JPEG-2000 Gradient-Based Coding: An Application To Object Detection

*Dae Yeol Lee **Guilherme O. Pinto ***Sheila S. Hemami
*ETRI **Cornell University ***Cornell University, Northeastern University
*daelee711@etri.re.kr

Abstract

Image distortions, such as quantization errors, can have a severe negative impact on the performance of computer vision algorithms, and, more specifically, on object detection algorithms. State-of-the-art implementations of the JPEG-2000 image coder commonly allocate the available bits to minimize the Mean-Squared-Error (MSE) distortion between the original image and the resulting compressed image. However, considering that some state-of-the-art object detection methods use the gradient information as the main image feature, an improved object detection performance is expected for JPEG-2000 image coders that allocate the available bits to minimize the distortions on the gradient content. Accordingly, in this work, the Gradient Mean-Squared-Error (GMSE) based JPEG-2000 coder presents an improved object detection performance over the MSE based JPEG-2000 image coder when the object of interest is located at the same spatial location of the image regions with the strongest gradients and also for high bit-rates. For low bit-rates (e.g. 0.07bpp), the GMSE based JPEG-2000 image coder becomes overly selective in choosing the gradients to preserve, and, as a result, there is a greater chance of mismatch between the spatial locations of the gradients that the coder is trying to preserve and the spatial locations of the objects of interest.

1. Introduction

Image distortions, such as JPEG or JPEG-2000 quantization errors, can have a severe negative impact on the performance of computer vision algorithms, and, more specifically, on object detection algorithms. The goal of this project is to modify the JPEG-2000 image coder [1] to provide an improved object detection performance for a given bit-rate. Some state-of-the-art object detectors rely heavily on the gradient information to detect the objects. In this case, an image compression scheme which focuses on the preservation of the gradient content is expected to provide a better performance at detecting objects.

State-of-the-art implementations of the JPEG-2000 image coder commonly allocate the available bits to minimize the Mean-Squared-Error (MSE) distortion between the original image and the resulting compressed image. The coefficients of each code-block are quantized on different levels so that the resulting image has the least MSE for a given bit budget.

Ref. [1] developed a rate-allocation algorithm for arbitrary quality and utility estimators within the Post-Compression Rate-Distortion Optimization (PCRD-opt) framework in JPEG-2000 image coding. This work will then use this rate-allocation algorithm to compress the images and allocate the available bits to minimize the Gradient Mean-Squared-Error (GMSE) distortion between the original image and the resulting compressed image. In this scenario, the bits are not allocated to preserve the perceived quality of the coded image, but to preserve the gradient information. The performance of a state-of-the-art object detector [2] will be evaluated on the set of images compressed with both the MSE and GMSE based JPEG-2000 image coders for various rates.

2. Background

2.1. JPEG-2000 Image Coding

The JPEG-2000 image coder uses the Embedded Block Coding with Optimal Truncation (EBCOT) framework in the encoder [5]. In this framework, the image is decomposed into different subbands by the discrete wavelet transform. Each subband of wavelet coefficients is divided into non-overlapping blocks, termed as code-blocks.

For an arbitrary code-block, there is a required number of bit to represent the coefficient with the greatest magnitude, and this number of bits, in turn, represents the number of bit planes for an arbitrary code-block. In addition, for each bit plane there are three different coding passes: a significance propagation pass, a magnitude refinement pass, and a cleanup pass. Roughly speaking,

each passes for each bit plane is seen as a possible truncation point and for each of those truncation points, the distortion and the length, in bits, are computed.

The Post-Compression Rate Distortion Optimization procedure chooses different truncation points on different code-blocks to minimize the distortion measure on a given bit budget. The optimal set of truncation points $\{z_i\}$ should satisfy,

$$Min \Sigma_{i=1}^{M} D(z_i), \text{ subject to } , \Sigma_{i=1}^{M} L_i(z_i) \le L_{\max},$$

where $D(Z_i)$ represents distortion measure on z_i, L_i represents the bit length on z_i, and L_{max} represents the target bit budget.

This work uses both the MSE and the Gradient Mean-Squared Error estimators to compute the distortions $D(z_i)$ for all code-blocks and truncation points. The resulting JPEG-2000 coders will be referred to as JPEG-2000-MSE and JPEG-2000-GMSE respectively.

2.2. Gradient Mean-Square-Error (GMSE)

The GMSE of a distorted image corresponds to the Mean-Squared-Error between the gradient-maps of the reference and distorted images. The gradient on each pixel is calculated by convolving the input image with the Sobel filter in the horizontal and vertical orientations. The computation of the GMSE for the whole image equals to

$$GMSE(R,T) = \frac{1}{M} \Sigma_{i=1}^{M} \sqrt{(\nabla R_x(i) - \nabla T_x(i))^2 + (\nabla R_y(i) - \nabla T_y(i))^2}$$

where M denotes the number of pixels in the image. R and T denote the reference and distorted images. ∇R_x and ∇T_x denote the horizontal gradient maps of images R and T, respectively. And ∇R_y and ∇T_y denote the vertical gradient maps of images R and T, respectively.

2.3. Object Detection

The object detection algorithm adopted in this work [2] uses the Histogram of Oriented Gradients (HOG) as the main feature for object detection. The HOG feature computes the strength of the gradient orientation in small portions of images. The histogram F of a particular pixel of the image is calculated in the following way:

$$B(x,y) = \mathrm{round}(\mathrm{P}\frac{\theta(\mathrm{x,y})}{2\pi})\mathrm{modp} \ ,$$

$$F(x,y)_b = r(x,y) \ \text{ if } \ b = B(x,y)$$

where $\theta(x,y)$ and $r(x,y)$ denote the orientation and the magnitude of the gradient on pixel at position (x, y), respectively. The p value is the bin quantization factor, which in this work is set to 18 to capture the strength of the gradients with orientation on every 20 degrees. The images are partitioned into small blocks of size 8x8 pixels and the HOG feature is calculated for each block. The HOG features of the object go through a matching process on the HOG features of the test image. The scores computed by this matching process give information about the expected location of a particular object.

2.4. Performance Evaluation

The object detection performance is evaluated using the Average Precision method which is effective in evaluating the accuracy of the information given by some ranked sequence. For this particular application, we are given a list of coordinates of the detected bounding boxes sorted by the scores. The locations of the detected bounding boxes are compared with those of the ground-truth bounding boxes, provided by the PASCAL VOC 2007 database. The calculated Average Precision score will serve as a metric for the object detection algorithm.

Each detected bounding box is compared with the ground-truth bounding boxes and is defined either as a true positive if the detected bounding box has an overlap amount of more than 50 percent with respect to any ground-truth bounding boxes, and false positive if otherwise. In the case of multiple detected bounding boxes having an overlap amount of more than 50 percent with respect to a single ground-truth bounding box, a non-maximum suppression procedure is applied to greedily select the detected bounding box with the highest score as the true positive.

The Average Precision score is calculated in the following way

$$\mathrm{average\,precision} = \frac{\Sigma_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{prec(i)} \times \mathrm{rel(i)}}{\mathrm{Number\,of\,ground\,truth\,objects}}$$

$$\mathrm{prec(i)} = \frac{\mathrm{cum\,tp(i)}}{\mathrm{cum\,tp(i)} + \mathrm{cum\,fp(i)}}$$

where n refers to the total number of detected bounding boxes, cum tp(i) and cum fp(i) refers to the cumulative number of true positives and false positives on the ith detected bounding box respectively, and rel(i) is an indicator function, which is 1 only if the ith detection is true positive and 0 otherwise. The prec(i) refers to the precision value at moment i. The Average Precision value is given as the sum of those precision values divided by the total number of ground-truth objects. The higher the Average Precision score is, the better the object detection performance is.

3. Experimental Setup

This work uses the PASCAL Visual Object Challenge 2007 (VOC2007) database [3], which contains a total of 4952 test images, 5011 train images and 20 different objects. Also, this work uses the state-of-the-art object detection algorithm, based on

Histogram of Oriented Gradients (HOG), presented in [2]. Specifically, this work evaluates the performance of the JPEG-2000-GMSE and JPEG-2000-MSE image coders for 8 different objects and for the rates of 0.2, 0.13, 0.1 and 0.07 bits per pixel(bpp).

The overall computational framework is as follows. First, the uncompressed trainset images are used to train the model filters for the different objects. After that, the test images are compressed for different bit-rates for both the JPEG-2000-GMSE and JPEG-2000-MSE image coders. Then, these compressed images are used by the object detection algorithm to compute the resulting bounding boxes and Average Precision scores for the two different image coders.

4. Result and Analysis

In this section, we analyze the object detection performances of the JPEG-2000-MSE and JPEG-2000-GMSE image coders for various objects and bit-rates. The Average Precision scores for various objects are presented for the rates of 0.2, 0.13, 0.1 and 0.07 bpp in Table I. In summary, the results indicate that the JPEG-2000-GMSE coder has an improved performance over the JPEG-2000-MSE coder for high bit-rates and also for selected objects, such as sheep and bottle. The results also show that the performance of the JPEG-2000-GMSE coder deteriorates for low-bit-rates, presenting a worse performance than the JPEG-2000-MSE coder.

4.1. Improved object detection by the JPEG-200-GMSE coder

The JPEG-2000-GMSE image coder shows improved object detection performance over the JPEG-2000-MSE coder when the locations of the objects of interest match the image locations with the strongest gradients.

Also, as the bit-rate increases, it is possible to preserve the gradient information in a greater portion of the image, which in turn increases the likelihood that the objects of interest will match the gradients preserved by the JPEG-2000-GMSE image coder. This is the reason why the JPEG-2000-GMSE image coder has a superior performance for higher rates.

Moreover, if the objects of interest occur in scenes without strong gradients or with a background composed mostly of low-frequency content, the JPEG-2000-GMSE coder outperforms the JPEG-2000-MSE coder for various rates. In this work, this scenario happens for the objects sheep and bottle. In general, the sheep object appears in rural areas with the grass at the bottom of the image and the sky at the top. The bottle object appears on tables or inside vending machines. Accordingly, JPEG-2000-GMSE coder provides an improved performance for the sheep object for all rates, and for the bottle object for all rates, with the exception of the lowest one.

4.2. Improved object detection by the JPEG-200-MSE coder

For low bit-rates, the JPEG-2000-MSE coder outperforms the JPEG-2000-GMSE coder when the object of interest does not lie on the region with the strongest gradients. In this work, this situation occurs primarily on images with urban backgrounds, with many strong edge structures. In this scenario, the JPEG-2000-GMSE image coder allocates most of the bits to the edge structures in the background, while allocating fewer bits to the objects of interest which leads to inferior object detection performances, when compared to the JPEG-2000-MSE coder.

4.3. JPEG-2000-GMSE coder for low bit-rates

If an image is compressed at a very low bit-rate, the JPEG-2000-GMSE image coder is only able to preserve a small portion of the image gradients, and, as a result, there is a greater chance of mismatch between the spatial locations of the gradients that the coder is trying to preserve and the spatial locations of the objects of interest. This will thus cause a decrease in performance of the JPEG-2000-GMSE coder, when compared to the JPEG-2000-MSE coder. Specifically, Fig. 1 and 2 show images that have relatively strong gradients on the object of interest. If these images are compressed at the moderate rate of 0.1 bpp, both images have superior object detection performance for the JPEG-2000-GMSE coder, when compared to the JPEG-2000-MSE coder. But for 0.07 bpp, the JPEG-2000-GMSE image coder has an inferior object detection performance, when compared to that of the JPEG-2000-MSE coder. Fig. 1(b) and 2(b) indicate that this probably happens because the strongest gradients are located at the fence structures and at the vehicles in the backgrounds, respectively rather than on the object of interest.

5. Conclusion

This project evaluates the performance of both the GMSE and MSE based JPEG-2000 image coders with respect to the object detection task.

The JPEG-2000-GMSE shows improved object detection performance when the locations of the strongest gradients of the image match the locations of the objects of interest. For higher bit-rates (e.g. 0.2 bpp) the JPEG-2000-GMSE coder is able to

Table I. Average precision scores on various rates

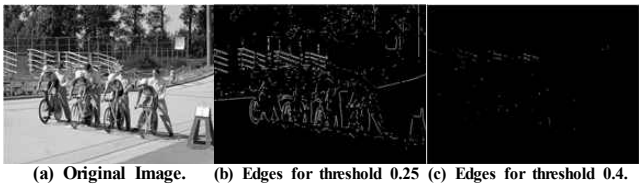
	Sheep	Car	Bike	Bird	Boat	Bus	Person	Bottle
Uncompressed	0.21	0.55	0.59	0.04	0.12	0.47	0.40	0.22
0.20 bpp								
JPEG-2000 MSE	0.15	0.48	0.52	0.02	0.07	0.42	0.33	0.13
JPEG-2000 GMSE	0.16	0.48	0.51	0.02	0.08	0.44	0.35	0.14
0.13 bpp								
JPEG-2000 MSE	0.12	0.45	0.48	0.02	0.07	0.42	0.31	0.10
JPEG-2000 GMSE	0.13	0.46	0.48	0.02	0.07	0.42	0.31	0.12
0.10 bpp								
JPEG-2000 MSE	0.13	0.44	0.44	0.02	0.07	0.39	0.28	0.08
JPEG-2000 GMSE	0.14	0.43	0.47	0.02	0.07	0.39	0.29	0.09
0.07 bpp								
JPEG-2000 MSE	0.10	0.40	0.42	0.02	0.07	0.36	0.25	0.07
JPEG-2000 GMSE	0.11	0.40	0.41	0.02	0.06	0.35	0.25	0.06

preserve the gradients in larger parts of the image, which increases its object detection performance when compared to the JPEG-2000-MSE coder. If the objects of interest occur in scenes without strong edges or strong gradients, the JPEG-2000-GMSE coder outperforms the JPEG-2000-MSE for various rates. In this work, this scenario happens for the objects sheep and bottle.

When the locations of the strong gradients of the image do not match the locations of the objects of interest, the JPEG-2000-GMSE spends extra bits to preserve the gradients in other image regions, and this may lead to inferior object detection performance when compared to the JPEG-2000-MSE coder. For lower bit-rates (e.g. 0.07bpp), the JPEG-2000-GMSE becomes selective in choosing the gradients to preserve, and, as a result, there is a greater chance of mismatch between the spatial locations of the significantly stronger gradients within the image that the coder is trying to preserve and the spatial locations of the objects of interest.

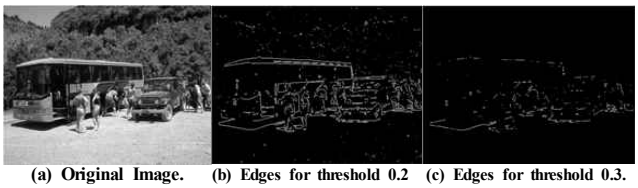
Acknowledgement

This research was partially funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.



(a) Original Image. (b) Edges for threshold 0.25 (c) Edges for threshold 0.4.

Fig. 1. Detection of object: *Bicycle*. Ground-truth with 4 bicycles. On 0.1 bpp, JPEG-2000-MSE and JPEG-2000-GMSE had 2 and 4 detected bicycles, respectively. On 0.07 bpp, JPEG-2000-MSE and JPEG-2000-GMSE had 2 and 0 detected bicycles, respectively.



(a) Original Image. (b) Edges for threshold 0.2 (c) Edges for threshold 0.3.

Fig. 2. Detection of object: *Person*. Ground-truth with 7 people. On 0.1 bpp, JPEG-2000-MSE and JPEG-2000-GMSE had 2 and 6 detected people, respectively. On 0.07 bpp, JPEG-2000-MSE and JPEG-2000-GMSE had 2 and 2 detected people, respectively.

References

[1] G.O. Pinto, S.S. Hemami, “Interplay between Image Coding and Quality Estimator” SPIE Human and Electronic Imaging (HVEI), 2013.

[2] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, “Object Detection with Discriminatively Trained Part Based Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, Sep. 2010.

[3] Everingham, M. and Van~Gool, L. and Williams, C. K. I. and Winn, J. and Zisserman, The PASCAL Visual Object Classes Challenge 2007(VOC2007), http://www.pascalnetwork.org/challenges/VOC/voc2007/workshop/index.html.

[4] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Discriminatively Trained Deformable Part Models, Release 5. http://people.cs.uchicago.edu/~rbg/latent-release5/.

[5] D. S. Taubman and M. W. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice, Kluwer Academic Publishers, Boston, 2002.