# ADAPTIVE LEAST SQUARES PREDICTION FOR STEREO IMAGE CODING

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

[1]Instituto de Telecomunicações; [2]ESTG, Instituto Politécnico de Leiria, Portugal;
[3]PEE/COPPE/DEL/Poli, Universidade Federal do Rio de Janeiro, Brazil
*e-mails: luisfrlucas@gmail.com, nuno.rodrigues@co.it.pt, eduardo@lps.ufrj.br, sergio.faria@co.it.pt*

## ABSTRACT

State-of-the art approaches towards stereo image coding exploit inter-view redundancy by employing block-matching methods for disparity estimation and compensation. However, the efficiency of these methods is affected by mismatched areas, due to occlusions, brightness variations, or perspective distortion between objects of the two views. In this paper we present a new prediction scheme for stereo image coding, that combines an implicit disparity estimation method, with an adaptive least squares (LS)-based filtering. The Multidimensional Multiscale Parser image coding algorithm was used to evaluate the efficiency of the proposed scheme. Experimental results demonstrate the advantage of LS prediction in stereo image coding. Furthermore, the rate-distortion performance of the MMP based stereo encoder is well above that of the state-of-the-art H.264/AVC Stereo Profile, especially at medium and high bit rates.

***Index Terms—*** Stereo Image Coding, Least Squares Method, Disparity Compensation, Inter-frame Prediction

## 1. INTRODUCTION

Stereoscopic vision can be experienced through the use of two images, obtained from the same scene using a pair of cameras. These two images are also referred to as a stereo pair. Depth is artificially perceived by the human brain, by processing the relative displacement between picture objects of the stereo pair, known as disparity. The vertical component of disparity is null, whenever the stereo pair is acquired with two aligned cameras, i.e., the cameras have parallel optical axes (alternatively, a rectification process may be applied).

However, generated data by stereo images requires twice as much bandwidth as a single image. With the recent growth of 3-D applications, efficient transmission and storage of stereo pairs is increasingly important. Thus, several compression techniques for this type of data have been proposed.

One simple approach is to independently encode each view of the stereo pair. This method guarantees backward compatibility and a minimum computation delay. A drawback of such strategy is its low coding efficiency, because the high correlation that may exist between both views is not exploited. This problem has been addressed in the state-of-the-art stereo image/video coding standard, Multiview Video Coding (MVC) [1]. The MVC is an extension of the H.264/MPEG-4 AVC standard, that improves the coding efficiency for stereo or multiview video, by exploiting the redundancy over time and across views.

This project was funded by FCT - "Fundação para a Ciência e Tecnologia", Portugal, under the grant SFRH/BD/45460/2008, and Project COMUVI (PTDC/EEA-TEL/099387/2008)

The main technique used by MVC for inter-view prediction is the block-based disparity compensation (DC). The DC approach is similar to motion estimation used for temporal prediction of video coding. The optimal disparity vector, that minimizes the encoding error for each block, is explicitly transmitted to the decoder.

Despite the success of the stereo image encoders based on DC, they do not fully exploit the redundancy between the existing views. The block-based disparity compensation is, in fact, unable to successfully represent some common effects between the stereo pair views, like: geometric distortion, due to the image acquisition cameras perspective; occlusion areas, due to variable scene depth; and brightness variations between views.

In this paper we propose a new paradigm for stereo image coding that does not perform explicit DC. The proposed algorithm consists on an implicit disparity estimation algorithm, improved by an adaptive LS-based prediction. The LS-based predictor locally optimizes its coefficients, based on causal decoded portions of data. Experimental results show that the proposed LS prediction (LSP) is much more efficient in compensating usual mismatches between views than the traditional algorithms.

LSP has been successfully applied both for image and video coding [2], [3], [4], [5]. In [6], LSP was used for stereo image coding. However, its authors present PSNR gains without considering the mandatory encoding of filter coefficients. Furthermore, the original left image is used as reference to the compression of the right image, which significantly improves the prediction results.

This paper is organized as follows. Section 2 presents the LSP derivation and its application on image coding. Section 3 presents the proposed prediction scheme and the LS-based prediction for stereo images. Experimental results are shown and discussed in Section 4, while Section 5 concludes the paper.

## 2. LEAST SQUARES PREDICTION OF IMAGE SIGNALS

LSP computes the predictor pixels of an image block by filtering a selected set of pixels, $X(n_i)$, that belong to a causal neighborhood of the image (already encoded), referenced to as the *filter support*. Since the causal pixels were previously encoded, the LSP filtering can be performed at both the encoder and decoder. If $n$ is the spatial coordinate of the image, the linear prediction of $X(n_0)$, $\hat{X}(n_0)$, is given by:

$$\hat{X}(n_0) = \sum_{i=1}^{N} a_i X(n_i), \tag{1}$$

where N is the linear filter order. The set of filter coefficients, $a_i$, are locally optimized in the least-squares sense using a previously encoded region of the image, known as the *training area*. Figure 1 shows an example of a training area.

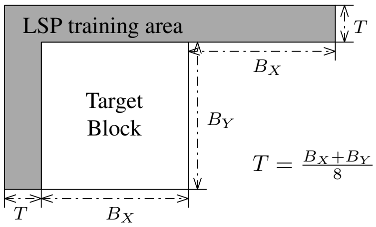



**Fig. 1**: LSP training area.

Considering a training area with $M$ elements, they can be rearranged in an $M \times 1$ column vector $\vec{y} = [X(n_1) \cdots X(n_M)]^T$. If the N causal neighbors for each pixel of the training window are joined in a row vector, a matrix $C$ with size $M \times N$ is formed:

$$C = \begin{bmatrix} X(n_{1,1}) & \cdots & X(n_{1,N}) \\ \vdots & & \vdots \\ X(n_{M,1}) & \cdots & X(n_{M,N}) \end{bmatrix} \quad (2)$$

The optimal filter coefficients are determined by least-squares optimization, finding the solution for

$$\min \left\{ \|\vec{y}_{M\times 1} - C_{M\times N}\vec{a}_{N\times 1}\|^2 \right\}. \quad (3)$$

A well-known closed form solution for this LS problem is

$$\vec{a} = (C^T C)^{-1}(C^T \vec{y}). \quad (4)$$

The high efficiency of LSP for intra prediction has been widely reported in the literature. In [2], the edge-directed property is the main motivation for applying LSP in image coding. LSP is able to adaptively learn the edge orientation information from the local causal data without requiring its explicit estimation.

It has also been demonstrated that LSP is able to efficiently model slow motion information between video frames [3]. In fact, a motion trajectory can be regarded as an edge contour in 2D, since both form an homogeneous intensity field along a continuous space. In order to enhance LSP performance for temporal prediction, a spatio-temporal filter has been used, with an adaptive update of the filter support. For fast motion, the use of a camera panning compensation, with an adaptive temporal warping algorithm, was proposed, in order to avoid the degradation of the LSP performance.

## 3. STEREO IMAGE PREDICTION USING LSP

### 3.1. Application of LSP to Stereo Pairs

Our major motivation to use LSP on stereo image prediction comes from its known efficiency in creating an accurate prediction model, based on a related reference area. The proposed algorithm combines an adaptive LSP with an implicit disparity compensation method. The disparity compensation step estimates the best matched region in the reference view for the training area of LSP. The aim of this procedure is to use co-located regions of the stereo pair for the LSP algorithm, improving its performance.

With the optimization process, applied in the causal training area, LSP finds the best set of coefficients, that transform each filter support pixels of the training area in the actual pixels of that area. In the proposed prediction method for stereo image coding, LSP uses a filter support, composed by pixels belonging to the reference image. Additionally, some neighbor causal pixels of the right image may also be used. Figure 2 shows an example of the filter support pixels used by the proposed LSP scheme, including 9 neighbors in the left image and 4 nearest causal neighbors in the right image. The predictor is trained to approximate each pixel, $X(n_0)$, belonging to the training area, represented in Figure 1.

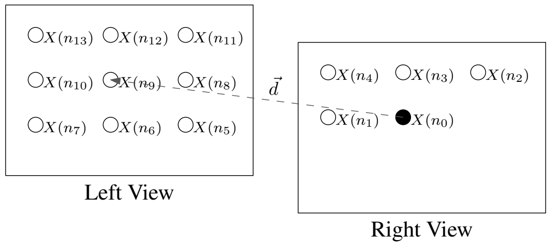



**Fig. 2**: Example of an LSP filter support.

The LSP training procedure is performed using only previously encoded regions of the images. Therefore, there is no need to transmit the prediction coefficients. The decoder simply repeats the training step, in order to determine the values of LSP coefficients. The width of the training area, $T$, represented in Figure 1, was adaptively set by using a value $T = (B_X + B_Y)/8$. This relation resulted from experimental observations. The computational complexity of the LSP method is also reduced by computing a set of prediction coefficients common to all pixels of the target block.

### 3.2. Disparity Compensation using Template Matching

Disparity estimation determines a corresponding pixel in the left (reference) image for each pixel in the right image. The distance between these pixels is given by a disparity vector $\vec{d}$. In order for LSP to perform efficiently, the training area and the region covered by the filter support for each of its pixels should have similar texture features. For stereo images, some care must be taken in the choice of the filter support, since some of its pixels will be placed in a different view (Figure 2). The efficiency of LSP method is thus increased by centering the portion of the filter support, that belongs to the left image, by a disparity vector, determined for the training area of the right image. This corresponds to vector $\vec{d}$ in Figure 2.

We propose to perform the disparity estimation using a template-matching (TM) based algorithm [7]. The main advantage of such algorithm is that no side information needs to be transmitted. TM acts in a similar way to the well-known block-matching method, used for motion compensation, where the estimated disparity value corresponds to the best-matched block. However, instead of using the temporal prediction block to perform the matching, TM uses the causal pixels that surround the block. In the proposed method, the size of the template area and of the training window used for the LSP optimization are the same (presented in Figure 1). The searching procedure to find the best template uses an integer precision of one pixel. Since the TM-based disparity estimation is unable to accurately predict deformed or occluded areas, LSP is tested using a support region centered on the best match, $\vec{d_b}$, returned by TM, but also for a $5 \times 3$ pixel area around this position. The LSP instance which gives the lowest training error in the referred area is chosen for the prediction.

### 3.3. Filter Support Adaptation of LSP

Depending on the original scene, each stereo image block may have different features. Furthermore, LSP may be used to predict mismatching blocks, where different deformations can be found. Thus, depending on each case, different LSP filter support may have impact on the prediction efficiency. One example is the prediction of

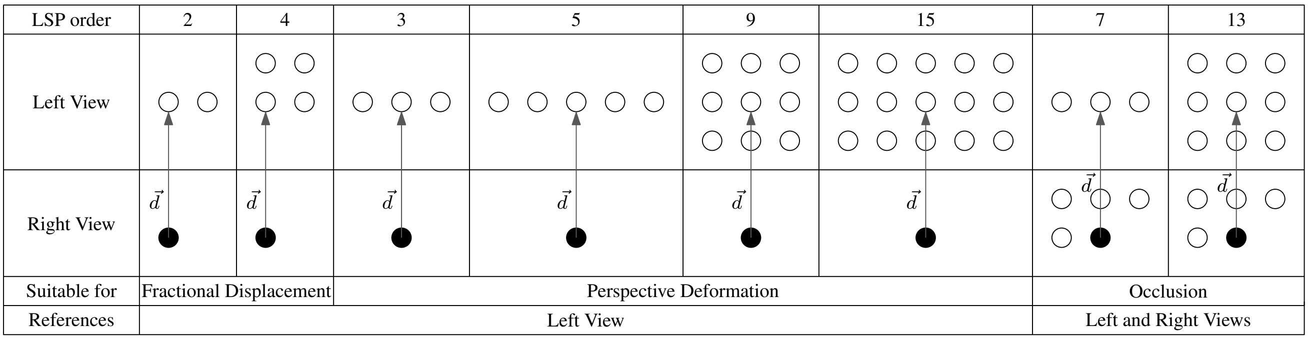



**Fig. 3**: Supports proposed for the explicit LSP selection.

occluded areas, that are common for stereo image pairs. In these areas, using a filter support that includes pixels from the target image as well as pixels from the left image, may improve the efficiency of the LSP prediction.

In order to optimize the performance of LSP, we propose to adaptively determine the filter support for each target block. Figure 3 shows the proposed filter support configurations, chosen accordingly to the results of our experimental tests. The first six support types use only elements from the reference view. The last two configurations use both disparity compensated pixels (from the reference view) and spatial neighborhood pixels (from the current view) for LSP training. LSP is thus able to use the available data from both regions, in an optimum manner. A flag is transmitted, that allows the decoder to choose which support region to use.

The used configurations for the support regions have different motivations. For well-matched blocks, low order supports are preferred, while higher order supports are more efficient for prediction in mismatched areas. Figure 3 shows some of the features that may be compensated by each of the eight proposed configurations. For example, when an occlusion occur, the latter modes, that use both disparity and spatial neighbors, may be used, with larger weights assigned to spatial neighbors than to the stereo neighbors. One interesting fact is the implementation of fractional disparity compensation by the low order models, represented in Figure 3. For example, the proposed second order model implements a traditional fractional disparity compensation in the horizontal direction.

When none of the filter supports improve the prediction, only the implicit TM step is performed. In this case, the block prediction is just given by the left image pixels pointed by the implicitly calculated disparity vector $\vec{d}$.

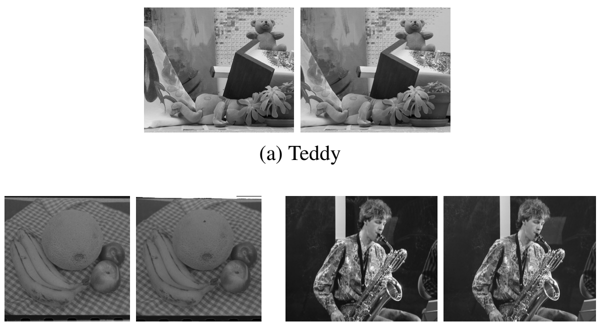
## 4. EXPERIMENTAL RESULTS

In order to demonstrate the efficiency of the proposed algorithm for stereo image prediction, we have implemented the new LSP method in the Multidimensional Multiscale Parser (MMP) encoder [8], [9]. MMP is a block based encoder that uses an hierarchical prediction scheme, based on the H.264/AVC standard prediction, with an additional LSP-based intra-frame prediction mode [4]. Image residue is encoded by using approximated patterns at different scales from an adaptive dictionary. MMP also uses an adaptive block size for prediction, with blocks ranging from $4 \times 4$ to $16 \times 16$. The flexible partitioning scheme used for prediction should also suit the proposed LSP method for stereo image coding with MMP, since it increases the accuracy of the prediction segmentation. It is known that the MMP algorithm presents a high computational complexity, which is inherent to its pattern matching approach. However, the complexity of LSP is not significant when compared to the MMP algorithm.

The proposed LSP method was added to the RD loop of MMP, in addition to the intra prediction modes. In order to evaluate the performance of LSP, we performed two different tests: the first test uses the proposed LSP algorithm for block prediction (MMP proposed); the second test uses only a TM-based disparity estimation (MMP only TM). The results of these tests demonstrate the relative performance of using LSP over a version that also uses TM disparity estimation, but does not apply LSP. We also compare the MMP rate-distortion (RD) performance with that of the state-of-the-art stereo image encoder H.264/AVC Stereo High Profile [10] using the default configuration file for stereo image coding in the JM-17.2 version of the reference software. The test set used in our simulations comprises the stereo images shown in Figure 4. Image Teddy was obtained with a parallel camera arrangement, while Fruit and Saxo were captured with convergent cameras.

Figures 5 to 7 show the RD results for both views of each stereo pair. For all image pairs, we may notice the superior RD performance of the MMP algorithm for the reference (left) images. As mentioned before, the proposed prediction scheme is only applied for the right image. For this image, at all rate distortion points, the use of the LSP prediction significantly improves the implicit TM-based disparity estimation. This results from the ability of LSP to compensate mismatched or occluded areas, regions with variable brightness, or even to perform disparity estimation with fractional accuracy. An analysis of the relative usage rate of all of the available prediction modes, shows that the LSP mode is used over 50% of the time. This measurement further demonstrates the usefulness of LSP for stereo coding.

The presented plots also show some promising results, when we compare the performance of the proposed MMP-based stereo image encoder with H.264/AVC. From medium to high bit rates, the pro-



(a) Teddy

(b) Fruit (c) Saxo

**Fig. 4**: Test images.

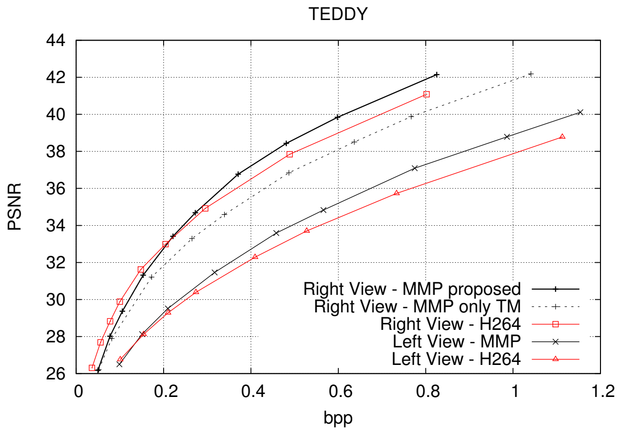



**Fig. 5**: Experimental results for stereo image Teddy.

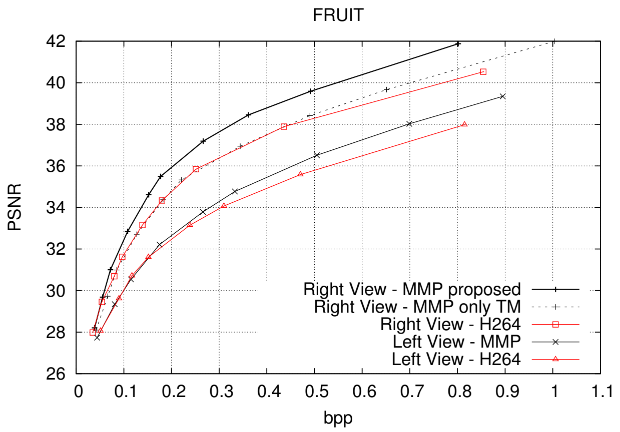



**Fig. 6**: Experimental results for stereo image Fruit.

posed scheme achieves gains of up to 1.5 dB, for the stereo image Fruit. At low bit rates, the quantization errors slightly impact on the performance of LSP, due to the backward adaptation process.

One may also notice that the achieved gains vary for different image pairs. For other images, with less mismatched and occluded blocks, as well as images without brightness variance, the advantage of the proposed scheme over H.264/AVC is smaller. However, the overall coding RD performance of the proposed prediction scheme, combined with MMP, is consistently superior to the one of the state-of-the-art H.264/AVC Stereo High Profile encoder.

## 5. CONCLUSION

In this paper we propose an LS-based prediction scheme, combined with template matching, for stereo image coding. Our experimental results have shown that this approach is able to efficiently predict stereo images, by reducing the effects of mismatching, occlusions and brightness variance. Also, for image pairs with small perspective deformation between views, the proposed algorithm is able to compensate fractional disparity values.

In order to test the performance of the proposed scheme, we have combined stereo LSP with the MMP algorithm. The resulting encoder achieves a high coding gain over the traditional MMP-intra scheme. Furthermore, the MMP based approach developed for stereo image coding outperforms the H.264/AVC Stereo High Profile encoder, with gains that can range up to 1.5dB at medium and high bit rates.

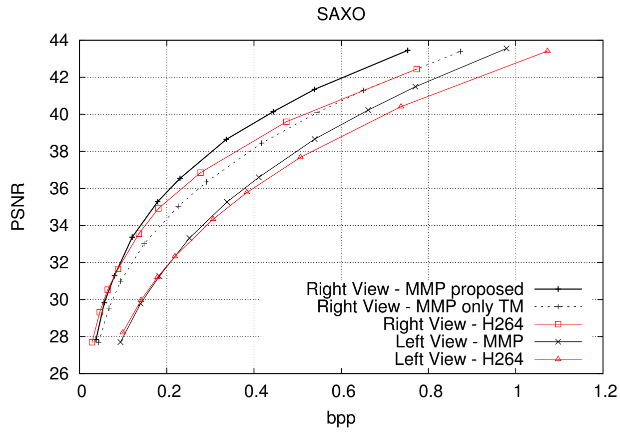



**Fig. 7**: Experimental results for stereo image Saxo.

## 6. REFERENCES

[1] Ying Chen, Ye-Kui Wang, Kemal Ugur, Miska M. Hannuksela, Jani Lainema, and Moncef Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP J. Appl. Signal Process.*, vol. 2009, pp. 8:1–8:13, January 2008.

[2] Xin Li and M.T. Orchard, "Edge-directed prediction for lossless compression of natural images," *Image Processing, IEEE Transactions on*, vol. 10, no. 6, pp. 813 –817, June 2001.

[3] Xin Li, "Least-square prediction for backward adaptive video coding," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 126–126, January 2006.

[4] D. B. Graziosi, N. Rodrigues, E. da Silva, S. de Faria, and M. de Carvalho, "Improving multiscale recurrent pattern image coding with least-squares prediction," in *IEEE Int. Conf. on Image Proc.*, Nov 2009.

[5] L.F.R. Lucas, N.M.M. Rodrigues, S.M.M. de Faria, E.A.B. da Silva, M.B. de Carvalho, and V.M.M. da Silva, "Intra-prediction for color image coding using YUV correlation," in *IEEE Int. Conf. on Image Proc.*, 2010, pp. 1329 –1332.

[6] Sang-Hoon Seo, M.R. Azimi-Sadjadi, and Bin Tian, "A least-squares-based 2-D filtering scheme for stereo image compression," *Image Processing, IEEE Transactions on*, vol. 9, no. 11, pp. 1967 –1972, Nov. 2000.

[7] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato, and C. Seng Boon, "Inter frame coding with template matching spatio-temporal prediction," *IEEE Int. Conf. on Image Proc.*, October 2004.

[8] N. Rodrigues, E. da Silva, M.de Carvalho, S. de Faria, and V. Silva, "On dictionary adaptation for recurrent pattern image coding," *Image Processing, IEEE Transactions on*, vol. 17, no. 9, pp. 1640–1653, September 2008.

[9] N.C. Francisco, N.M.M. Rodrigues, E.A.B. da Silva, M.B. de Carvalho, S.M.M. de Faria, and V.M.M. Silva, "Scanned compound document encoding using multiscale recurrent patterns," *Image Processing, IEEE Transactions on*, vol. 19, no. 10, pp. 2712–2724, October 2010.

[10] http://iphome.hhi.de/suehring/tml/download/.