Volume 2, Issue 1, January 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An analytical study on stereo image compression using DCT II and vector quantization

Ramaprabha M Sc MPhil
HOD, Dept of Computer Science
Sarah Tucker College (Autonomous)
Tirunelveli– 627007 ,South India

Dr.M.Mohamed Sathik
Principal
Sadakkathulla Appa College(Autonomous)
Tirunelveli-627 011,South India

Abstract: A stereoscopic system with a single left and right image pair should be transmitted simultaneously is twice of the raw data in a monoscopic image system, It needs an effective coding algorithm. Disparity compensation is a good method for stereo image compression that compresses the first image independently and then compresses the second image by estimating the disparity between the two images. The first image was transformed using 2-D Discrete Cosine Transform, quantized using the JPEG quantization matrix and finally encoded into a bit stream using arithmetic encoding. The disparity between the two images was estimated by the Block Matching Algorithm and the resulting disparity vector was then encoded into a bit stream. The images were then decoded and were compared with the original images.

Keywords: stereo image compression, stereoscopic image, vector quantization, wavelet transform, disparity compensation

I. Introduction

A pair of stereo images is very similar to each other that is the images of a object taken from two different angles. So we compress both images independently and an efficient way to compress a pair of stereoscopic images. The calculation of the difference between the two images is known as disparity estimation, then we compress one of the image independently. This image is known as the reference image and either the right image or the left image that reference image and the disparity vectors are then used to reconstruct the second image. A stereo image is produced by taking two cameras, and recording the perspectives of the right eye and the left eye using different lenses. The left image is seen through the left lens and the right image is seen through the right lens. The brain then merges the two images into one and also perceives the depth of the object. Stereo images can be produced cheaply using inexpensive digital cameras and are used widely in clinical applications, environmental science and entertainment. In this paper, the left image is taken as the reference image and the disparity vectors between the two images are estimated using the Block Matching Algorithm. The reference image is transformed using two-dimensional discrete cosine transform (DCT-II) and quantized using the quantization matrix. The resulting matrix is compressed into a bit stream using arithmetic coding An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

II. Compressing Stereo Images by DCT II

a) Encoding the image

The left image is taken as the reference image and is transformed using two dimensional forward discrete cosine transform . The resulting image matrix is quantized using the quantization matrix and then compressed using Arithmetic coding. The second part of the encoder involves compressing the right image. Since the two images are very similar to each other, disparity vectors between the two images are estimated. The resulting disparity vectors are compressed into a bit stream using arithmetic encoding .

Transform coding is a component of image processing and is very useful in the compression of still images . Transformation is used to remove the duplicate information between the neighboring pixels in the image. It depends on the how much a pixel in an image is correlated to its neighbor, and this correlation is used to predict the value of the neighboring pixel. To perform DCT-II, the first step is to ensure that the pixels are centered around zero. Since the value of a pixel is in the range [0, 255], it can be centered around zero by extracting the value of each pixel and subtracting from it 128. Then the image is divided into NxN blocks, and the 2-D DCT for each block is calculated by

$$C(u,v)=\alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

for u,v = 0,1,2…..,N-1 and alpha(u), alpha(v) are the horizontal and vertical spatial frequencies and are defined by

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & for \quad u = 0 \\ \sqrt{\frac{2}{N}} & for \quad u \neq 0. \end{cases}$$

In Quantization The human eye is not sensitive to variations in brightness of high-frequency components over a large area. Therefore, the high frequency values in the image matrix can be rounded off to zero without the user noticing any difference in the quality of the image .Quantization achieves this compression by dividing the DCT output for each block by a quantization and then rounding the result into the closest integer . The 8x8 quantization matrix is given below

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

Once DCT has been performed, the image is divided into 8x8 blocks, and each pixel is quantized as

Cij=round(Cij/Qij)

Arithmetic coding achieves lossless compression by assigning short codes to the frequencies that have a higher probability of occurrence and longer codes to frequencies whose probability of occurrence is low. The Block Matching Algorithm is the most accurate and easiest to implement. The size of a macro block was taken as 16x16 pixels and each block in the anchor frame was compared with each block in the target frame. The block with the minimum error was calculated by the equation

$$E_{\mathrm{DFD}}(\mathbf{d}_m) = \sum_{\mathbf{x} \in B_m} \left| \psi_2(\mathbf{x} + \mathbf{d}_m) - \psi_1(\mathbf{x}) \right|^p$$

The displacement between the current block and the best matching block is the displacement vector dm and is known as the disparity.

The blocks will have M rows and N columns and the search size is assumed to be R pixels. The search size is the range over which the reference image will search for the best matching block in the second image. This means that if the search size is 8 pixels, the best matching block will be searched over 8 pixels relative to the position of the current block. The x-component and the y-component of the image vectors were stored in xc and yc respectively. Therefore,xc(i, j) stores the x-component of the disparity vector corresponding to (i,j) -th and yc (i,j) stores the y-component of the disparity vector corresponding to the particular block. The algorithm to estimate the disparity isas follows: determine the size of the reference frame . calculate the size of disparity vectors xc and yc, based on the size of refrence image, M and N.For each block in the reference image , Search the second image, over a region of pixels relative to the position of the current block and calculate the sum of absolute difference between the pixels of the reference image and the second image. The minimum difference is the best matching block. The difference between the reference image and the second image is the disparity vector for the two images. This value is stored in the appropriate locations in xc and yc.

a) Decoding the Image

To decode the image, all the steps in encoding were performed in reverse order. The image matrix is divided into 8x8 blocks and each block is multiplied by the Quantization matrix. This is step is referred to as inverse quantization . The image is divided into NxN blocks and the inverse 2-D DCT of each pixel is calculated as

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \alpha(u)\alpha(v)C(u,v)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

The number 128 was subtracted from each pixel while calculating the forward DCT-II, therefore 128 is now added to each pixel value. To reconstruct the second image, we need image matrix of the reconstructed left image and the disparity vectors calculated. To reconstruct the right image is calculate M and N, based on the size of the reference image and mvx, then initialize disparity compensation vector and second image as the same size as the reference image. For each block in the reference image, Extract the current block from reference image . Extract the disparity vectors corresponding to the current location. Add the disparity vectors to the current block . The decoded left and right images were compared with the original left and right images. The Mean Square Error (MSE) between the original and decoded left and right images was calculated by

$$MSE = \frac{1}{MN} \sum_{y=1}^{M} \sum_{x=1}^{N} [I(x,y) - I'(x,y)]^2$$

The MSE of the image is the average of the MSE of the left and right image.

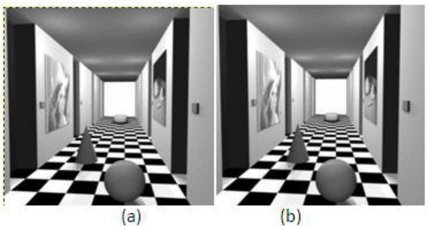
MSE = (MSEL + MSER)/2

The MSE was converted into Peak-Signal to Noise Ratio according to the formula

$$PSNR = 10\log_{10}\frac{255^2}{MSE}$$

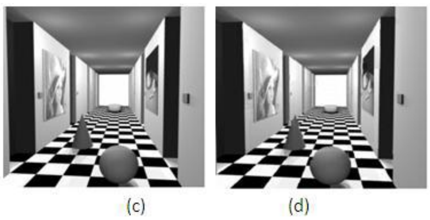
When we apply the method for the corridor stereoscopic image we have the following result

Fig 1.CORRIDOR:original



left Image right Image

Fig 2.CORRIDOR: Reconstructed



left image *Right Image*

III.Compressing Stereo Images by Vector quantization

a) Encoding the image

The encoder consists of two processes. The left image is compressed by using VQ techniques, then the reconstructed left image is used as reference to do disparity compensation with the right image. Here the left image is decomposed by wavelet transform. The transformed result used to generate the codebook. In wavelet transform, the input left image will be divided into small blocks of the same size as 4x4 pixels. The number of divided blocks is depend on the original size of the left image. For example, if the original size of the left image is 64x64 pixels, then we have 256 divided blocks, and each block's size is 4x4 pixels. Then, each block of the original left image is decomposed two times by wavelet transform (WT) as shown in Figure 3.

<table>
<tr><td>L</td><td>L</td><td rowspan="2">HL</td></tr>
<tr><td>H</td><td>H</td></tr>
<tr><td colspan="2">LH</td><td>HH</td></tr>
</table>

Fig 3.Wavelet decomposition

The wavelet decomposition is a multiresolution representation of a signal by using a set of basis functions that generated by the dilation and translation of a unique wavelet function Let Φ(t)be a low pass scaling function and Ψ (t) be an associated band pass wavelet function. Then, the two dimensional wavelet decomposition can be constructed by using the separable products of Φ(t)and Ψ (t). The discrete two-dimensional wavelet transform of the image function f(x,y) in one level decomposition can be written as follow

$$A_1^1f = ((f(x,y)*\Phi(-x)\Phi(-y))(2n,2m))(n,m)€z^2 \text{ ----(1)}$$

$$D_1^1f= ((f(x,y)*\Phi(-x)\Psi(-y))(2n,2m))(n,m)€z^2 \text{ -----(2)}$$

$$D_1^2f=((f(x,y)*\Psi(-x)\Phi(-y))(2n,2m))(n,m)€z^2 \text{ ------(3)}$$

$$D_1^3f = ((f(x,y)*\Psi(-x)\Psi(-y))(2n,2m))(n,m)€z^2 \text{ -----(4)}$$

As shown in Figure 3 and equations (1), (2), (3) and (4), the image function f(x,y) is decomposed into four components. The component A_1f (LL channel) is an approximation part of image and the other components D_1^1f (LH channel), D_1^2f (HL channel) and D_1^3f (HH channel) are the details of the image. Therefore, most energy of the image f(x,y) is condensed in the wavelet transform image, i.e., A_1f component. In the first level of wavelet decomposition, the individual image block of size 4x4 pixels is decomposed into 4 small blocks of size 2x2 pixels. Therefore, A1f will be a block of size 2x2 pixels, at the first time.

After decomposition in the second level, the smallest wavelet coefficients inside four components are obtained as shown in Figure 4. Each coefficient pixel has its sign, positive or negative, based on wavelet transform. We will use these characteristics for training the vector and then generating compact codebook.

<table>
<tr><td>A_1</td><td>D_1^1</td><td rowspan="2">D_2^1</td></tr>
<tr><td>D_1^2</td><td>D_1^3</td></tr>
<tr><td colspan="2">D_2^2</td><td>D_2^3</td></tr>
</table>

Fig 4.Wavelet Coefficients

codebook is generated based on wavelet decomposition and training vector operation. Once the closest codeword is found, the index of that codeword is sent through a channel (the channel can be computer storage, communications channel, and so on) followed by all codebooks. The importance of this method is the creation of the smallest effective codebook that can be used as the reference to reconstruct the satisfied quality image.

b) Disparity Compensation

After VQ process, the reconstructed left image is used as a reference to predict right image. And by using disparity compensation process, the error of reconstructing the right image can be reduced.

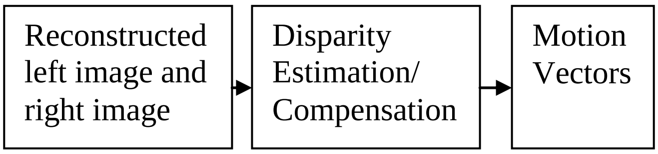


Fig 5.Disparity Compensation

Figure 5 shows the disparity compensation for encoding right image. Initially, the classical block matching algorithm is employed in order to estimate correspondence between blocks of the two images. The right image is partitioned into 8x8 pixels-block and each block is scanned to match with a corresponding block in the reference feedback reconstructed left image.

c) Decoder

Decoder is used to reconstructing the left and right image. The left image is reconstructed from VQ process and the right image is reconstructed from disparity compensation process. For reconstructing the left image, when the decoder receives the codebook and the index of the codeword from encoder part, it replaces the index with the associated codeword.The reconstructed left image is used as the reference to reconstruct the right image by searching match vector with motion vector and compensation. The implementation of the proposed compression technique and its simulation results are described as followes

IV. Implementation and Results

The input sources are stereo image pairs in bitmap format and their sizes are 64x64 and 128x128 pixels, gray scale images. The “Room” image represents virtual reality, while the two scenes provide different distance scenarios that exhibit different disparity properties. The left and right image of the “Room” pair differ mainly in the left edge of the left image where a piece of the wall is visible that cannot be found in the right image. We compress the left image by VQ with haar wavelet and sent its reconstructed image to compress right image with vector estimation and compensation. The feedback reconstructed left image will Reduce error for the reconstructing right image because it is also using in the decoding part. The example of reconstructed right image from disparity compensation process is shown in Figure 6. And the results of reconstruct left image by VQ coding technique using haar wavelet are represented in Table 1. The mean peak-signal-to-noise ratio is used as a measure of the stereo pair’s reconstruction quality (in decibels) as follow;

(a) (b) (c)

Figure 6: (a) and (b) are the original “Room” in left and right stereo images, Fig (c) is a reconstructed right image from disparity compensation method.

$$PSNR = 10\log_{10} * 255^2/(D_L+D_R)/2$$

where D_L and D_R are the difference between the original and reconstructed of left and right image,. The results of stereo image compression obtained by vector quantization are represented as bit per pixel (bpp), compression ratio (CR) and Peak Signal to Noise Ratio (PSNR) in decibels

Table 1: Reconstructed stereo image quality in size

64x64 pixels in 8 bits (decibels for PSNR) in DCT II

Bits/pixel	CR	PSNR
0.49190	16.264	33.281
0.48257	16.578	25.260

Table 2: Reconstructed stereo image quality in size

128x128 pixels in 8 bits (decibels for PSNR) using VQ.

Bits/pixel	CR	PSNR
0.661537	17.5266	35.8647
0.859580	18.6791	38.2231

V. Conclusions

The disparity compensation can reduce stereo image size. The application of quadrant vector quantization that uses the advantage of coefficients symbol obtained by wavelet transform to make compact codebook. The algorithm does not only reduce the number of codeword for training vector, but also gives good quality of reconstructed image. Higher quality is obtained by smaller image experiment. Both a methods are well suited for stereo image compression .

IV. References

[1] W.Beil andI.C.Carlsen, “Surface reconstruction from stereoscopy and “shape from shading” in SEM images” in Machine Vision and Applications, 1991, pp 271-285

[2] G.Strang “The Discrete Cosine Transform” in SIAM Review, 1999, pp 135-147

[3] S.Bique "New Characterizations of 2D Discrete Cosine Transform" in IEEE Transactions on Computers, 2005, pp 1054-1060

[4] A.Gersho and R.M.Gray "Vector quantization and signal compression",Springer, 1991

[5] W.Niehsen "Fast full-search block matching" in IEEE Transactions on circuits and systems for video technology, 2001, pp 241-247

[6] N.S.Sulthana, M.Chandra "Image Compression with Adaptive Arithmeti Coding" in International Journal of Computer Applications, 2010, pp 31- 34

[7] Mark S. Moellenhoff, Mark W. Maier "Characteristics of disparity-compensated stereo image pair residuals", Signal Processing Image Communication, vol. 14, pp. 55-69

[8] Michael G. Perkins, "Data Compression of Stereopairs", IEEE Trans. Commun., vol.40,pp. 684-696, Apr. (1992).

[9] M. G. Strintzis and S. Malassiotis, "Objectbased coding of stereoscopic and 3D image sequences: A review", IEEE Signal Processing Mag., vol.16, pp. 14 -29, May. (1999).

[10] Qin Jiang, Joon Jae Lee and Monson H. Hayes, "A Wavelet Based Stereo Image Coding Algorithm", ICASSP'99 Proceedings, pp.3157-3160, (1999).