A fast two-stage OMP algorithm for coding stereo image residuals

Wan-Fung Cheung and Yuk-Hee Chan

Center for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

Over-complete signal decomposition has been shown to be an efficient technique for coding disparity compensated residual images in stereo image pairs. This paper presents a residual image coding scheme based on a recently proposed MP-based coding scheme called SOSU. The proposed scheme decomposes an orthogonal matching pursuit process into two stages. In different stages, dictionaries of different nature are used. Simulation results show that the proposed scheme can significantly reduce the computation effort while maintaining a similar rate-distortion performance as compared with SOSU.

1. INTRODUCTION

Stereo imaging is being studied actively due to its capability in providing stereoscopic pictures with high resolution and great sensation of reality[1]. As it carries much more information than a single image, it takes more data bits to represent it and hence an efficient encoding algorithm for its representation is necessary.

A stereo image pair contains a left image and a right image, and they are usually compressed in different ways. In general, one of them, say, the left image, is encoded with typical image coding techniques such as JPEG and then its encoded version is used as a reference image to encode the right image by utilizing the characteristics that there are very strong correlation between the two images.

The utilization of the correlation between the two images in coding a stereo image pair can be different in different approaches[2-7]. However, they are in common in a way as follows. In general, the image to be coded is estimated with the reference image. The estimation can be block-based[2-5] or region-based[6]. The estimation error of the image, which is termed as compensated residual image, is then coded with conventional image coding techniques.

In conventional stereo image coding algorithms, a compensated residual image is typically encoded with the discrete cosine transform (DCT) [1-3] due to its simplicity, low coding overhead requirements, and suitability for hardware implementation. However, this simple scheme suffers from several limitations such as blocking artifacts on the reconstructed images and poor compensation ability for the mismatched areas. Accordingly, Matching Pursuit (MP) has recently been applied to solve the above-mentioned problems and provide better compensation ability[4]. Matching Pursuit expands the disparity compensated residual image with an overcomplete dictionary of basis vectors. The superior performance is achieved at a huge cost of complexity.

In [7], Seo and Azimi-Sadjadi introduced a block-based coding scheme called SOSU to encode a disparity compensated residual image. Two major contributions were made in this scheme. First, it makes use of orthogonal matching pursuit (OMP) instead of MP to successively approximate the residual image with an overcomplete dictionary, which improves the coding efficiency. Second, an adaptive dictionary composed of an extended set of neighboring blocks of the best matching block in the reference picture and a set of typical edge blocks is used in the successive approximation. This allows a more flexible and accurate representation of the block of interest with a few basis vectors. However, the gain in coding performance is achieved at a cost of computation effort.

In this paper, based on SOSU, we proposed a Two-stage Orthogonal Matching Pursuit algorithm (TSOMP) to reduce the computation effort without sacrificing the coding performance. The paper is organized as follows. In Section 2, some background information of SOSU is provided. The proposed TSOMP scheme is then presented in Section 3. Section 4 provides some simulation results for performance evaluation. Finally, some concluding remarks are given in Section 5.

2. BACKGROUND

In SOSU, the right image is partitioned into a number of non-overlapping blocks of equal size. Each block is lexicographically ordered to form a column residual vector $\vec{e}_o$. This vector is then iteratively decomposed into a weighted sum of vectors selected from a dynamic dictionary Ω. Let $\vec{e}_{(k)}$ is the residual vector after the k^{th} iteration and $\vec{e}_{(0)}=\vec{e}_o$. At a particular iteration k, a basis vector in Ω, say $\vec{b}_{(k)}$, is selected as follows.

0-7803-7750-8/03/$17.00 ©2003 IEEE

$$\vec{b}_{(k)} = \arg\max_{\vec{b}_j \in \Omega} \left| \vec{e}^{\,t}_{(k-1)} \vec{b}_j \right| / \left\| \vec{b}_j \right\| \quad (1)$$

where $\vec{e}^{\,t}_{(k)}$ is the transpose of $\vec{e}_{(k)}$. The residual vector $\vec{e}_{(k)}$ and the dictionary are then updated with the selected $\vec{b}_{(k)}$. This selection process involves a number of inner products and repeats until a particular criterion is satisfied. Accordingly, it contributes significantly to the complexity of the scheme. Comparatively speaking, the realization effort paid for other activities is neglectable.

Obviously, the number of inner products involved in searching a basis vector in Ω is directly proportional to the number of elements in Ω. In SOSU, the dictionary is combined with two sets of basis vectors. One comes from an extended set of neighboring blocks of the best matching block in the reference picture and the other comes from a set of typical edge blocks[8] as shown in Fig.1. The nature of the two sets is completely different and their targets are different. If a block is better represented with an element of one of the sets, it will be unlikely that it can also be represented with an element coming from the other set. Hence, one can reduce the size of the dictionary by, at each iteration, selecting one of the sets as the dictionary to search the best basis vector for the current residual vector.

The selection cannot be achieved with a simple classification of the residual vector as both the residual vector and the dictionary change every iteration in OMP. Besides, it is difficult to find a simple criterion to classify a block into the corresponding classes. In the following Section, a modification to SOSU is made to achieve this goal.

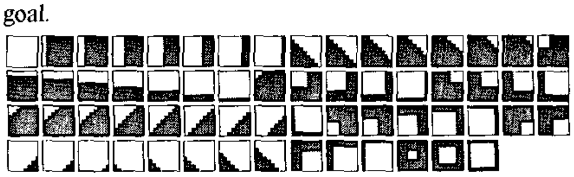
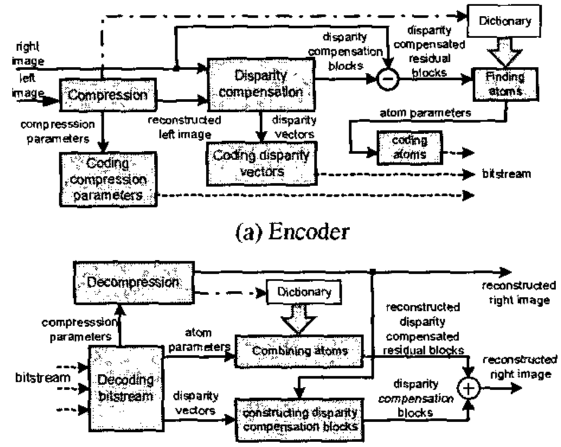


Fig.1 A set of typical edge blocks

3. TWO-STAGE OMP

Fig.2 shows the proposed Two-stage Orthogonal Matching Pursuit (TSOMP) stereo image coding system. The left image is first encoded with a typical image coding scheme such as JPEG and the reconstructed left image is used as the reference image for coding the right image. The right image is partitioned into non-overlapping blocks of equal size. For each block, a searching region is defined in the reference image and, from the searching region, the best-matching block of the block is located subject to the MSE criterion. The disparity vector of the block is then determined. The best-matching block is subtracted from the block of interest to produce a disparity compensated residual block.



(a) Encoder

(b) Decoder

Fig.2 A stereo image coding system using Two-stage Orthogonal Matching Pursuit (**TSOMP**)

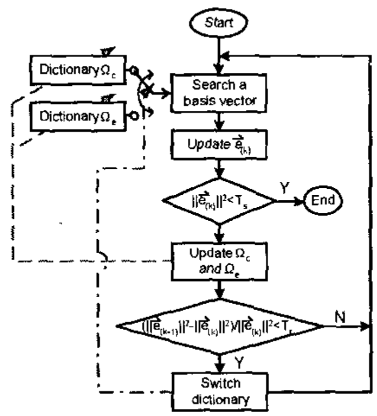


Fig.3 Flow diagram of the proposed Two-stage OMP scheme

The disparity compensated residual block is lexicographically ordered to form a column residual vector $\vec{e}_o$. This vector is then successively approximated with OMP. The dictionary Ω used for the approximation is composed of two sets of basis vectors and both of them are defined as in [7]. Let Ω_c be the set contains the basis vectors extracted from the support region inside the searching region and Ω_e be the set of lexicographically ordered edge blocks shown in Fig.1. Unlike SOSU, only one of them is used to search a basis vector at each iteration. Fig.3 shows the flow diagram of the proposed scheme.

At a particular iteration k, either Ω_c or Ω_e is selected to be Λ, the active dictionary for the iteration. A basis vector in Λ, say $\vec{b}_{(k)}$, is selected as follows.

$$\vec{b}_{(k)} = \arg\max_{\vec{b}_j \in \Lambda} \left| \vec{e}^{\,t}_{(k-1)} \vec{b}_j \right| / \left\| \vec{b}_j \right\| \quad (2)$$

where $\vec{e}_{(k)}$ is the residual vector after the k^{th} iteration and $\vec{e}_{(0)} = \vec{e}_o$. After $\vec{b}_{(k)}$ is determined, the residual vector is updated to be

$$\vec{e}_{(k)} = \vec{e}_{(k-1)} - \alpha_{(k)} \vec{b}_{(k)} \quad (3)$$

where

$$\alpha_{(k)} = \left| \vec{e}^{\,t}_{(k-1)} \vec{b}_{(k)} \right| / \left\| \vec{b}_{(k)} \right\|^2 \quad (4)$$

The basis vector $\vec{b}_{(k)}$ is then dropped and all elements remained in Ω_e and Ω_c are updated with

$$\vec{b}_i := \vec{b}_i - \left(\vec{b}_i^{\,t} \vec{b}_{(k)} / \left\| \vec{b}_{(k)} \right\|^2 \right) \vec{b}_{(k)} \qquad \forall \vec{b}_i \in \Omega_e \cup \Omega_c \quad (5)$$

At the end of each iteration, the relative marginal improvement in distortion, ΔI_r, which is defined as

$$\Delta I_r = \left(\| \vec{e}_{(k-1)} \|^2 - \| \vec{e}_{(k)} \|^2 \right) / \| \vec{e}_{(k-1)} \|^2 \quad (6)$$

is computed to judge if one should switch the dictionaries for next iteration. They are switched when ΔI_r is less than a predefined threshold T_r as it implies that the basis vectors in the active dictionary are not suitable for representing the residual vector any more.

The active dictionary for searching the first basis vector is selected to be Ω_e. This is based on our observation that, for typical stereo image pairs, the first basis vector obtained in the successive approximation usually comes from Ω_e when Ω is used in our simulations.

The successive approximation repeated until $\| \vec{e}_{(k)} \|^2 < T_s$ is satisfied, where T_s is a predefined threshold which controls the quality of the reconstructed image block. The disparity compensated vector $\vec{e}_o$ is then approximated to be

$$\hat{\vec{e}}_o = \sum_{k=1}^{N} \alpha_{(k)} \vec{b}_{(k)} \quad (7)$$

where N is the total number of selected basis vectors.

The index of $\vec{b}_{(k)}$ in Ω and its associated weight $\alpha_{(k)}$ are paired up to form the k^{th} atom parameter of $\hat{\vec{e}}_o$. All atom parameters and the disparity vector of the block are encoded with entropy coding. At the decoder, these parameters can be recovered and used to reconstruct the left reconstructed image with the reconstructed right image as shown in Fig.2b.

4. SIMULATION RESULTS

Simulation was carried out to evaluate the performance of the proposed coding system. The scheme proposed in [7] was also evaluated for comparison. In both cases, the dictionary is of size 126 where 64 is the size of Ω_e and 62 is the size of Ω_c. In the simulation, we set the search region for disparity estimation to be left margin 48, right margin 8, upper margin 3 and lower margin 3. Note that more margin was given to the left side in the left image since an object in the left image is typically shifted to the left side with respect to the right image. A support region of size 15×15 was defined as in [7] and the block size was chosen to be 8×8. Atoms were coded with the probabilistic model proposed in [9].

Fig.4 shows two stereo image pairs used in our simulation. *Room* is synthesized while the other is natural. A study was first made to investigate the probability of having an element in Ω_e as the first basis vector during the successive approximation of a disparity compensated residual block. In our simulation, we found that the percentage of blocks whose first basis vectors came from Ω_e were 72% for *Room* and 72% for *Hcktraip* respectively. This supports that Ω_e should be used as the active dictionary of the first iteration.

Fig.5 shows the rate-distortion performance of various schemes. The evaluation of the distortion performance was based on the peak signal to noise ratio (PSNR) of the reconstructed right image. In particular, it is defined as

$$PSNR = 20\log_{10}\left(255 / \| D - \hat{D} \|\right) \quad (8)$$

where D is the original right image and $\hat{D}$ is the reconstructed right image. Fig.6 shows the rate-complexity performance of various schemes. The complexity is measured in terms of number of inner products involved in encoding the right image. As a matter of fact, in the realization of the schemes, more than 97% of the computation effort was paid for computing inner products of vectors. Each curve in the Figures was obtained by varying the value of T_s when a particular scheme was simulated.

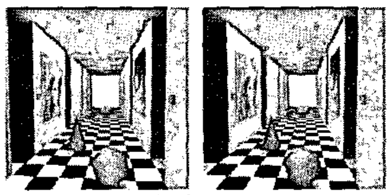
SOSU-IE and SOSU-I are, respectively, Seo and Azimi-Sadjadi's original scheme[7] and its variant where Ω_e is not included in Ω. TSOMP(x) is the proposed scheme, where x is the value of T_r used. By comparing SOSU-IE and SOSU-I, one can see that including Ω_e in Ω can significantly improve the rate-distortion performance at a cost of complexity. The rate-distortion performance of the proposed scheme is very close to that of SOSU-IE. However, the cost for the same improvement in rate-distortion performance is much lower than that required in SOSU-IE. As seen in Fig.5, the increase in the computation effort is only 60 to 70% of that required in SOSU-IE.

It was also found that the performance of the proposed scheme was not sensitive to T_r. However, as

Fig.5 and Fig.6 show that the best performance of the proposed scheme was obtained at T_r =0.1, this threshold value is suggested for the proposed scheme.

5. CONCLUSIONS

In this paper, a two-stage OMP algorithm for coding stereo image residuals is proposed based on a recently proposed stereo image coding scheme[7]. This algorithm can achieve a similar rate-distortion performance at a reduced cost of complexity as compared with [7]. In particular, around 30-40% of the computation effort required in [7] for achieving the gain in rate-distortion performance as compared with conventional coding algorithms can be reduced.



(a) *Room*

(b) *Hcktraip*

Fig.4 Testing stereo image pairs (Left, Right): (a) *Room*, (b) *Hcktraip* and (c) *Cliff*

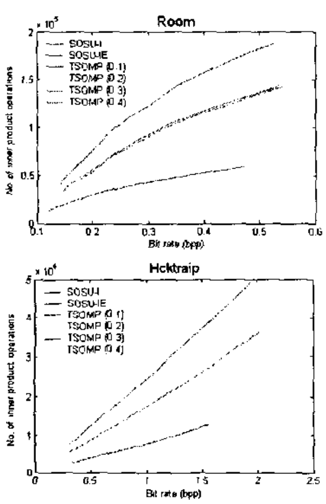


Fig.5 Rate-complexity performance of various schemes

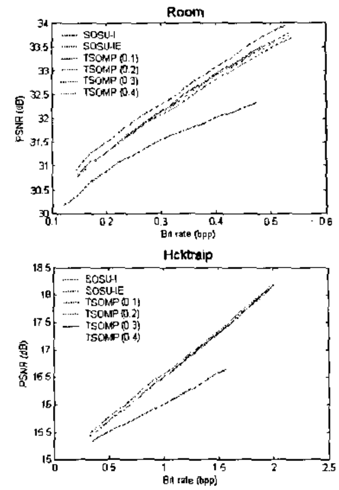


Fig.6 Rate-distortion performance of various schemes

6. REFERENCE

[1] T. Motoki, H. Isono, and I. Yuyama, "Present stage of three-dimensional television research," Proc. IEEE, vol.83, pp.1009-1021, Jul. 1995

[2] D. Tzovaras, M. G. Strintzis, and H. Sanhinoglou, "Evaluation of multiresolution block matching technique for motion and disparity estimation," Signal Processing: Image Commun., vol.6, pp.59-67, 1994

[3] M.S. Moellenhoff and. M.W. Maier, "Transform coding of stereo image residuals," IEEE trans. on Image Processing, Vol.7, No.6, pp.804-812, Jun 1998

[4] H. Aydinoglu and M. H. Hayes, "Stereo image coding: a projection approach," IEEE Trans. on Image Processing, Vol.7, No.4, pp.506-516, Apr 1998

[5] W.Woo and A.Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," IEEE trans. on CASVT, Vol.10, No.2, pp.194-200. Mar 2000

[6] S. Sethuraman, A.G. Jordan and M.W. Siegel, "A multiresolutional region based segmentation scheme for stereoscopic image compression," SPIE, Vol.2419, pp.265-274,1995

[7] S.H. Seo and M.R. Azimi-Sadjadi, "A 2-D filtering scheme for stereo image compression using sequential orthogonal subspace updating," IEEE trans. on CASVT, Vol.11, No.1, pp.52-66, Jan 2001

[8] B. Ramamurthi and A. Gersho, "Classified vector quantization of images," IEEE trans. on Communication, Vol.COM-34, No.11, pp.1105-1115, Nov 1986

[9] M. Gharavi-Akikhanasari, "A model for entropy coding in matching pursuit," Proceedings, IEEE ICIP'98, Vol.1, pp.778-782, 1998