ABSTRACT

Many research works have been developed for stereo image compression purpose by focusing on the disparity compensation technique. For this reason, a great attention should be paid to the generation of the disparity-compensated residual image. Generally, the residual image is computed through a simple subtraction of the disparity-compensated reference image from the target one. In this paper, we investigate two techniques for optimizing the computation of the residual image. The obtained results confirm the benefits of these optimization approaches in the context of stereo image coding.

Index Terms— Stereo image, compression, disparity compensation, optimization, \( \ell_2 \) and \( \ell_1 \) minimization.

1. INTRODUCTION

Stereoscopic imaging has gained a growing interest in many emerging applications such as 3D TV, 3D Digital cinema, immersive games and videoconferencing. The main advantage of stereoscopic technology is that it enhances the depth perception and makes the 3D experience more vivid. This technology requires the acquisition of two images, called left and right images, by recording two slightly different views angles of the same scene. Thus, adding the third dimension to the viewer/player leads to the doubling of the image data size compared to the monoscopic case, and consequently, involves a large amounts of stereo data. Therefore, it becomes necessary to design efficient stereo compression techniques for storing and transmitting purpose.

To this end, the simplest way for compressing stereo image (SI) is the independent coding scheme where the left and right images are encoded separately by using existing still image coders. However, since these images result from the projection of the same 3D scene and so present similar visual contents, it has been shown that more efficient joint stereo image compression scheme. Most of the reported methods apply a simple subtraction between each pixel of the target image and its homologous one in the reference image. In order to better exploit the inter-view dependencies, we propose in this paper to use the neighborhood of the homologous pixel to predict the pixel of the target image. To this end, we investigate two minimization techniques, based on the use of \( \ell_2 \) and \( \ell_1 \) criteria, for optimizing the generation of the residual image. Note that the benefits of such optimization criteria have been recently shown in the context of the design of lifting operators in a wavelet-based coding scheme [7].

The rest of this paper is organized as follows. In Section 2, the disparity compensation technique for stereo image coding is presented. The developed optimization techniques for generating the residual image are described in Section 3. In Section 4, experimental results are given and conclusions are drawn in Section 5.

2. DISPARITY COMPENSATED RESIDUAL CODING

Stereo matching algorithms consist of assigning to each pixel of the target image \( I^{(r)} \) a displacement value that allows to find its homologous point in the reference image \( I^{(l)} \). Such problem has been extensively studied in computer vision, and a review of disparity estimation techniques can be found in
In the context of stereo image coding, block-matching algorithms are often used because all pixels belonging to the same block can be represented by only one value, and so requires a few bits to encode the resulting disparity map. More precisely, after partitioning the target image into non-overlapping blocks, the disparity value associated to each block is obtained by minimizing a similarity criterion \( D \) as follows:

\[
d = \arg \min_{d_0 \in \{1, \ldots, d_{\text{max}}\}} D(I^{(r)}(i, j), I^{(l)}(i + d_0, j)),
\]

where \((i, j)\) are the spatial coordinates of the top left pixel in the block, and \(d_{\text{max}}\) represents the potential maximum disparity value. In general, the similarity criterion is the sum of square difference (SSD) or the sum of absolute difference (SAD).

Once the disparity map is estimated, the target image \( I^{(r)} \) can be predicted by shifting the reference image \( I^{(l)} \) along the horizontal direction thanks to the disparity information. Then, the residual image \( I^{(e)} \) is generated as follows:

\[
I^{(e)}(i, j) = I^{(r)}(i, j) - I^{(l)}(i + d(i, j), j),
\]

Thus, the reference image, the residual one and the disparity map are encoded and transmitted to the decoder side. Note that the reconstruction of the stereo images from these data is straightforward. Indeed, the reference and residual images as well as the disparity map are firstly decoded. Then, the target image is predicted using the decoded reference image and the disparity information. Finally, the target image is re-constructed by adding the decoded residual image to the predicted target one.

3. OPTIMIZED RESIDUAL IMAGE

3.1. Motivation and computation strategy

Although most of the reported stereo image coding methods generate the residual image according to Eq. (2), it is worth pointing out that this computation strategy may be suboptimal. Indeed, some authors have proposed to weight the compensation term to take into account the illumination variation between the two views [9]:

\[
I^{(e)}(i, j) = I^{(r)}(i, j) - \alpha I^{(l)}(i + d(i, j), j),
\]

where \(\alpha \in \mathbb{R}_+\).

Most importantly, the homologous pixel \( I^{(l)}(i + d(i, j), j) \) may not appear exactly on the epipolar line due to the presence of noise, the numerical rectification error, and the deviation from the pinhole camera model. Therefore, during the subtraction (i.e prediction) operation, instead of only taking the homologous pixel \( I^{(l)}(i + d(i, j), j) \), it becomes more interesting to use also its neighboring samples. More precisely, we will consider the following equation for computing the residual image:

\[
I^{(e)}(i, j) = I^{(r)}(i, j) - p^T I^{(e)}(i, j),
\]

where

\[
I^{(e)}(i, j) = \left( I^{(l)}(i + d(i, j) + m, j + n) \right)_{-M/2 \leq m \leq M/2}^{-N/2 \leq n \leq N/2}
\]

is the vector, of dimension \((M + 1) \times (N + 1)\), containing the pixels of the left image used in the prediction step, and \(p \) is the prediction vector.

Once the computation strategy of the residual image is described, we focus now on the optimal design of the vector \( p \). The objective is to produce a compact representation of the residual image by minimizing a given cost functional \( J \):

\[
J(p) = \sum_{i=1}^{W} \sum_{j=1}^{H} \left( I^{(r)}(i, j) - p^T I^{(e)}(i, j) \right)^2
\]

where \(W\) and \(H\) are the width and the height of the stereo images.

By minimizing this criterion, it can be checked that the optimal prediction vector \( p^{\text{opt}} \) must satisfy the well-known Yule-Walker equations:

\[
E[I^{(e)}(i, j) I^{(e)}(i, j)^T] p^{\text{opt}} = E[I^{(r)}(i, j) I^{(e)}(i, j)].
\]

where \(E[\cdot] \) denotes the mathematical expectation.

3.3. \( \ell_1 \) optimization technique

Sparse representations have attracted a great deal of attention in many application fields during the last years such as compressive sensing, image deblurring and image compression [7, 10]. Therefore, with the ultimate aim of increasing the sparsity of the residual coefficients, we propose to study also an \( \ell_1 \) optimization technique for designing the optimal prediction vector. More precisely, the objective consists in minimizing the following \( \ell_1 \) criterion:

\[
J(p) = \sum_{i=1}^{W} \sum_{j=1}^{H} |I^{(r)}(i, j) - p^T I^{(e)}(i, j)|.
\]
It should be noted that the sparsest residual coefficients could be obtained by minimizing an $\ell_0$ criterion. However, such a problem is inherently non convex and NP-hard. For this reason, we have focused on the use of $\ell_1$ criterion which is convex and can be efficiently solved thanks to a class of proximal optimization algorithms [11]. Among them, we adopt the Douglas-Rachford algorithm which was found to be simple and effective for this problem [12].

Before describing the algorithm, let us first address the necessary backgrounds on convex analysis and proximity operators [11].

3.3.1. Background on convex optimization tools

We denote by $\mathbb{R}^K$ the $K$-dimensional Euclidean space with norm $\| \cdot \|$. The main definitions which will be used in this work are the followings:

- The indicator function of a convex set $C \subset \mathbb{R}^K$ is given by:
  \[
  \forall x \in \mathbb{R}^K, \quad \iota_C(x) = \begin{cases} 
  0 & \text{if } x \in C, \\
  +\infty & \text{otherwise}.
  \end{cases}
  \]

- The distance function to a nonempty set $C \subset \mathbb{R}^K$ is defined by:
  \[
  \forall x \in \mathbb{R}^K, \quad d_C(x) = \inf_{y \in C} \|x - y\|.
  \]

- The projection of $x \in \mathbb{R}^K$ onto a nonempty closed convex set $C \subset \mathbb{R}^K$ is the unique point $P_C(x) \in C$ such that $d_C(x) = \|x - P_C(x)\|$.

- $\Gamma_0(\mathbb{R}^K)$ is the class of lower semicontinuous convex function from $\mathbb{R}^K$ to $]-\infty, +\infty]$ and not identically equal to $+\infty$.

- The proximity operator of a function $f \in \Gamma_0(\mathbb{R}^K)$, denoted by prox$_f$, is defined as follows:
  \[
  \text{prox}_f : \mathbb{R}^K \to \mathbb{R}^K
  \]
  \[
  x \mapsto \arg \min_{y \in \mathbb{R}^K} f(y) + \frac{1}{2}\|x - y\|^2.
  \]

3.3.2. Douglas-Rachford algorithm

In the following, the algorithm adopted to generate the residual image is described. We recall that the aim consists of minimizing the $\ell_1$-norm of the difference between the current pixel of the right image $I^{(r)}(i, j)$ and its predicted value. In this context, the vector $\mathbf{I}^{(r)} = (I^{(r)}(i, j))_{1 \leq i \leq W, 1 \leq j \leq H}$ can be seen as an element of the Euclidean space $\mathbb{R}^{W \times H}$. The minimization problem (9) can be rewritten as follows:

\[
\min_{x \in V} \sum_{i=1}^W \sum_{j=1}^H |I^{(r)}(i, j) - X(i, j)|,
\]

where $V = \{X = (X(i, j))_{1 \leq i \leq W, 1 \leq j \leq H} \mid \exists p \in \mathbb{R}^L, \forall (i, j) \in \{1, \ldots, W\} \times \{1, \ldots, H\}, X(i, j) = p^T \mathbf{1}^{(c)}(i, j)\}$.

According to the definition of the indicator function (10), the minimization problem (13) is equivalent to the following one:

\[
\min_{x \in \mathbb{R}^{W \times H}} f_1(X) + f_2(X),
\]

where

\[
f_1(X) = \|I^{(r)} - X\|_{\ell_1}
= \sum_{i=1}^W \sum_{j=1}^H |I^{(r)}(i, j) - X(i, j)|,
\]

and

\[
f_2(X) = \iota_C(X).
\]

By writing it under this form, the minimization problem (14) can be solved by using the Douglas-Rachford algorithm. The obtained numerical solution can be obtained based on the following iterative algorithm:

1. Fix $S_0 \in \mathbb{R}^{W \times H}$, $\gamma > 0$, $\epsilon \in ]0, 1]$ and $\lambda \in [\epsilon, 2 - \epsilon]$
2. for $k = 0, 1, \ldots$ do
   \[
   \begin{align*}
   X_k &= \text{prox}_{\gamma f_1} S_k, \\
   S_{k+1} &= S_k + \lambda (\text{prox}_{\gamma f_2}(2X_k - S_k) - X_k).
   \end{align*}
   \]
3. end

Algorithm 1: Douglas-Rachford algorithm.

As it can be seen in this algorithm, each iteration requires the computation of two proximity operators for the functions $f_1$ and $f_2$. Note that closed-form expressions of the proximity operator of some functions belonging to $\Gamma_0(\mathbb{R}^K)$ are developed in [11].

In our case, the proximity operator of the function $\gamma f_1$ is given by:

\[
\forall S_k \in \mathbb{R}^{W \times H}, \quad \text{prox}_{\gamma f_1}(S_k) = (\pi_k(i, j))_{1 \leq i \leq W, 1 \leq j \leq H},
\]

where \(\forall (i, j) \in \{1, \ldots, W\} \times \{1, \ldots, H\}\),

\[
\pi_k(i, j) = \text{soft}_{[-\lambda, \lambda]}(S_k(i, j) - I^{(r)}(i, j)) + I^{(r)}(i, j)
\]

and \(\forall t \in \mathbb{R}\),

\[
\text{soft}_{[-\lambda, \lambda]}(t) = \begin{cases} 
  \text{sign}(t)(|t| - \gamma) & \text{if } |t| > \gamma \\
  0 & \text{otherwise}.
  \end{cases}
\]
Concerning $\gamma f_2$, the proximity operator is given by:

$$\forall S_k \in \mathbb{R}^{W \times H}, \quad \text{pro}_{\gamma f_2}(S_k) = P_{\gamma f_2}(S_k) = \left(p^T I^{(c)}(i,j)\right)_{1 \leq i \leq W, \ 1 \leq j \leq H}$$

where

$$p = \left(\sum_{i,j} I^{(c)}(i,j)I^{(c)}(i,j)^T\right)^{-1} \sum_{i,j} S_k(i,j)I^{(c)}(i,j).$$

Finally, we should note that it has been shown in [13] that every sequence $X_k$ generated by the Douglas-Rachford algorithm converges to a solution to problem (14) if the selected parameters $\lambda$ and $\gamma$ satisfy the conditions given in Algorithm 1.

### 4. EXPERIMENTAL RESULTS

The simulations are performed on different standard stereo images downloaded from some public stereovision datasets $^1$,$^2$ and $^3$. A block-matching technique with a $8 \times 8$ block size is employed to estimate firstly the residual image. The latter is then losslessly encoded by using a DPCM technique followed by an arithmetic coder. Finally, to encode the reference image as well as the residual image, the 9/7 (resp. 5/3) wavelet transform, retained for the lossy (resp. lossless) compression mode of JPEG2000 [14], is applied over three resolution levels. The resulting wavelet coefficients are then encoded by using the entropy coder EBCOT.

In order to study the proposed optimization strategies, we focus on the generation of the residual image by considering the three following experiments:

1. The first one uses the standard computation procedure defined by Eq. (2). We refer to this method by “Standard”.

2. The second one consists of applying a weighting term as given by Eq. (3). The weight is computed by minimizing the variance of the residual coefficients. This method will be denoted by “Weighted-version”.

3. The third and the fourth ones correspond to the proposed optimization strategies. More precisely, according to the notations used in Eq. (5), the pixels taken during the disparity compensation process are obtained by setting $m \in \{-1,0,1\}$ and $n \in \{-1,0,1\}$. The optimization based on the $\ell_2$ and $\ell_1$ criteria will be designated by “Proposed-$\ell_2$-OPT” and “Proposed-$\ell_1$-OPT”, respectively.

In addition to these methods, we have also considered the independent coding scheme, denoted by “Independent”, where the retained wavelet transform is applied separately to the left and right images.

First, the performance of these different methods will be evaluated in the context of lossless coding mode. Since the reference image is encoded in the same way for all the tested methods, we give in Table 1 the entropy of the multiresolution representations of the target image. The advantages of such measure are that it is easily computed and it is independent of the performance of any embedded coder. Compared to the standard approach, the proposed optimized computation strategy of the residual image results in a gain which can reach 0.25 bits per pixel (bpp). It can be also observed that the $\ell_1$ optimization technique achieves a further improvement compared to the $\ell_2$ optimization approach.

<table>
<thead>
<tr>
<th>Method</th>
<th>Shrub</th>
<th>Art</th>
<th>Cones</th>
<th>Drumsticks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>4.95</td>
<td>4.38</td>
<td>5.39</td>
<td>4.49</td>
</tr>
<tr>
<td>Standard</td>
<td>3.73</td>
<td>3.51</td>
<td>4.27</td>
<td>3.88</td>
</tr>
<tr>
<td>Weighted-version</td>
<td>3.70</td>
<td>3.51</td>
<td>4.26</td>
<td>3.87</td>
</tr>
<tr>
<td>Proposed-$\ell_2$-OPT</td>
<td>3.39</td>
<td>3.51</td>
<td>4.22</td>
<td>3.88</td>
</tr>
<tr>
<td>Proposed-$\ell_1$-OPT</td>
<td>3.37</td>
<td>3.46</td>
<td>4.19</td>
<td>3.83</td>
</tr>
</tbody>
</table>

Moreover, these methods have also been evaluated in the context of lossy coding mode. Note that a closed loop-based coding structure has been employed [5]. The reported results are given in terms of the average bitrate $R_{av}$ and its corresponding PSNR measure:

$$R_{av} = \frac{R^{(l)} + R^{(e)} + R^{(d)}}{2},$$

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{(\text{MSE}^{(l)} + \text{MSE}^{(r)})/2},$$

where $R^{(l)}$, $R^{(e)}$ and $R^{(d)}$ denote respectively the bitrate of the left, target, and disparity images. $\text{MSE}^{(l)}$ and $\text{MSE}^{(r)}$ correspond respectively to the mean squared error of the reconstructed left and right images. Figure 1 illustrates the performance of the different considered stereo images compression methods for the stereo image “Shrub” and “Houseof”. It can be observed that the proposed strategy outperforms the standard computation procedure by about 0.3-1 dB. Furthermore, we have also compared the $\ell_1$-based optimization technique to the standard one in terms of Bjontegaard metric, often used to measure the distance between two R-D curves [15]. The PSNR differences and the bitrate saving between these two approaches are provided in Table 2 for low, middle and high bitrates corresponding to the four target bitrate points \{0.15, 0.2, 0.3, 0.4\}, \{0.5, 0.6, 0.7, 0.8\} and \{0.9, 1, 1.1, 1.2\} bpp, respectively. This table shows that the average of the PSNR differences between the R-D results varies between 0.14-1.2 dB.

Finally, we illustrate in Fig. 2 the reconstructed “Pentagon” stereo image. The reconstruction quality is evaluated in

---

1. http://vision.middlebury.edu/stereo/
Fig. 1: PSNR (in dB) versus the bitrate (bpp) after JPEG 2000 encoding for the stereo images “Shrub” and “Houseof”.

terms of PSNR. As it can be seen in Fig. 2, the proposed \( \ell_1 \) optimization strategy leads to better reconstruction quality compared to the standard approach.

All these results confirm the benefits which can be drawn from the proposed techniques for optimizing the residual image.

5. CONCLUSION

In this work, we have focused on the optimization of the residual image to improve the stereo coding performance. More precisely, for each pixel in the target image, we have proposed to use the homologous pixel in the reference image as well its neighboring to generate the residual image. To this end, two optimization techniques including the \( \ell_2 \) and \( \ell_1 \) criteria have been investigated. Experimental results show the inter-

Table 2: The average PSNR differences and the bitrate saving at low, middle and high bitrates for different Stereo images. The gain of the Proposed \( \ell_1 \)-OPT method w.r.t to the standard one.

<table>
<thead>
<tr>
<th>SI</th>
<th>PSNR gain (dB)</th>
<th>Bitrate saving (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low</td>
<td>middle</td>
</tr>
<tr>
<td>Shrub</td>
<td>0.60</td>
<td>0.91</td>
</tr>
<tr>
<td>Houseof</td>
<td>0.34</td>
<td>0.59</td>
</tr>
<tr>
<td>Pentagon</td>
<td>0.52</td>
<td>0.91</td>
</tr>
<tr>
<td>Art</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>Cones</td>
<td>0.14</td>
<td>0.27</td>
</tr>
<tr>
<td>Average</td>
<td>0.37</td>
<td>0.58</td>
</tr>
</tbody>
</table>

6. REFERENCES


Fig. 2: Reconstructed “Pentagon” stereo image coded at $R_{av}=0.23$ bpp, (a) Standard and (b) Proposed-$\ell_1$-OPT.


