Adaptive Compression of Stereoscopic Images

Alessandro Ortis[1], Francesco Rundo[2],
Giuseppe Di Giore[2], and Sebastiano Battiato[1]

[1] Università degli Studi di Catania - Dipartimento di Matematica ed Informatica
Viale A. Doria 6, 95125 – Catania, Italy
`alessandro.ortis@gmail.com,`
`battiato@dmi.unict.it`
[2] STMicroelectronics s.r.l. - Digital Convergence Group/CSP
Stradale Primosole 50, 95121 – Catania
`{francesco.rundo,giuseppe.di-giore}@st.com`

Abstract. Nowadays the growing availability of stereo cameras for common applications is becoming a commodity. This paper addresses the problem of stereoscopic images data compression proposing an innovative algorithm for compressing Multi Picture Object coded stereopairs. By means of self organizing reconstruction algorithm based on image redundancy we are able to reduce the size of the enclosed JPEG images. The overall perceived (and measured) quality is managed by considering that a stereoscopic image represents the same scene acquired from two different perspectives. In particular we achieve some compression gain just encoding the two images with different quality factors. The reported results and test benchmarks show the robustness and efficiency of the proposed algorithm.

1   Introduction

Stereoscopy (also called stereoscopics or 3D imaging) is a technique for representation and projection of images and videos for creating the illusion of depth in order to simulate the human binocular vision. The binocular vision is based on the principle that we can present two slightly different images separately to the left and the right eye whereas such images are then combined by the viewer's brain to give the perception of 3D vision. By comparing these two images the HVS (Human Vision System) is able to infer depth distances in the scene. Using this principle, the images are usually projected on the same screen with different polarization in order to present a different image to each eye. The viewer wears eyeglasses which contain a pair of opposite polarizing filters, each filter blocks the opposite polarized light and each eye sees only one image [1][2].

Multi Picture Object (MPO) is a file format specification to store multi-picture images [3]. The included object images are constituted by a chain of still JPEGs [4] merged together in a single file with specific tags allowing the images to be handled.

Fig. 1. Left and right shot taken as stored in an MPO file

One of the most used scenario is the so called "stereoscopic image representations". In that scenario, MPO tags contain both information about single images (independent JPEG images) and data about the global position of each image in the real environment. 3D viewers (e.g., 3D set top box, TV, etc.) make use of such peculiarities to properly render the stereoscopic image.

A stereoscopic image is composed by two images (left and right ones) of the same subject (called target object) captured from two different perspectives by using two cameras (or a 3D camera) where the reciprocal distance corresponds to the interocular distance (Figure1). The distance and the angle between these perspectives are called *baseline length* and *convergence angle* respectively and their measures are established in order to reproduce artificially the human binocular vision [3].

The needed storage for each MPO file is approximately the space needed to store the embedded JPEGs, so that, taking into account that the involved scene is the same for both images, we have a lot of common and redundant information.
In literature there are no other similar approaches in terms of adaptive coding denoted to attack such redundancy in MPO compliant way. Perkins's approach [5] requires a transmission format which could not replace established image formats. Unlike Perkins's approach, methods for achieve an improvement without departing from JPEG encoding are available [6][7]. In [8] a joint decoding scheme is proposed in order to enhance the quality of the reconstructed image pairs independently compressed with JPEG. But some regions of the processed image cannot possibly be reconstructed by this method and, with a middle quality of JPEG compression, there are PSNR [9] decrease and some ghosting artifacts appear seldom [8].

In the proposed method, during encoding, one of the two images is JPEG compressed with a low quality rate [4]. Then, in decoding process, we apply a block based reconstruction algorithm to enhance its quality by using the redundant information of the high level encoded image (where some block matching is exploited). If no reliable matching is achieved, i.e. the difference between the images is high, the algorithm

Fig. 2. An example of MPO file that contains two stereoscopic images. Differences between left and right images are located in the margin, in the left margin of the right image are shown some objects such a window with a tree in the background, a lamp and some details of a table which are not present in the left image.
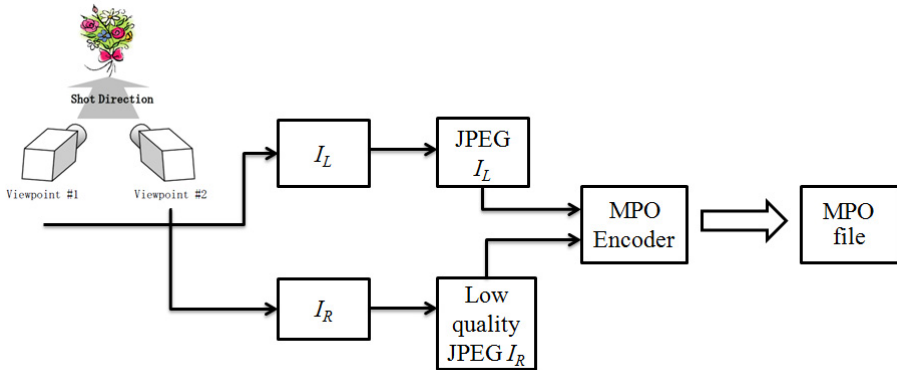

Fig. 3. Modified encoding pipeline

uses the low quality image information. This causes a losing rate which is not perceptible by the human vision system.

Experiments over a significant number of MPO images [10] confirm the effectiveness of the proposed approach. In addition, as further suggested in [11] our HVS can tolerate a slight amount of asymmetric image quality for stereo viewing (i.e. the perceived quality lies between that of the two views) thus the JPEG factor of the low quality image could be reduced even more than we've described allowing to obtain better results both in terms of bitrate saving and PSNR.

The paper is organized as follows. Section 2 describes the proposed algorithm, in Section 3 the results in terms of measured quality and bitrate saving are discussed, while a conclusion section gives direction for future works.

2   Proposed Method

In stereoscopic pair images the target object and all its main visual details appears in both cases, only a few differences are evident on the border (left and right), as shown in Figure 2. Figure 3 reports the encoding pipeline of the proposed approach where we split the processing according to two different encoding settings (Figure 3). As shown in Figure 3, we conventionally use the left image (denoted by I_L) as the high quality image and the right one (denoted by I_R) as the low quality image. Each block (of each channel) of the compressed image I_R is enhanced by using the image I_L as a reference; the same approach could be used inverting the role of the two involved images. This is a clear evidence of the robustness of the proposed method as it works without specific hypothesis related to the source images included in the original stereoscopic MPO. We limit the algorithm description to luminance plane only, but the same processing is applied also to the chromatic components.

During decoding, both images are subdivided in not overlapping blocks of equal dimensions ($NxMx3$ with $N<W$ and $M<H$ where W and H are the dimensions of both images), the proposed scheme works block by block on each channel so it operates on blocks of NxM samples for each channel plane.

For the reconstruction of the generic i^{th} block of I_R, called b^R_i, the first step consists in finding the b^L_i block of I_L which best approximates the considered b^R_i block of I_R; to this purpose we consider two candidate blocks from the reference image: the first one is obtained by using a cross-correlation based template matching method and the other one is taken considering the b^L_i block which has the same position of b^R_i. This task is called *"Match similar"*. The first candidate block considered by *Match similar* is denoted by Y^{corr} and is obtained using a well known template matching method based on the computation of the normalized cross correlation (NCC) [12].

A common way to calculate the position of the pattern template t in the image f is to evaluate the NCC value $\gamma(u,v)$ at each point (u,v) for f and t, which has been shifted by u and v steps in the x and y direction respectively. The following equation gives a definition for the NCC coefficients:

$$\gamma(u,v) = \frac{\sum_{x,y}\left\{\left[f(x,y) - \bar{f}_{u,v}\right]\left[t(x-u, y-v) - \bar{t}\right]\right\}}{\left\{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]^2 \sum_{x,y}\left[t(x-u, y-v) - \bar{t}\right]^2\right\}^{0.5}} \tag{1}$$

where $\bar{f}_{u,v}$ denotes the mean value of $f(x,y)$ within the area of template t shifted to (u,v) and $\bar{t}$ is the mean value of the template t. The normalization makes coefficients independent to changes in brightness or contrast of the image [12]. To reduce the computational cost of NCC we used an optimization approach which calculate the NCC using a subimage of the high quality image as search range instead of using all the image as reference image, this optimization permitted us to reduce the time required for reconstruction but it doesn't affected the results in terms of quality.
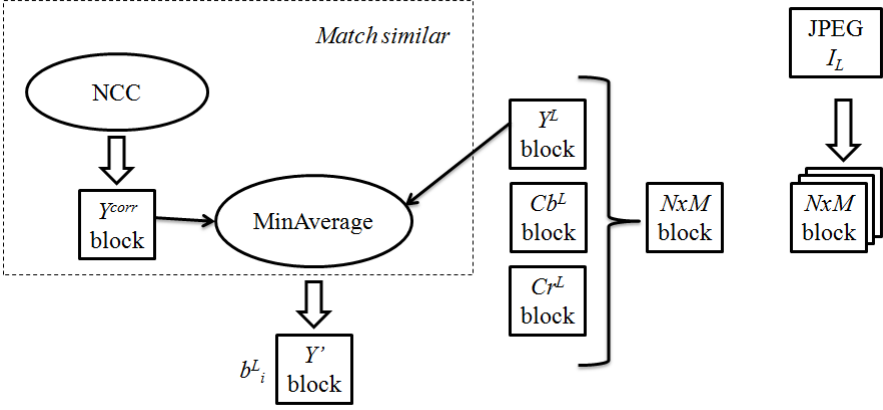
Fig. 4. Schematic view of the *Match similar* function where Y^{corr} and Y^L denotes the first and the second candidate block respectively, this function gets the candidate which minimize the average difference with respect to b^R_i.

The second candidate block, denoted by Y^L, is the *NxM* block in the luminance channel of I_L which is located in the same position of b^R_i. Figure 4 illustrates how *Match similar* works: for each candidate we calculate the average of the differences between candidate samples with respect to b^R_i samples (in the considered channel) with the following equations:

$$avg_{Y^{corr}} = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| Y^{corr}_{i,j} - Y_{i,j} \right| \tag{2}$$

$$avg_{Y^L} = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| Y^L_{i,j} - Y_{i,j} \right| \tag{3}$$

Match similar gets the candidate block which minimize this value (denoted by *Y'*).

The second step of the reconstruction consists of enhancing the b^R_i block using the redundant information between the luminance block of b^R_i and the luminance block of b^L_i (i.e., the block *Y'* returned by *Match similar*).

This function is based on a simplified version of Kohonen reconstruction [13]. In fact, in this step we "pull" the value of some samples from b^R_i to b^L_i depending of the similarity between these two samples.

Our reconstruction uses the following equation:

$$\bar{b}^R_i(u,v) = \begin{cases} b^R_i(u,v) + \alpha \cdot \left(b^R_i(u,v) - b^L_i(u,v) \right) & if \left(b^R_i(u,v) - b^L_i(u,v) \right) < limit \\ b^R_i(u,v) & otherwise \end{cases} \tag{4}$$

where $\overline{b}_i^{R}(u,v)$ is the enhanced sample, $b_i^{L}(u,v)$ is the sample of Y' block chosen by *Match similar* and $b_i^{R}(u,v)$ is the sample which has to be enhanced (taken from Y block of b^{R}_i). The constants α and *limit* (in our cases $\alpha = 0,25$ and *limit* $= 0,043$) are two coefficients which determine the amount of "pulling" and the amount of "similarity" of the Kohonen reconstruction respectively [13]. This function works sample by sample and returns the enhanced block $\overline{Y}$ (Figure 5).
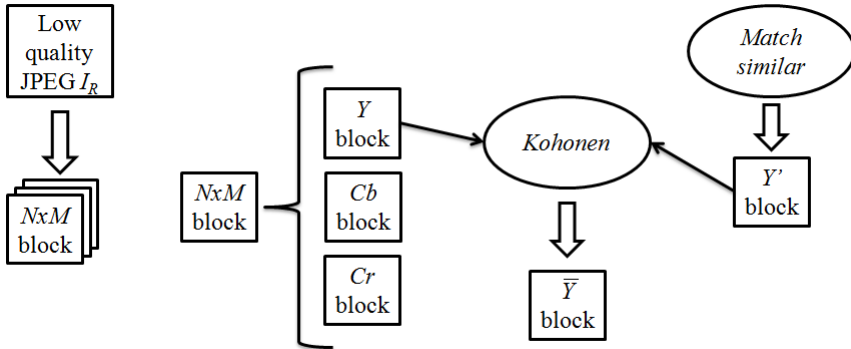

Fig. 5. $\overline{Y}$ block is the result of the reconstruction obtained enhancing Y through Kohonen reconstruction of Y using Y' as reference

3   Experimental Results

To evaluate the performances of the proposed MPO pipeline we have used for our experiments, 23 stereo MPO images conform with [3] specifications taken from [10] at various resolution sizes (1440x1080, 1444x1080, 1620x1080, 1920x1080 or 1924x1080). In our settings JPEG quality factor has been established to 85 for the high quality image and 65 (or 70) for the low quality image. Using a quality compression factor less than 65 causes an excessive degradation, viceversa using a quality compression factor greater than 70 involves a slight of bitrate saving. Reported results show compression gain in terms of total bitrate while the quality is measured considering the PSNR and the SSIM [14]. Table 1 list also the parameters used in our tests (e.g., the block size). For each MPO image and for each value of JPEG compression factor used to compress the low quality image (65 or 70) the following parameters are reported: file name, dimensions of the blocks used in decoding, the lossy (in terms of PSNR dB) measured after reconstruction and the rate of saving with respect to the same image codified using standard as high quality factor equal to 85.

Table 1. Comparison of the parameters and results of tests using 65 and 70 as JPEG quality factor for the low quality image

MPO image	NxM	Low quality 65		Low quality 70	
		Lossy (dB)	Bitrate saving	Lossy (dB)	Bitrate saving
Flowers1	360x360	2,17	40,7%	1,65	34,6%
Flowers2	360x481	2	40,6%	1,32	34,5%
Flowers3	360x481	2,76	48,7%	2,76	41,7%
Castle	360x481	2,62	38,3%	2,18	32,5%
Dorm	360x360	2,73	37,1%	2,64	31,1%
Pelion	360x481	2,34	37,6%	2,34	31,8%
Hallway	360x482	2,33	37,6%	2,33	32,1%
Statue	360x483	2,59	41,9%	2,60	35,7%
Library	360x270	1,98	38,7%	1,71	32,7%
Hall	360x360	1,66	41,1%	1,40	34,9%
Garden bridge	360x360	2,11	39,5%	1,82	33,5%
Autumn1	360x361	2,73	35,3%	2,53	29,8%
Autumn2	360x361	2,6	36,4%	2,40	30,6%
Autumn3	360x361	2,38	37,0%	2,15	31,2%
Autumn4	360x361	2,65	36,1%	2,44	30,4%
Animals1	360x240	2,16	38,8%	2,16	32,8%
Animals2	360x240	2,47	37,0%	2,18	31,3%
Cube	360x360	2,33	39,3%	2,00	33,3%
Covered	360x360	1,88	39,2%	1,73	33,4%
Garden	360x360	2,41	38,5%	2,15	32,5%
Snow	360x481	2,62	36,8%	2,45	31,2%
Tree	360x360	2,69	37,4%	2,52	31,4%
Zoo	360x240	2,67	36,9%	2,33	31,1%

Table 2. Comparison between overall results obtained using 65 or 70 as JPEG quality factor

Low quality	Average lossy (dB)	Average saving
65	2.39	38,7%
70	1.16	32,8%

   Although the mean values reported in Table 2 are similar, the performances change for each image and the difference between using 65 than using 70 as JPEG quality factor appears more evident comparing this variability (see Table 1).

Compressing using 65 the bitrate saving is between 35% and 48,7% and the lossy is between 1,66 dB and 2,76 dB. Compressing using 70 the bitrate saving is between 30% and 42% and the lossy is between 1,32 dB and 2,76 dB. The visual degradation of the reconstructed images are so low that the SSIM values are all close to 1 (average value 0,93), therefore we considered the PSNR as the most significant quality metric.

4   Conclusion and Future Works

We have presented a reconstruction algorithm applied on a previous degraded image which use the redundant information between stereo pairs in order to reduce the bitrate of one view of a stereo image. The results show how this approach permits to reduce the size of MPOs with low visual degradation.

The proposed algorithm provides effective results both in terms of viewing and PSNR values, the only one disadvantage of this scheme is the computational cost caused by NCC. However other more powerful approaches [12] could be used instead of NCC in order to reduce the runtime of the reconstruction algorithm.

Furthermore, it would be interesting to evaluate the performance of the proposed algorithm by means of 3D quality metrics instead of PSNR or other 2D quality metrics, but very little research has been done to design a quality metric for 2-view case scenario. The HV3D metric [15] is a 3D video quality metric which takes into account the depth effect and the binocular properties of the HVS.

Future works will be focused on the improvement of the performances; this could be obtained using a different blocking approach or adding to the JPEG coding of the low quality image a previous compression scheme based on a different coding or a motion estimation compression based on the spatial parameters of the acquisition device or both. Moreover when the image has small fluctuation in depth it would be interesting to evaluate alternative linear template matching methods based on the asymmetry of the distortion. The encoding pipeline proposed by this scheme should be implemented directly in the digital cameras thus all alternative solutions must take into account that the latency of acquisition is a critical factor.

References

1. Anderson, P.: Advanced Display Technologies JISC Technology and Standards Watch. JISC, Bristol (2005)
2. Wikipedia contributors, Polarized 3D system, Wikipedia The Free Encyclopedia (2013), http://en.wikipedia.org/wiki/Polarized_3D_system
3. Multi-Picture format, Camera & Imaging Products Association Standardization Committee, CIPA (2009)

4. Wallace, G.K.: The JPEG still picture compression standard. Communications of the ACM 34(4), 30–44 (1991)
5. Perkins, M.G.: Data compression of stereopairs. IEEE Transactions on Communications 40(4), 684–696 (1992)
6. Battiato, S., Mancuso, M., Bosco, A., Guarnera, M.: Psychovisual and Statistical Optimization of Quantization Tables for DCT Compression Engines. In: Proceedings of IEEE International Conference on Image Analysis and Processing, pp. 602–606 (2001)
7. Battiato, S., Bosco, C., Bruna, A., Di Blasi, G., Gallo, G.: Statistical Modeling of Huffman Tables Coding. In: Roli, F., Vitulano, S. (eds.) ICIAP 2005. LNCS, vol. 3617, pp. 711–718. Springer, Heidelberg (2005)
8. Schenkel, M.B., Luo, C., Frossard, P., Wu, F.: Joint decoding of stereo JPEG image Pairs. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 2633–2636 (2010)
9. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice Hall, Upper Saddle River (2008)
10. 3DMedia – 3D Technology and Software (2013),
    `http://www.3dmedia.com/gallery`
11. Pieter, S., Meesters, L., Ijsselsteijn, W.: Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation. ACM Transactions on Applied Perception (TAP) 3(2), 95–109 (2006)
12. Briechle, K., Hanebeck, U.D.: Template matching using fast normalized cross correlation. In: International Society for Optics and Photonics, Aerospace/Defense Sensing, Simulation, and Controls, pp. 95–102 (2001)
13. Kohonen, T.: The self-organizing map. Proceedings of the IEEE 78(9), 1464–1480 (1990)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
15. Banitalebi-Dehkordi, A., Pourazad, M.T., Nasiopoulos, P.: A human visual system based 3D video quality metric. In: The 2nd International Conference on 3D Imaging, IC3D, Liege, Belgium (2012)