to the left and above the block to be predicted). This template area allows to find the most correlated samples between the TW of LSP and the samples of the left image belonging to the filter support. The displacement vector returned by the TM algorithm is used to position the filter support in the left image, so that the filtered samples are the most correlated ones with the training samples present in the TW. In this way, the LSP algorithm is able to better exploit the similarities between the left and right images. For an improved adaptation of LSP for stereo image coding the algorithm in Lucas et al. (2011a) also proposes the use of a varied set of filter supports. The idea is to test several filter support configurations and choose the one that generates the best prediction for the target block. The chosen support is explicitly signalled to the decoder. These filters provide different modelling capabilities which can be adapted to different regions of stereo images. For example, filters that include samples from both left and right images can be advantageous to predict blocks with partially occluded regions.

3 Proposed contributions to the MMP algorithm

Unlike the dictionary-design approach presented in Duarte et al. (2005), in this paper we propose to use efficient predictive methods to exploit the stereo redundancy. Our approach firstly encodes the reference/left image using intra prediction methods and the dictionary approximation paradigm for residue representation. Then, the right image is encoded using the same intra prediction methods plus the proposed linear predictive DC scheme based on the left image. The resulting residue is encoded similarly to the left image, using the MMP paradigm.

Besides the proposed DC scheme, based on linear predictors, we propose an improved intra prediction framework for the MMP encoder. Some of the features of the MMP algorithm, like the block segmentation, were also revised and improved. The techniques and improvements, which resulted in an increased rate-distortion performance for both the left and right views, are described in this section.

3.1 Initial block size and flexible segmentation

MMP has long been using the 16 × 16 initial block size, as the H.264/AVC standard. However, with the advent of high resolution formats, larger block sizes may be beneficial for efficient coding. Since in this work MMP is used for the compression of high resolution stereo pairs, the MMP algorithm was adapted for the initial block size of 64 × 64. Keeping the MMP flexible block partitioning rule, the new initial block size would enable 49 possible block sizes. Nevertheless, a different block segmentation scheme was studied in this work. Empirical observations demonstrated that larger block sizes improve MMP rate-distortion performance mainly at lower bitrates, while smaller block sizes continue to be frequently used at higher bitrates or for image regions with more complex structures or textures. In order to avoid the transmission of a large number of segmentation flags needed to reach smaller block sizes, a bi-level approach for the initial block size is proposed. This approach enables the partitioning of the initial 64 × 64 block size into 16 square blocks with 16 × 16 size, which are independently optimized and encoded. MMP decides the best initial block level by optimizing both the coding of the 64 × 64 block size and the 16 smaller 16 × 16 blocks. One flag is transmitted to indicate the best solution.

Figure 1 illustrates the proposed segmentation trees for the two possible initial block sizes in MMP. Gray blocks indicate the MMP scales where intra prediction methods can be used. While the 16 × 16 initial block size uses the same segmentation tree as the original MMP algorithm, a pruned tree was defined for the 64 × 64 size. The pruning for the initial block size of 64 × 64 was applied on smaller block sizes. This is motivated by the fact that small block sizes are mainly used for the cases of initial block size 16 × 16.

3.2 Improved intra prediction framework

Based on the recent developments of H.265/HEVC, we also improved the intra-prediction framework of the MMP algorithm. These improvements were highly motivated by the availability of larger block sizes, for which new directional correlations may be exploited. Thus, the 8 directional modes used in the original MMP algorithm were replaced by 33 directional modes, similar to the ones used in HEVC (ITU-T and ISO/IEC JTC 1/SC 29 (MPEG) 2013), illustrated in Fig. 2. Besides the directional modes, MMP uses the planar, DC and an intra-based LSP mode (Graziosi et al. 2009), totalling 36 intra modes. By using a more sophisticated prediction framework, MMP residue probability distribution tends to be narrower, leading to better encoding results. As demonstrated in Rodrigues et al. (2008), this happens because more uniform elements tend to be used at larger sizes favouring the adaptation of the dictionary and its efficiency in representing the encoded blocks.

The proposed intra prediction framework uses 36 modes which should be efficiently transmitted to the decoder. This can be done by exploiting correlations that may occur between adjacent blocks of the image, namely when the same directional modes are used across several blocks. In this work, we propose a prediction process for the intra prediction modes in the MMP algorithm, in order to improve their coding performance.

The proposed scheme starts by deriving three candidate prediction modes from the block neighbourhood, using an implicit approach that can be executed in both the encoder and the decoder. When some candidate matches the chosen prediction mode, an index is transmitted, otherwise the prediction mode value is fully encoded. The candidate modes are based on the causal neighbour samples adjacent to the current block. Since these neighbouring samples may correspond to several prediction modes, the method chooses the three most frequent prediction modes among the adjacent neighbouring samples. If the number of available modes in the neighbourhood is inferior to 3, the existing modes are used as candidates. To encode the mode into the bitstream, a binary flag is first transmitted indicating whether the

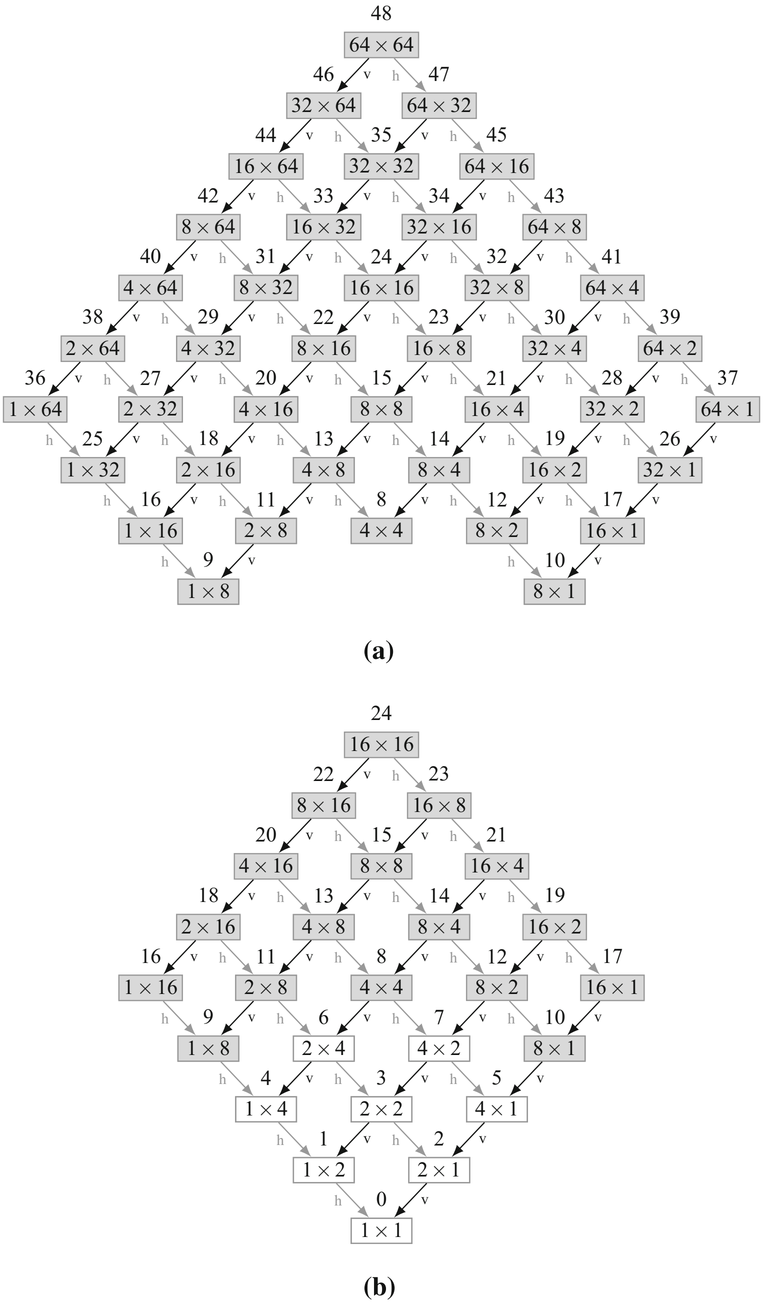

Fig. 1 Block segmentation tree for the bi-level initial block size: **a** 64 × 64 and **b** 16 × 16

prediction mode matches one of the candidates. Depending on the value of this flag, the next transmitted symbol is either the candidate index (with three possible values) or the actual prediction mode value (36 modes minus the candidate modes).

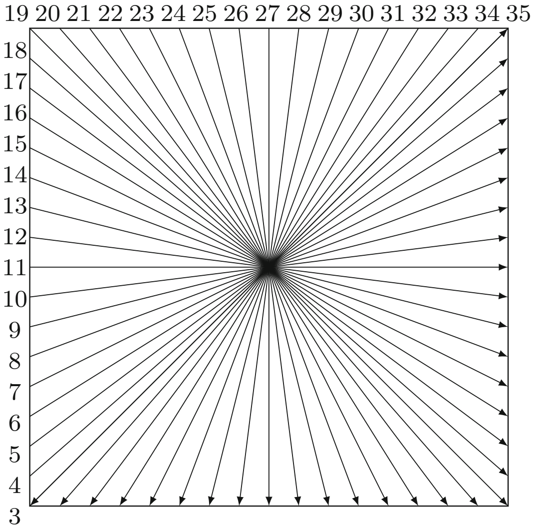

Fig. 2 Set of prediction directions of the 33 directional modes (3–35) for the Intra Prediction in MMP algorithm. The modes 0, 1, and 2, not shown in the figure, refer to the DC, planar and LSP modes, respectively

4 Disparity compensation framework

Previous research works on predictive coding methods using the MMP algorithm have demonstrated their importance for efficient still and stereo image coding (Rodrigues et al. 2005; Lucas et al. 2010, 2011a, b). As discussed in Sect. 2, LSP presents a high degree of adaptation providing successful results for disparity estimation. Unlike the traditional BM-based disparity estimation, that explicitly transmits the average disparity value of the block, LSP is able to adaptively learn more complex linear representations of the block disparity, using linear combinations of the left image samples. This approach may be advantageous when the block disparity is not simply given by an uniform displacement of the whole block samples, e.g. blocks that present perspective distortions or partially occluded blocks. The implicit estimation of linear predictors through training procedure may provide efficient prediction results with low bitrate overhead. However, explicit transmission of the block disparity may be advantageous when the causal information is not correlated with the unknown block to be predicted.

In this context, we propose a new linear predictive method for DC which combines both explicit and implicit DC approaches in order to generate more efficient prediction results and provide state-of-the-art rate-distortion performance. Although we combine two different methodologies, based on LSP and BM algorithms, we show that these methods can be interpreted as particular cases of linear prediction with distinct characteristics. While LSP provides linear prediction using adaptively estimated coefficients for each block, the BM algorithm can be viewed as a linear prediction method that uses fixed coefficients associated with sub-sample interpolation. Another difference between the two methods is related with the positioning of the filter support in the left image. These characteristics make LSP and BM algorithms function as complementary linear predictors, being able to benefit from each other when used in the same framework. The proposed generalized DC framework is described by the following linear predictor:

$$\hat{I}_R(\mathbf{n}) = \sum_{i=1}^{N} a_i \tilde{I}_L\left(\mathbf{n} - \mathbf{g}_s^L(i) - \mathbf{d}\right) + \sum_{j=1}^{M} b_j \tilde{I}_R\left(\mathbf{n} - \mathbf{g}_s^R(j)\right), \quad N > 0, M \geq 0 \qquad (2)$$

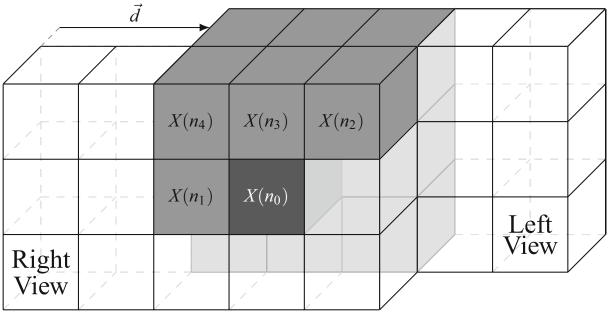

Fig. 3 Tri-dimensional representation of the filter support

given the linear filter support shape s and displaced by vector $\mathbf{d}$ in the left image, with $\hat{I}_R(\mathbf{n})$ being the right image predicted sample at position $\mathbf{n}$. The two summations represent two types of linear combinations that use reconstructed samples from the left and right images, based on the filter supports $\mathbf{g}_s^L$ and $\mathbf{g}_s^R$, and linear coefficients a_i and b_i, respectively. The optimal filter support shape s and displacement vector $\mathbf{d}$ should be determined. Note that the first summation with reconstructed samples from the left image always exists ($N > 0$). However, the weighted sum of right image samples can be null, specifically $M = 0$. This is so because the spatial neighbouring samples (of the right image) are only used by some linear predictors. An example illustrating the proposed linear model, with a filter support formed by four samples in the right image and nine samples in the left image is presented in Fig. 3.

In the proposed framework, the filter support shapes associated with both left ($\mathbf{g}_s^L$) and right ($\mathbf{g}_s^R$) images are chosen from a previously defined set with various possibilities. The coefficient values of the filter support can be estimated in a training window by LSP or given by the quarter sample interpolation filter of BM algorithm. Note that the proposed generalized formulation for LSP and BM algorithms allows a better integration of these methods, namely for predicting common information between neighbouring blocks. As we will discuss, our method predicts the displacement vector $\mathbf{d}$ across blocks that use either adaptive or fixed linear predictors.

The adaptive and fixed linear predictors are evaluated in the rate-distortion loop of MMP, along with the intra prediction modes for all available block sizes superior or equal to 4 × 4. The best prediction mode is selected, according to a Lagrangian rate-distortion criterion, by minimizing the weighted sum of the residue energy and the bitrate used to encode the mode and residue data. We present a detailed description of the proposed LSP and BM-based predictors in the sequel.

4.1 Adaptive linear predictors using LSP

The proposed adaptive linear predictors are based on the LSP methods previously presented in Lucas et al. (2011a). Here we describe the improvements made to them, and we also investigate the impact of the TW shape in the prediction process. In this approach, linear coefficients are adaptively estimated in a causal TW defined in the block neighbourhood, using the LSP algorithm (Li and Orchard 2001). The proposed LSP method may use several filter support shapes, similarly to our previous proposal in Lucas et al. (2011a). The available filter supports are illustrated in Fig. 4, where the filled black circle represents the unknown sample to predict and the empty circles represent the filter support positions. The left view row of Fig. 4 represents filter support positions located in the reference left view, properly

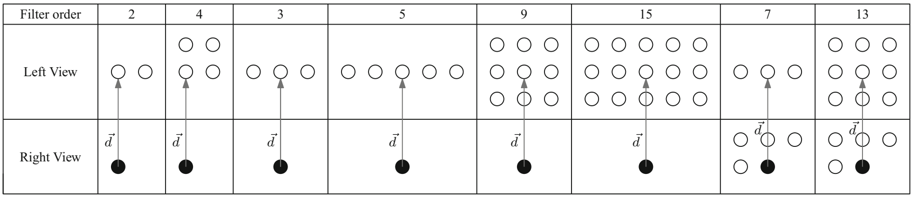

Fig. 4 Filter context configurations used in the proposed LSP algorithm for stereo image coding (Lucas et al. 2011a)

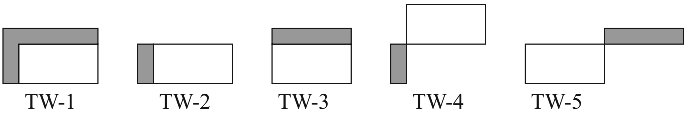

Fig. 5 Example of proposed training windows used in LSP algorithm

displaced by **d** from the co-located position of unknown sample. Regarding the spatial filter support positions, represented in the right view row of Fig. 4, only the last two filters include some neighbouring spatial positions. These filter supports are intended to provide an efficient representation of disparity information based on the left image. By exploiting the spatial filter positions, the filters of 7th and 13th orders may provide a more advanced representation, addressing not only disparity redundancy, but also occlusion areas and luminance variations.

In the proposed method we optimize the size of the TW used by LSP for the new variety of block sizes available in the MMP algorithm. It is a fact that the larger blocks tend to be used when disparity is constant or varies smoothly, and the smaller blocks are used in the presence of large disparity variations. Therefore, we propose an adaptive size for TW, which tries to increase the similarity between the TW and the unknown block. The proposed solution consists in using smaller TW sizes for smaller unknown blocks, so that the TW for these blocks includes less causal samples that may be unrelated with the unknown block (due to the highly changing characteristics of these regions). We define the TW size based on its thickness, expressed by $T_{th} = \min((B_w + B_h)/4, 4)$, where B_w and B_h match the block's width and height, respectively. This solution returns a TW size proportional to the unknown block size, until a minimum of $T_{th} = 4$. The minimum thickness limit keeps the TW size within a reasonable value for which stationary statistics within TW can be assumed.

We also investigated an improved LSP algorithm with better learning capabilities, by testing various TWs to estimate the linear coefficients. The idea of this approach is to select the TW that generates the best set of linear coefficients to predict the unknown block. Thus, in addition to the TW proposed in Lucas et al. (2011a), we propose four additional neighbouring TWs with different shapes. Figure 5 illustrates the selected TW shapes, which exploit the available regions around the unknown block. The TW-1 shape of Fig. 5 corresponds to the original TW proposed in Lucas et al. (2011a). When top and left neighbour regions are uncorrelated, TW-1 can be inefficient and LSP has the possibility to choose a smaller neighbour TW, e.g., only left (TW-2) or only top (TW-3). The inclusion of TW-4 and TW-5 regions is justified by the larger block sizes present in MMP (e.g. 64 × 64), which increase the availability of the top-right and bottom-left neighbourhoods. The optimal TW for predicting the unknown block should be explicitly signalled to the decoder using an index.

Regarding the derivation of the displacement vector **d**, used to position the filter support in the left image, an implicit approach that does not require the signalling of the vector is

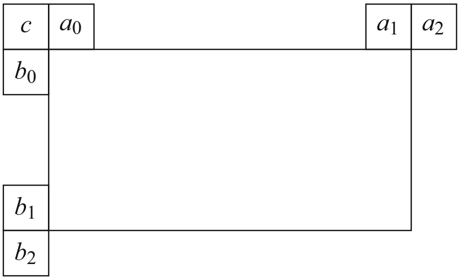

Fig. 6 Spatial candidates for disparity vector prediction in a block typically used in the MMP algorithm

Table 1 Relative importance of spatial candidates for each TW shape, represented as sorted sequences from the most to the least important candidates

TW shape	Candidate importance (+ to −)
1	$c, a_0, b_0, a_1, b_1, a_2, b_2$
2	$b_0, b_1, c, b_2, a_0, a_1, a_2$
3	$a_0, a_1, c, a_2, b_0, b_1, b_2$
4	$b_2, b_1, b_0, c, a_0, a_1, a_2$
5	$a_2, a_1, a_0, c, b_0, b_1, b_2$

used. As the original LSP proposal for stereo image coding in Lucas et al. (2011a), we try to maximize the correlation between the TW of LSP and the filter support samples located in the left image. This is because the training procedure tries to approximate the TW samples by linearly combining the samples of the left image (and right image for a few filter supports). In most cases, the displacement vector corresponds approximately to the average disparity of the TW. Since the TW samples are available in both decoder and encoder sides, we can implicitly derive the displacement vector $\mathbf{d}$ in both sides, without requiring the transmission of additional information. The proposed procedure uses a preliminary estimation of vector $\mathbf{d}$, based on previously encoded blocks, that is later refined using the TM algorithm with a small search window of size 20×4. Note that the larger horizontal dimension is motivated by the fact that disparity mainly varies in that direction, being even exactly horizontal for stereo pairs obtained by parallel camera arrangements.

In order to derive the preliminary estimation of the displacement vector, the seven spatial candidates of Fig. 6, placed around the unknown block, are considered. The algorithm selects one spatial candidate and uses the displacement vector associated with the encoded block that contains the selected candidate. The choice of the spatial candidate depends on the level of importance attributed to the spatial candidates, as well as their availability. A candidate is defined as available, if the block which it belongs to was predicted using the proposed linear predictive method. Spatial candidates associated to intra-coded blocks are marked as unavailable. The relative importance of each spatial candidate is defined according to the TW shape being used. Table 1 defines the candidate importances, depending on the TW shape, by listing the proposed candidates from the most to the least important. Note that the proposed method always prefers the candidates closest to the region of the corresponding TW shape. The first available spatial candidate in the presented strings of Table 1 is used, according to the current TW shape. When no candidate is available, the null vector is assumed.

The position pointed by the displacement vector, $\mathbf{d_e}$, estimated from neighbouring blocks plus TM algorithm tends to be the position where LSP performs more efficiently, mainly due to the high correlation between TW and left image samples on that position. However, the neighbouring positions can be equally efficient or even better. In the proposed algorithm, not only the position pointed by the estimated displacement vector $\mathbf{d_e}$ is evaluated for linear prediction, but also the eight positions that are defined around that vector, in an 8-connected neighbourhood. Therefore, the proposed adaptive linear predictors are evaluated at nine possible positions in the left image, for the displacement vectors given by $\mathbf{d} = \mathbf{d_e} + \mathbf{u}$, where $\mathbf{u} = (i, j)$, with $i, j \in \{-1, 0, 1\}$. Each filter support shape of Fig. 4 is tested for the nine possible positions in the left image, and the one that generates the lowest modelling error in the TW is selected (vector $\mathbf{d} = \mathbf{d_e} + \mathbf{u_s}$). It is important to note that this is an implicit selection, i.e. it does not require the transmission of the chosen displacement vector $\mathbf{d}$.

Algorithm 1: Prediction algorithm based on adaptive linear predictors (encoder).

```
Input: causal reconstructed image
Output: $t_s$, $f_s$, predicted block $\hat{P}$
for each TW shape $t$ do
    derive approximation for vector $\mathbf{d}$ from neighbouring blocks;
    obtain enhanced vector $\mathbf{d_e}$ based on a search procedure using TW shape $t$;
    for each filter support $f$ do
        estimate 9 linear models within TW shape $t$ using filter support $f$ displaced by different vectors
        given by $\mathbf{d} = \mathbf{d_e} + \mathbf{u}$, where $\mathbf{u} = (i, j)$, with $(i, j) \in \{-1, 0, 1\}$;
        select linear model that produced the lowest training error (among the 9 possibilities) and save
        associated vector $\mathbf{d} = \mathbf{d_e} + \mathbf{u_s}$;
        compute predicted block $\hat{P}$ using selected linear model based on filter support $f$ and vector
        $\mathbf{d} = \mathbf{d_e} + \mathbf{u_s}$;
        compute prediction error $E(t, f)$;
    end
end
find optimal TW shape and filter support by: $t_s, f_s = \min_{t,f} E(t, f)$;
save $t_s$, $f_s$ and predicted block $\hat{P}$;
```

Algorithm 1 summarizes the procedure of the proposed method based on adaptive linear predictors in the encoder side. This algorithm is used after intra prediction and before block matching-based disparity prediction. Thus, the obtained predicted block and chosen parameters are saved (see last step of Algorithm 1), in order to be compared to the remaining prediction methods and select the optimal one in terms of rate-distortion performance. As explained before, this algorithm tests five different TW shapes (Fig. 5) and eight filter supports (Fig. 4), which results in two nested loops. Each possible combination of filter support f and TW shape t is tested and the best one, represented by f_s and t_s is explicitly transmitted. The displacement vector $\mathbf{d}$ is estimated to position the filter support in the reference left image. As shown in Algorithm 1, this process involves an approximate derivation from neighbouring blocks, an enhancement step based on search procedure, resulting in vector $\mathbf{d_e}$ and a final selection from nine possible vectors given by $\mathbf{d} = \mathbf{d_e} + \mathbf{u}$.

In order to be reproduced in the decoder side, the LSP algorithm only requires two signalling flags, indicating the chosen filter support shape, f_s, and the TW shape, t_s. These flags are compressed using adaptive arithmetic coding. Both the values of the filter coefficients and the position of the filter in the left image (vector $\mathbf{d}$) are implicitly derived from causal reconstructed samples as previously explained. The proposed method for the decoder side is described in Algorithm 2.

Algorithm 2: Prediction algorithm based on adaptive linear predictors (decoder).

```
Input: $t_s$, $f_s$, causal reconstructed image
Output: predicted block $\hat{P}$
derive approximation for vector $\mathbf{d}$ from neighbouring blocks;
obtain enhanced vector $\mathbf{d_e}$ based on a search procedure using TW shape $t_s$;
estimate 9 linear models within TW shape $t_s$ using filter support $f_s$ displaced by different vectors given by $\mathbf{d} = \mathbf{d_e} + \mathbf{u}$, where $\mathbf{u} = (i, j)$, with $(i, j) \in \{-1, 0, 1\}$;
select linear model that produced the lowest training error (among the 9 possibilities) and save associated vector $\mathbf{d} = \mathbf{d_e} + \mathbf{u_s}$;
compute predicted block $\hat{P}$ using selected linear model based on filter support $f_s$ and vector $\mathbf{d} = \mathbf{d_e} + \mathbf{u_s}$;
```

4.2 Fixed linear predictors using BM algorithm

The use of adaptive linear predictors trained in a causal window (TW) of the unknown block provides good prediction results for most encoded blocks of the stereo pair. However, in some situations, these predictors may present inefficiencies, namely when the TW samples are fully decorrelated with the unknown block. In these situations, the image features and disparity information learnt by the LSP training procedure may be not be useful for the block prediction. In our method, we include fixed linear predictors with explicit transmission of displacement vector **d**, in order to cope with situations that cannot be well represented using implicit adaptive predictors. The proposed fixed predictors are based on the well-known BM matching algorithm, commonly used in current state-of-the-art video coding standards, e.g. H.264/AVC and H.265/HEVC.

In order to implement the fixed linear predictors, we use the full search BM algorithm with quarter sample interpolation using a search window that varies in the interval between -96 and 96 for horizontal direction and between -16 and 16 for vertical direction. Note that disparity vectors mainly vary in the horizontal direction. In the proposed scheme, we use the same interpolation filters of H.264/AVC standard (ITU-T and ISO/IEC JTC1 2010).

As we will explain, the BM algorithm can be viewed as a particular case of the proposed generalized linear prediction model that uses explicit estimation of the displacement vector **d** and uses fixed predictors. This is an interesting interpretation of BM algorithm that highlights its similarities with the previous presented adaptive linear predictors, supporting its incorporation in the proposed DC framework.

We may describe BM algorithm as a set of linear filter supports, with predefined shapes and coefficient values, that are positioned in the left image using a displacement vector **d**. The BM-based disparity compensation relies on the assumption that disparity is constant for the whole block. Integer precision BM is achieved by using a first order linear filter with coefficient value equal to one, positioned in the left image using the estimated vector **d**. In order to provide fractional DC, proper linear filters that provide half and quarter sample interpolations are commonly used. The sample interpolation is typically based on the neighbouring full samples, depending on the fractional sample position. Thus, associated with each integer sample of the left image, there are 16 possible interpolation filters that generate the fractional sample positions, including a first-order filter support with coefficient equal to one for no interpolation (at full sample position). Figure 7 illustrates the full samples (gray blocks) and 15 fractional sample positions associated to the full sample $X(x, y)$.

Therefore, we can use quarter sample precision BM algorithm as a particular case of the proposed linear prediction framework, with explicit transmission of the displacement vector **d** plus linear filtering based on 16 available fixed filters. Note that, with BM algorithm the

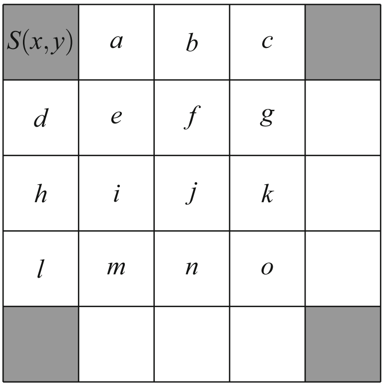

Fig. 7 Full samples (*gray blocks*) and fractional samples (*white blocks*) used for BM-based DC with quarter sample interpolation

meaning of displacement vector **d** is not the average disparity of the TW, as for adaptive linear predictors, but it corresponds approximately to the disparity of the unknown block in integer sample precision.

Regarding to entropy coding, we perform a differential compression of the quarter sample precision disparity vectors obtained by BM algorithm, instead of transmitting two independent symbols (vector **d** and linear filter s). This approach is motivated by the physical meaning of the interpolation filters, which represent a sub-sample displacement. Thus, some correlation may exist for both vector **d** and linear filters s between multiple neighbouring blocks. We propose an efficient coding solution based on the recent Advanced Motion Vector Prediction (AMVP) method used in H.265/HEVC, that considers two spatial disparity candidates selected among seven candidates, according to their availability. The proposed candidates are illustrated in Fig. 6 for a non-square block example typically used in the MMP algorithm. Note that these candidates are obtained from previous encoded blocks using the proposed linear predictors. In the case of blocks encoded by adaptive linear predictors, the displacement vector **d** is used as the candidate disparity vector.

The first candidate is chosen among the disparity vectors of top samples $\{a_0, a_1, a_2\}$ (Fig. 6), according to their availability and presented order. The second candidate is chosen in the same way among the disparity vector of the left samples $\{b_0, b_1, b_2\}$. Candidates corresponding to intra prediction modes are considered as unavailable. The top-left c candidate is used as an alternative when all the top or left candidates are unavailable. At the end, one binary flag signalling the best candidate is transmitted to the decoder, followed by the vector difference between the estimated disparity vector and the chosen candidate. For entropy coding of the differential disparity vectors, the Context-based Adaptive Binary Arithmetic Coding (CABAC) (Marpe et al. 2003) has been used based on the algorithm developed for H.264/AVC standard.

5 Experimental results

Experimental tests were performed in order to evaluate the rate-distortion (RD) performance of the proposed predictive scheme and its application in the framework of the MMP encoder (referred to as *MMP-stereo-proposed*). Some comparisons using MMP with different configurations of adaptive (LSP) and fixed (BM) linear predictors are also presented, in order to demonstrate the effectiveness of the proposed framework for DC. MMP has also been compared with the state-of-the-art H.264/AVC standard using the Stereo High profile and the MV-HEVC standard. The first frame of two of the views of the selected multiview sequences [mainly based on the sequences proposed by MPEG (Muller and Vetro 2014)] were chosen, to form a test-set of stereo pairs, represented in Fig. 8.[1] The first six sequences in Fig. 8 have 1024 × 768 resolution (8a–f). The remaining have resolution of 1920 × 1088 pixels.

5.1 Evaluation of the proposed DC framework

The overall improvements in coding performance provided by the proposed DC framework can be evaluated by comparing the *MMP-stereo-proposed* with the MMP-based simulcast approach. This MMP version, without the ability to exploit inter-view redundancy, is denominated *MMP-intra-proposed* in these experiments. Note that this is the version that has been used to encode the reference/left image in all of our tests. A comparison with the results of the previous intra version of MMP (*MMP-intra*) (Rodrigues et al. 2008) is presented for the left image, in order to highlight the advantages of the new improved intra prediction model and new block sizes, previously described in Sect. 3.

Experiments were also conducted in order to evaluate the coding performance gains provided by the linear predictors based on LSP and BM algorithms. The following MMP configurations were evaluated: *MMP-stereo-BM* only uses the fixed linear predictors DC based on BM algorithm for inter-view prediction, similarly to current video coding standards; *MMP-stereo-staticLSP* uses all linear predictors based on BM and LSP methods, but adaptive coefficients are estimated using only TW-1 of Fig. 5, and *MMP-stereo-proposed* refers to the main proposal of this work using all linear predictors with improved adaptive predictors based on multiple TWs, as explained in Sect. 4.1. Table 2 summarizes the evaluated MMP configurations. In the experiments, the same λ values (used in the MMP Lagrangian cost function) were considered for both left and right images, specifically the values: 300, 75, 25 and 10. We adopted this strategy because efficient rate allocation between views was not subject of research in this work, and, if used, it could mask the differences between the prediction strategies.

Experimental results are presented in terms of Bjontegaard Delta PSNR (BDPSNR) results (Bjøntegaard 2001) for all stereo pairs of the test set in Table 3. In order to illustrate the gains of the new proposed intra prediction model, BDPSNR results for the left image encoded by *MMP-intra-proposed* relative to the *MMP-intra* are presented. For the right image, BDPSNR results for *MMP-stereo-BM*, *MMP-stereo-staticLSP* and *MMP-stereo-proposed* are computed relative to *MMP-intra-proposed*, to show the advantage of DC techniques over the fully-intra approach. In addition to these results, Fig. 9 shows the RD curve for Kendo stereo pair, as a representative result that illustrates the algorithm's behaviour from lower to higher bitrates. For the presented results, the bitrate is computed independently for the left and right images, even when inter-view prediction techniques are used.

In order to compress right image using DC methods, the encoded left image using *MMP-intra-proposed* algorithm was used as reference image for all experiments. We may notice that the RD results of the right image using DC techniques are usually significantly superior relatively to the results of *MMP-intra-proposed*. As shown in Fig. 9, the RD performance of MMP using the BM algorithm (*MMP-stereo-BM*) for the right image is far superior to the one of *MMP-intra-proposed*, with an average BDPSNR gain of 2.99 dB. These results

---

[1] The authors would like to thank Poznan University of Technology, Nagoya University-Tanimoto Lab, HHI, GIST, NICT, Nokia and Microsoft for providing *Poznan Street* and *Poznan Hall2*, *Kendo* and *Balloons*, *Book Arrival*, *Newspaper*, *Shark*, *GT Fly* and *Undo Dancer sequences*, *Ballet* and *Breakdancers*, respectively.

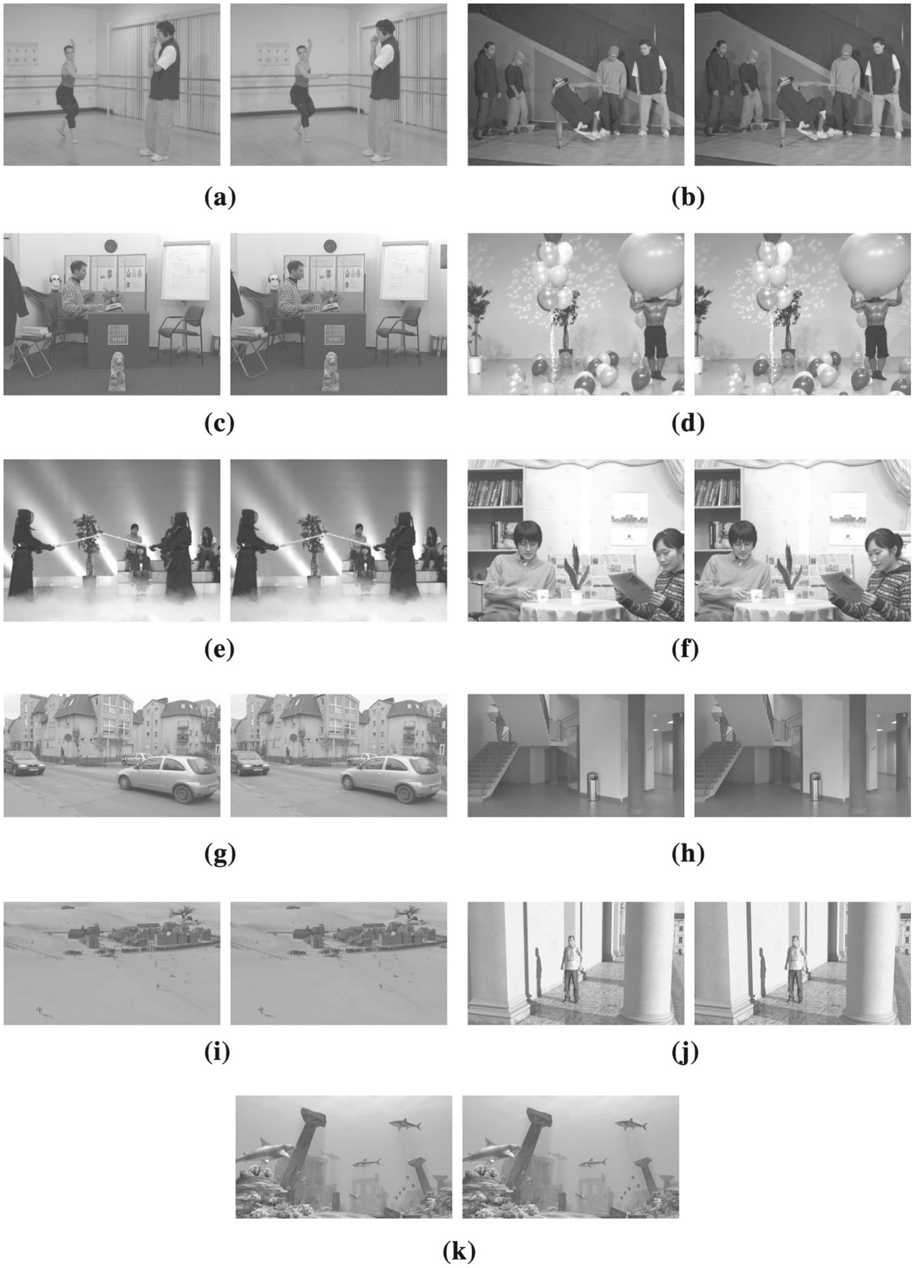

Fig. 8 Used test stereo pairs given by the frame 0 of the selected views of the multiview sequences. **a** Ballet (cameras 5 and 3), **b** Breakdancers (cameras 5 and 3), **c** Book Arrival (cameras 10 and 8), **d** Balloons (cameras 3 and 5), **e** Kendo (cameras 3 and 5), **f** Newspaper (cameras 4 and 6), **g** Poznan Street (cameras 4 and 3), **h** Poznan Hall2 (cameras 7 and 6), **i** GT Fly (cameras 5 and 1), **j** Undo Dancer (cameras 5 and 9), **k** Shark (cameras 1 and 5)

Table 2 MMP configurations for different intra and inter prediction techniques that have been evaluated in the presented experiments

MMP approach	Description
MMP-intra	MMP encodes one image using the intra techniques as presented in Rodrigues et al. (2008)
MMP-intra-proposed	MMP encodes one image using the proposed improved intra techniques
MMP-stereo-BM	MMP encodes right image using the proposed intra techniques and BM algorithm
MMP-stereo-staticLSP	MMP encodes right image using the proposed intra techniques and linear prediction using fixed TW for LSP algorithm
MMP-stereo-proposed	MMP encodes right image using the proposed intra techniques and linear prediction with improved adaptation

Table 3 BDPSNR results of MMP algorithm using different prediction framework configurations, for the left and right views of the presented test set

Stereo pairs	Left	Right		
	MMP-intra	*MMP-stereo*		
	proposed	*BM*	*StaticLSP*	*Proposed*
Ballet	0.8550	0.1483	0.2716	0.2796
Breakdancers	0.3012	0.4716	0.5723	0.5908
Book Arrival	0.4472	2.6809	2.8817	2.9185
Balloons	0.5267	2.7064	3.3856	3.4313
Kendo	0.6512	3.2705	4.2456	4.2800
Newspaper	0.4844	2.4452	3.0118	3.0969
Poznan Street	0.4840	2.2671	2.4976	2.5236
Poznan Hall2	0.7543	1.0098	1.1542	1.1845
GT Fly	0.4909	4.2707	4.3811	4.3830
Undo Dancer	0.2987	5.4748	5.5486	5.5896
Shark	0.4304	8.1917	8.2855	8.2755
Average	0.5204	2.9943	3.2941	3.3230

confirm the advantage of inter prediction techniques that exploit stereo redundancy between views.

In order to demonstrate the advantage of the proposed improved TW adaptation for LSP method, *MMP-stereo-staticLSP* is evaluated using BM and LSP predictors, with fixed TW for coefficient estimation. The results of both Table 3 and Fig. 9 show that *MMP-stereo-staticLSP* significantly improves the RD performance for the right image when comparing with *MMP-stereo-BM*. BDPSNR results show an average gain of 3.29 dB over *MMP-intra-proposed* which corresponds to an improvement of 0.3 dB relative to *MMP-stereo-BM*. This advantage is confirmed by the Wilcoxon signed-rank test (Siegel 1956), where the performance difference (*MMP-stereo-BM* minus *MMP-stereo-staticLSP*) results in a Z-value of −2.9341 based on positive rank, corresponding to a p value of 0.00338, which is significant enough to reject the null hypothesis (inferior to 0.05).

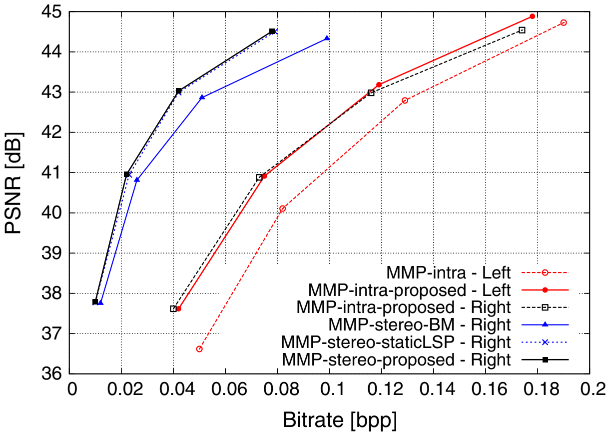

Fig. 9 RD performance evaluation of the proposed prediction techniques for the *Kendo* stereo pair

When we enable the proposed TW adaptation for LSP method (*MMP-stereo-proposed*), experiments show that further improvements to RD results can be achieved. Table 3 shows that the average PSNR gain of *MMP-stereo-proposed* over *MMP-intra-proposed* is 3.32 dB, which is superior to the gains of BM-based MMP approach (*MMP-stereo-BM*). This result is also confirmed by the Wilcoxon signed-rank test, where the performance difference (*MMP-stereo-BM* minus *MMP-stereo-proposed*) results in a Z-value of -2.9341 based on positive rank, corresponding to a p value of 0.00338, which is significant enough to reject the null hypothesis (inferior to 0.05). From these results, one may conclude that the proposed predictive framework is able to better exploit the stereo redundancies than the state-of-the-art BM algorithm, used in the most recent stereo image coding standards.

By comparing the BDPSNR results of *MMP-stereo-staticLSP* and *MMP-stereo-proposed*, one may conclude that the selection of the optimal TW shape improves LSP adaptation. This is reflected in a slight increase of the overall encoder RD performance for most stereo pairs. Analysing the performance difference (*MMP-stereo-staticLSP* minus *MMP-stereo-proposed*) using the Wilcoxon signed-rank test resulted in a Z-value of -2.667 based on positive rank, corresponding to a p value of 0.00758, which is significant enough to reject the null hypothesis (inferior to 0.05) and accept the superiority of *MMP-stereo-proposed*.

When evaluating the RD performance of these methods from low to high bitrates in Fig. 9, one may observe that LSP-based approaches provide the highest coding gains mainly at medium and high bit rates. This is because at lower rates the reference/left image samples used for LSP optimization have higher distortion, which compromises the performance of LSP training procedure for coefficient estimation.

Regarding the results for the left image, Table 3 and Fig. 9 show that significant performance gains were achieved by using the proposed improved intra prediction framework, as well as the new block sizes for MMP. One observes average BDPSNR gain of 0.52 dB, for *MMP-intra-proposed* relative to the previous *MMP-intra* algorithm, which demonstrates that the proposed improvements are effective for the MMP algorithm, either for intra or stereo image coding. Figure 9 shows that observed gains are superior at lower rates mainly due to the fact that larger blocks are mostly used at these rates.

The simulcast approach for stereo image coding uses *MMP-intra-proposed* to encode both the left and right images of the stereo pair. As observed in Fig. 9, *MMP-intra-proposed*

RD curves are very similar for both the left and right images, which was expected, since no redundancy is exploited between views and the information contained in each image is similar.

5.2 Comparison with the state-of-the-art video coding algorithms

The H.264/AVC reference software JM-18.5 was employed in these experiments using the Stereo High profile (ITU-T and ISO/IEC JTC1 2010). The default configuration file *encoder_stereo.cfg* under JM source for stereo coding was considered. The *QPISlice* and *QPPSlice* parameters were configured using the equal QP values: 25, 30, 35 and 40.

We also performed experiments using the MV-HEVC with reference software HTM-11.2 for comparison purposes. The default configuration file for multiview video coding was used, setting the *NumberOfLayers* parameter to 2, for stereo image coding. The used QP values were the same as the ones used with H.264/AVC, for both the left and right images, corresponding to the values recommended in common test conditions document (Muller and Vetro 2014).

Both the MV-HEVC and H.264/AVC configurations used *FramesToBeEncoded* parameter equal to one, the disparity search range was set to 96, and the *SearchMode* parameter, associated to the disparity compensation mode, was set to *Full search*, as done in the MMP algorithm. In order to fairly compare the performance of the *MMP-stereo-proposed* with its state-of-the-art counterparts, the PSNR information of the luminance of each view is given in function of the global rate used to encode both views. The global rate is used to generate both the RD curves and BDPSNR results (Fig. 10; Table 4, respectively). Similarly to the transform-based standards, MMP also used a post-deblocking filter algorithm on both image views, based on the method proposed in Francisco et al. (2012).

The summary of the BDPSNR results for all images of the test set is presented in Table 4. We can notice that the proposed MMP algorithm presents consistent RD gains over the H.264/AVC standard. The BDPSNR results between both algorithms show that MMP is superior, with gains ranging approximately from 0.38 dB up to 1.12 dB for the left image, and from 0.76 dB up to 1.73 dB for the right image. Relatively to the MV-HEVC algorithm, MMP presents an equivalent performance for the right image, despite the reference/left image being less efficiently encoded. On average, MV-HEVC presents a BDPSNR gain of 0.14 dB for the left image, while MMP outperforms it by almost 0.05 dB for the right image. When applying the Wilcoxon signed-rank test to the Bjontegaard results comparing *MMP-stereo-proposed* and MV-HEVC algorithms, we conclude that there is no evidence to reject the null hypothesis (stating that average difference is null), since the p value is 0.09102 for the left image and 0.4777 for the right image, being greater than the significance level (0.05 by default). Thus, in statistical terms, we can not rely on the presented average Bjontegaard results comparing MMP with the MV-HEVC algorithm. Nevertheless, these results suggest the advantage of the proposed disparity compensation model, which is able to present a competitive coding performance for the right image, even using a reference image worse than MV-HEVC.

The rate-distortion curves in Fig. 10 compare the MMP, H.264/AVC and MV-HEVC algorithms, for the stereo pairs *GT Fly* and *Breakdancers*. These curves provide more detailed results than BDPSNR values, illustrating the algorithms' performance from low to high bitrates. Only two stereo pairs, one natural and one synthetic, were chosen due to space constraints.

The difference between the H.264/AVC and MMP curves in Fig. 10 shows a clear advantage of the MMP algorithm for both left and right images, at all bitrates. These results agree

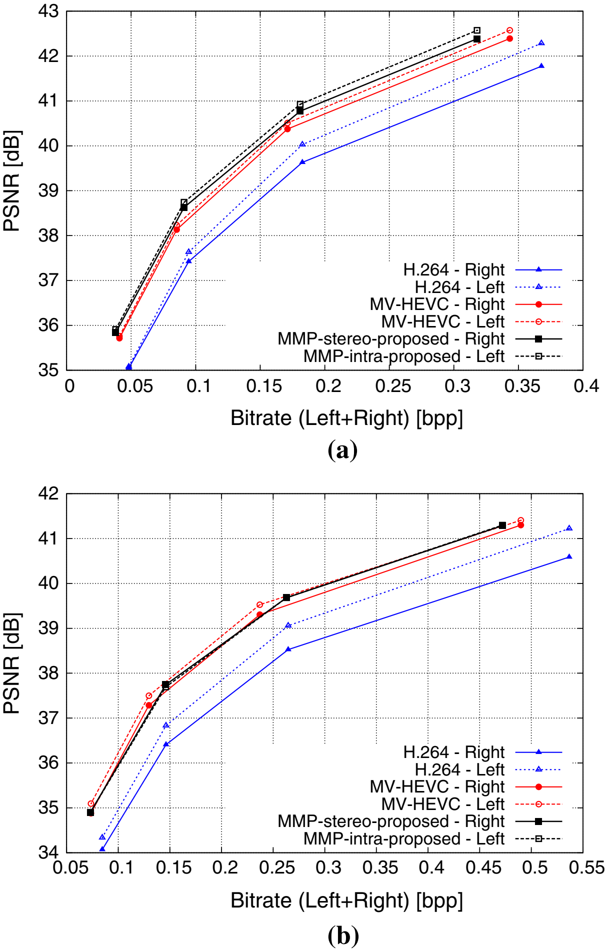

Fig. 10 RD coding performance evaluation of the proposed MMP versus H.264/AVC and MV-HEVC. **a** GT Fly and **b** Book Arrival

with the BDPSNR values, demonstrating the advantage of the proposed intra and inter coding methods. Relatively to the MV-HEVC algorithm, one may observe that the rate distortion performance of MMP proposal is competitive, outperforming it for the *GT Fly* stereo pair and for some rate-distortion points of *Breakdancers*.

The proposed prediction method uses a higher number of operations than the block-matching disparity estimation methods, because it needs to estimate the adaptive linear predictors using least-squares algorithm in addition to the fixed predictors based on block matching algorithm. The least-squares algorithm tends to be more complex than the block matching algorithm, however, the pattern-matching residue coding method has the higher impact in the computational complexity of MMP algorithm. In this work, this issue was aggravated, not only by the new linear predictors for disparity compensation, but also by the increased number of intra prediction modes and larger initial block size, which increased the number of encoded residue blocks during rate-distortion optimization loop. Thus, when compared to the transform-based HEVC standard, the computational complexity of the pre-

Table 4 BDPSNR results of the proposed MMP encoder over H.264/AVC and MV-HEVC, for the used test set

Stereo pairs	H.264/AVC		MV-HEVC	
	Left	Right	Left	Right
Ballet	1.0964	1.4045	−0.0926	0.0989
Breakdancers	0.3831	0.7578	−0.1615	−0.0090
Book Arrival	0.7644	1.2549	−0.1730	0.0507
Balloons	0.9465	1.7336	−0.4231	−0.1443
Kendo	1.0887	1.7671	−0.4152	−0.0343
Newspaper	0.8409	1.5340	−0.2633	−0.0548
Poznan Street	0.6115	1.2828	−0.2500	−0.0207
Poznan Hall2	0.8144	1.2490	−0.0170	0.1291
GT Fly	1.1209	1.2689	0.2831	0.2664
Undo Dancer	0.9222	1.0450	0.2267	0.2367
Shark	0.9265	1.2099	−0.2374	−0.0256
Average	0.8650	1.3189	−0.1385	0.0448

sented MMP algorithm, is about three orders of magnitude higher. However we believe that the execution speed of the MMP algorithm could be largely improved by using more efficient algorithmic implementations as well as exploiting parallel processing architectures.

6 Conclusion

A novel linear predictive algorithm for disparity compensation based on least-squares prediction and block-matching algorithms is proposed in this paper. Stereo redundancy is reduced by means of adaptive linear predictors implicitly estimated using least-squares methods and by explicit linear predictors based on well-known block-matching algorithm using quarter sample accuracy. The proposed disparity compensation framework has been implemented on the Multidimensional Multiscale Parser paradigm. This algorithm has been also improved using a more advanced intra prediction framework and new initial block size.

Experimental rate-distortion results suggest that, for stereo image coding, the proposed pattern-matching-based encoder can have competitive rate-distortion performances when compared to the traditional transform-quantization-entropy coding paradigm. Furthermore, the presented linear prediction solution for disparity compensation was shown to be a worthy generalization of the block-matching algorithm, typically used in current state-of-the-art image coding standards, presenting average BDPSNR gains superior to 0.3 dB over the traditional disparity compensation techniques.

References

Accame, M., De Natale, F., & Giusto, D. (1995). Hierarchical block matching for disparity estimation in stereo sequences. *International Conference on Image Processing*, *2*, 374–377.

Bjøntegaard, G. (2001). Calculation of average psnr differences between RD-curves. *ITU-T SG 16 Q6 VCEG, Doc VCEG-M33*.

Carvalho, M., da Silva, E., & Finamore, W. (2002). Multidimensional signal compression using multiscale recurrent patterns. *Elsevier Signal Processing*, *82*, 1559–1580.

Chen, Y., Wang, Y. K., Ugur, K., Hannuksela, M. M., Lainema, J., & Gabbouj, M. (2008). The emerging MVC standard for 3D video services. *EURASIP Journal on Advances in Signal Processing*, *2009*, 1–13.

Dinstein, I., Guy, G., Rabany, J., Tzelgov, J., & Henik, A. (1988). On stereo image coding. *9th International Conference on Pattern Recognition*, *1*, 357–359.

Duarte, M., Carvalho, M., da Silva, E., Pagliari, C., & Mendonca, G. (2005). Multiscale recurrent patterns applied to stereo image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *15*(11), 1434–1447.

Ellinas, J. N., & Sangriotis, M. S. (2006). Stereo image coder based on the MRF model for disparity compensation. *EURASIP Journal on Advances in Signal Processing*, *2006*(1), 1–13. doi:10.1155/ASP/2006/73950.

Frajka, A., & Zeger, K. (2002). Residual image coding for stereo image compression. In *IEEE international conference on image processing* (Vol. 2). http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1039926.

Francisco, N. C., Rodrigues, N. M. M., da Silva, E. A. B., de Carvalho, M. B., de Faria, S. M. M., da Silva, V. M. M., & Reis, M. J. C. S. (2008). Multiscale recurrent pattern image coding with a flexible partition scheme. In *15th IEEE international conference on image processing* (pp. 141–144). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4711711.

Francisco, N., Rodrigues, N., da Silva, E., Carvalho, M., Faria, S., & Silva, V. (2010). Scanned compound document encoding using multiscale recurrent patterns. *IEEE Transactions on Image Processing*, *19*(10), 2712–2724.

Francisco, N., Rodrigues, N., da Silva, E., & Faria, S. (2012). A generic post-deblocking filter for block based image compression algorithms. *Signal Processing: Image Communication*, *27*(9), 985–997.

Graziosi, D. B., Rodrigues, N. M. M., da Silva, E. A. B., de Faria, S. M. M., & de Carvalho, M. B. (2009). Improving multiscale recurrent pattern image coding with least-squares prediction mode. In *16th IEEE international conference on image processing* (ICIP) (pp. 2813–2816). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5414219.

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

ITU-T, ISO/IEC JTC1 (2010) Advanced video coding for generic audiovisual services. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC).

Kaup, A., & Fecker, U. (2006). Analysis of multi-reference block matching for multi-view video coding. In *Proceedings of 7th workshop digital broadcasting* (pp. 33–39).

Li, X., & Orchard, M. (2001). Edge-directed prediction for lossless compression of natural images. *IEEE Transactions on Image Processing*, *10*(6), 813–817.

Lucas, L., Rodrigues, N., de Faria, S., da Silva, E., Carvalho, M., & da Silva, V. (2010). Intra-prediction for color image coding using YUV correlation. *17th IEEE international conference on image processing* (pp. 1329–1332).

Lucas, L., Rodrigues, N., da Silva, E., & Faria, S. (2011a). Adaptive least squares prediction for stereo image coding. *18th IEEE international conference on image processing* (pp. 2013–2016).

Lucas, L., Rodrigues, N., da Silva, E., & Faria, S. (2011b). Stereo image coding using dynamic template-matching prediction. *IEEE EUROCON2011—International conference on computer as a tool* (pp. 1–4).

Marpe, D., Schwarz, H., & Wiegand, T. (2003). Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 620–636.

Merkle, P., Smolic, A., Muller, K., & Wiegand, T. (2007). Efficient prediction structures for multiview video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *17*(11), 1461–1473.

Mueller, K., Schwarz, H., Marpe, D., Bartnik, C., Bosse, S., Brust, H., et al. (2013). 3D high efficiency video coding for multi-view video and depth data. *IEEE Transactions on Image Processing*, *22*(9), 3366–3378.

Muller, K., & Vetro, A. (2014). Common test conditions of 3DV core experiments. *Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 7th Meeting*, San Jos, USA.

Muller, K., Merkle, P., & Wiegand, T. (2011). 3-D video representation using depth maps. *Proceedings of the IEEE*, *99*(4), 643–656.

Palaz, D., Tosic, I., & Frossard, P. (2011). Sparse stereo image coding with learned dictionaries. *2011 18th IEEE international conference on image processing (ICIP)* (pp. 133–136).

Perkins, M. (1992). Data compression of stereopairs. *IEEE Transactions on Communications*, *40*(4), 684–696.

Rodrigues, N., da Silva, E., Carvalho, M., Faria, S., & Silva, V. (2005). Universal image coding using multiscale recurrent patterns and prediction. *IEEE international conference on image processing*.

Rodrigues, N., da Silva, E., Carvalho, M., Faria, S., & Silva, V. (2008). On dictionary adaptation for recurrent pattern image coding. *IEEE Transactions on Image Processing*, *17*(9), 1640–1653.

Seo, S. H., Azimi-Sadjadi, M., & Tian, B. (2000). A least-squares-based 2-D filtering scheme for stereo image compression. *IEEE Transactions on Image Processing*, *9*(11), 1967–1972.

Sethuraman, S., Siegel, M., & Jordan, A. G. (1995). Multiresolutional region-based segmentation scheme for stereoscopic image compression. *Proceedings of SPIE*, *2419*, 265–274. doi:10.1117/12.206365.

Siegel, S. (1956). *Non-parametric statistics for the behavioral sciences* (pp. 75–83). New York: McGraw-Hill.

Stankowski, J., Domanski, M., Stankiewicz, O., Konieczny, J., Siast, J., & Wegner, K. (2012). Extensions of the HEVC technology for efficient multiview video coding. *2012 19th IEEE international conference on image processing (ICIP)* (pp. 225–228).

Tech, G., Wegner, K., Chen, Y., Hannuksela, M. M., & Boyce, J. (2013). MV-HEVC draft text 5. *Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3V-E1004, 5th Meeting*. Vienna, Austria

Vetro, A. (2010). Frame compatible formats for 3d video distribution. *17th IEEE international conference on image processing* (pp. 2405–2408).

Woo, W., & Ortega, A. (2000). Overlapped block disparity compensation with adaptive windows for stereo image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *10*(2), 194–200.

Woo, W., & Ortega, A. (1996). Stereo image compression with disparity compensation using the MRF model. *Proc SPIE VCIP* (pp. 28–41)

Woo, W., & Ortega, A. (1999). Optimal blockwise dependent quantization for stereo image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *9*(6), 861–867.