Structural Similarity As a Prediction Metric in Lossy Image Set Compression

Jason Ranger, Howard Cheng[1]
Department of Mathematics and Computer Science
University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, Canada
jason.ranger@uleth.ca, howard.cheng@uleth.ca
[1]corresponding author

Abstract—*An automatic compression strategy proposed by Gergel* et al. *is a near-optimal lossy compression scheme for a given collection of similar images whose inter-image relationships are unknown. That algorithm uses the root mean square error (RMSE) as a measure of the similarity between two images. Since RMSE is a metric, provable guarantees on the quality of the decompressed images can be made. However, it is well known that the RMSE does not correspond well to the human visual system. Recently, Brunet* et al. *introduced a metric based on structural similarity (SSIM). In this work, we show that the application of a SSIM-based metric instead of RMSE in the lossy image set compression scheme give improvements on some types of image sets.*

Keywords: Image set compression, root mean square error, structural similarity index, prediction metric.

1. Introduction

Many modern applications, such as medical imaging centers, store and generate enormously large volumes of images [20]. A number of different strategies to remove inter-image redundancies in sets of similar images have been proposed [2], [4], [5], [14], [15], [16], [17], [19], [21], [22], [23], [24], [25]. Many of these techniques, such as the Centroid method [15], [17], perform well on image sets with particular inter-image relationships, but are less effective on others. It is not clear which method will perform best *a priori* for any particular image set.

The automatic lossy compression strategy of Gergel *et al.* [7], [9] allows for the efficient storage of collections of similar images without any prior knowledge of the images. The unifying graph theoretical framework allows for the comparison of all previous techniques that look at the relationship between pairs of images [7], [8], [9]. This framework led to the discovery of an automatic compression strategy performing no worse than any previous strategy, and often performing better. Instead of storing n original images, a subset of $n-1$ difference images are stored. The subset of difference images are selected by studying the compressibility of each of the $\binom{n}{2}$ difference images. The root mean square error (RMSE) is used to predict the compressibility of the difference images, and the accuracy of this prediction directly affects the performance of the image set compression scheme. It was assumed heuristically that the RMSE between two images is small if and only if the difference image is easy to compress. In addition, the RMSE is a metric in the mathematical sense, which implies that a guarantee on the overall quality of the decompressed images can be made from quality guarantees of the decompressed difference images. It is important to note that the RMSE is used in two different ways here—for estimating compressibility of difference images and for measuring the quality of each decompressed image.

Despite the fact that RMSE is easy to analyze mathematically, it is well known that it does not always correspond to the human visual system (HVS)—two images that are similar to human may have a very large RMSE between them. Moreover, two images may have a large RMSE between them and yet their difference can be easily compressed. For example, two images whose difference image is a large constant will have a large RMSE even though the difference image can be compressed extremely well. Many other measures for image quality have been proposed to model the HVS more closely but they are generally not metrics [26]. In particular, they do not satisfy the triangle inequality.

The structural similarity (SSIM) index is a measure designed to provide better assessments of visual distortions between two images [27], and recently a metric based on the SSIM measure have been proposed [3]. This paper examines the application of a SSIM-based metric in lossy image set compression to see if there are benefits in replacing RMSE with this metric. It will be shown that for some types of image sets, using a SSIM-based metric gives slightly better results. We also discuss when the simple application of a SSIM-based metric may not be useful for other types of image sets. This points to the need for more research in a better way to incorporate SSIM-based metrics into the lossy image set compression framework, so that their properties can be better exploited.

2. Preliminaries

2.1 Image Set Compression

Traditionally images in a set are compressed individually using standard image compression algorithms (see, for example, [12]). Often a set of images are similar and inter-image redundancies can be used to further improve compression performance. For example, a video can be considered a sequence of images whose inter-image redundancies are defined by the time index. In these cases, each image can be predicted from another one known to be very similar (e.g. previous frame in a video), so that the overall storage requirement is reduced.

In some applications, the images in the collection are similar but the relationships are not known *a priori*. Gergel *et al.* [7], [9] modelled this problem as a complete weighted graph, so that each image is represented as a vertex. A "zero image" is also added to the set as a known starting point for both the encoder and the decoder. The edge between a pair of images has a weight that is used to predict the difficulty in compressing one image when the other one is known. Some choices of edge weights include the entropy (different orders) of the difference image, the root mean square error (RMSE) between the two images, or even the actual bit rate obtained by a compression algorithm on the difference image. Other measures are studied in [10]. In cases in which images in the set are very similar, it may also be helpful to insert an additional average image that is the centroid of the entire set.

The automatic compression scheme examines this graph and computes a minimum spanning tree (MST). The edges in the MST represent the difference images that are compressed. These edges must form a connected subgraph to ensure that all images in the set can be decompressed. When the edge weights correspond exactly to the compressibility of difference images, this gives an optimal compression scheme. As in any predictive coding scheme, error propagation is avoided by predicted the next image using the reconstructed version of the previous image instead of the original image. We refer to the automatic compression scheme the MST method, or the MST$_A$ method if the average image is included.

The difference images are compressed using "standard" image compression algorithms. In our experiments, we use JPEG2000 [1] and wavelet packet compression [18]. It was shown that wavelet packet compression generally outperforms JPEG2000 [7] since the statistical properties of difference images are very different from those of the photographic images that JPEG2000 is designed for.

Since the prediction is based on the reconstructed version of the previous image while the MST calculation is based on the original images, the optimality of the computed MST can no longer guarantee that the selected difference images result in the best compression scheme. However, near-optimality can still be obtained when the distortion measure (between original and reconstructed images) is a metric and is the same as the chosen edge weight. In particular, if the distortion between each original image and the corresponding reconstructed image is bounded by Δ, then the total weight of the edges chosen by the MST methods is at most $O(n\Delta)$ worse than the MST of the "hidden" graph where all images can be distorted by at most Δ where n is the number of images [7], [9]. Thus, the scheme is at most $O(\Delta)$ away from the optimal compression scheme per image, on average. When the prediction measure correlates well with the actual bit rate used, this can be used as a guarantee on the optimality (in terms of bit rate) given a particular distortion bound.

2.2 Structural Similarity

It is well known that while the RMSE has many useful mathematical properties, it is not necessarily suitable for comparing images. In our application, we may have two images whose difference image is easy to compress and yet have a large RMSE. The main reason is that RMSE simply averages the pixelwise differences without regard to the trends and structures in the images.

Structural similarity (SSIM) was introduced to overcome some of these issues and more closely correlate to human perception [27]. Given two signals $\mathbf{x}$ and $\mathbf{y}$ of N samples each, a simplified form of SSIM that is commonly used in application is

$$\begin{aligned} S(\mathbf{x},\mathbf{y}) &= S_1(\mathbf{x},\mathbf{y})S_2(\mathbf{x},\mathbf{y}) \\ &= \left[\frac{2\bar{\mathbf{x}}\bar{\mathbf{y}}+\epsilon_1}{\bar{\mathbf{x}}^2+\bar{\mathbf{y}}^2+\epsilon_1}\right]\left[\frac{2s_{\mathbf{xy}}+\epsilon_2}{s_{\mathbf{x}}^2+s_{\mathbf{y}}^2+\epsilon_2}\right], \end{aligned} \tag{1}$$

where

$$\begin{aligned} &\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \bar{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{N} y_i \\ &s_{\mathbf{x}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{\mathbf{x}})^2, \quad s_{\mathbf{y}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{\mathbf{y}})^2 \\ &s_{\mathbf{xy}} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{\mathbf{x}})(y_i-\bar{\mathbf{y}}), \end{aligned} \tag{2}$$

and ϵ_1 and ϵ_2 are small constants to prevent numerical instability. Intuitively, the first component $S_1(\mathbf{x},\mathbf{y})$ compares the means of the signal while the second component $S_2(\mathbf{x},\mathbf{y})$ measures the correlation and contrast distortion. The value of $S(\mathbf{x},\mathbf{y})$ is in the range $[-1,1]$, and it is not a metric. It was shown that pairs of images with the same value of $S(\mathbf{x},\mathbf{y})$ have a similar amount of visual distortions subjectively [27]. The same could not be said about RMSE.

Brunet *et al.* [3] constructed a metric (in the mathematical sense) based on SSIM that also corresponds well with SSIM. We will only present one of the many related metrics defined in [3] and simplify it for our application. Given the signals

$\mathbf{x}$ and $\mathbf{y}$, define the function

$$\bar{d}(\mathbf{x},\mathbf{y}) = \begin{cases} \frac{\|\mathbf{x}-\mathbf{y}\|_2}{\sqrt{\|\mathbf{x}\|_2^2+\|\mathbf{y}\|_2^2}} & (\mathbf{x},\mathbf{y}) \neq (\mathbf{0},\mathbf{0}) \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The D_2 metric is defined as

$$D_2(\mathbf{x},\mathbf{y}) = \sqrt{\bar{d}^2(\bar{\mathbf{x}},\bar{\mathbf{y}}) + \bar{d}^2(\mathbf{x}-\bar{\mathbf{x}},\mathbf{y}-\bar{\mathbf{y}})}. \tag{4}$$

The D_2 metric is a good approximation for the SSIM measure [3].

3. Structural Similarity in Image Set Compression

The SSIM measure (as well as the D_2 metric as defined in Equation (4)) are often used in “maps” in which the measure is applied to (possibly smoothed) windows to show areas in a reconstructed image with the most distortion from the original image. In our application, however, we must arrive at a single numerical value for measuring the differences between two images. In this work, we have chosen to apply the D_2 metric on the entire images to arrive at a single value.

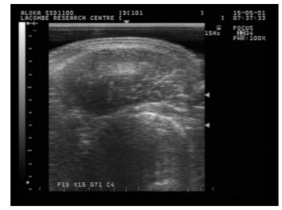
As described in Section 2.1, the MST and MST_A lossy image compression schemes use image distortion measures in two ways—for predicting the compressibility of the difference image and to assess the quality of the reconstructed images. In order for near-optimality guarantees to be made, the same measure must be used in both settings. However, we also wish to compare our results with those when RMSE are used. As a result, experiments will be done in all four combinations using RMSE and D_2 in each of the two components.

Four image sets were used in the experiments. They were also used in previous works by Gergel *et al.* [7], [9] and Nielsen *et al.* [21], [22]. Figure 1 shows a typical image from the first four image sets. The Galway set contains webcam images from a street in Galway City, Ireland [6]. The Pig set is composed of ultrasound images of pig rib cages. The Joe set is another webcam image set taken from a camera directed at a beach in Victoria, British Columbia [13]. Satellite images from the GOES project [11] make up the GOES set. All the images were 8-bit gray scale images.

We present the results for the experiments in Figures 2– 9. For brevity, we only present the results for the MST scheme as the results for the MST_A scheme are very similar. We show the rate-distortion curves using RMSE and D_2 metrics as the prediction measures in the MST scheme. Here the distortion is the average distortion between each reconstructed image in the set and the corresponding original image. Intuitively, a “lower” curve indicates a better compression scheme.

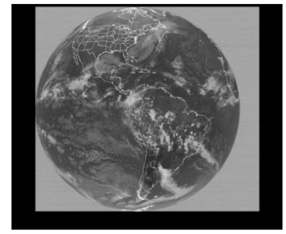
The results show that for every image set except the Joe set, using RMSE as the prediction measure in the MST lossy image set compression scheme is better than using D_2. The results are similar whether RMSE or D_2 is used

(a) Galway



(b) Pig

(c) Joe



(d) GOES

Fig. 1
TYPICAL IMAGES FROM EACH SET.

to measure the average distortion, and whether JPEG2000 or wavelet packet compression is used to compress the difference images. As expected, the results using wavelet packet compression are generally better than using JPEG2000.

For the Joe set, however, the results are slightly better when the D_2 metric is used. This can be understood by examining the properties of the image set. The images in this set consists of webcam images of a natural scene at different times of the day, over a number of days. Since the main differences among the images are the illumination, many difference images are easily compressible. However, the RMSE values are large and so the easily compressible difference images are not chosen by the MST scheme. Since the D_2 metric compares both the means and the signal with the means removed, it allows the easily compressible difference images to be chosen.

The Galway set also consists of webcam images but there are many more differences than simply variations of illumination (e.g. moving pedestrians). Although the RMSE metric performs better on this set, the gap is smaller here than in the remaining two sets—Pig and GOES. The images in these two sets have many more differences throughout the entire images (e.g. noise in ultrasound images and cloud patterns), so the correlation part of the SSIM measure (and hence the D_2 metric) will result in a large value even though the differences are in the high-frequency part of the spectrum and may be quantized away anyway, especially at lower bit rates.

We have also attempted to assess the quality of the reconstructed images visually, but neither method is consistently better for an entire set. Even within the same image, we find that some parts may be better when RMSE is used but some parts are worse compared to when the D_2 metric is used.

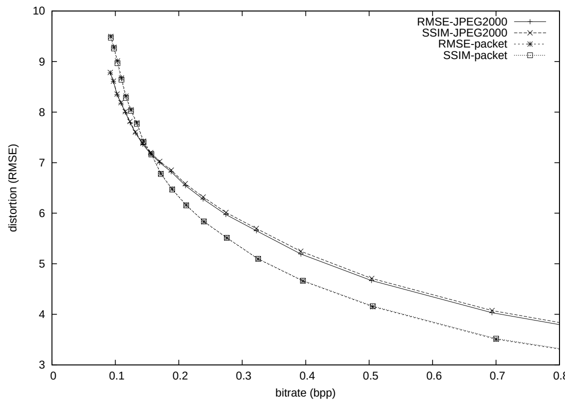


Fig. 2
RMSE RATE-DISTORTION RESULTS FOR THE GALWAY IMAGE SET.

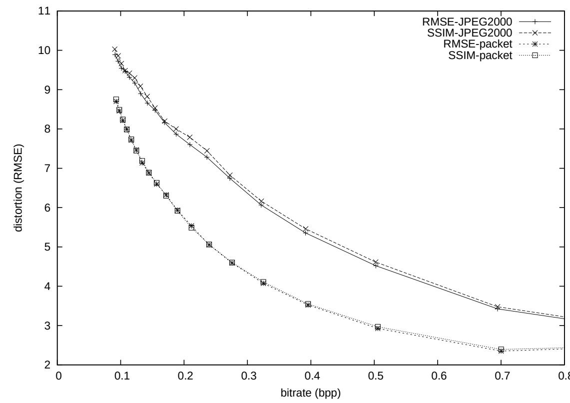


Fig. 4
RMSE RATE-DISTORTION RESULTS FOR THE PIG IMAGE SET.

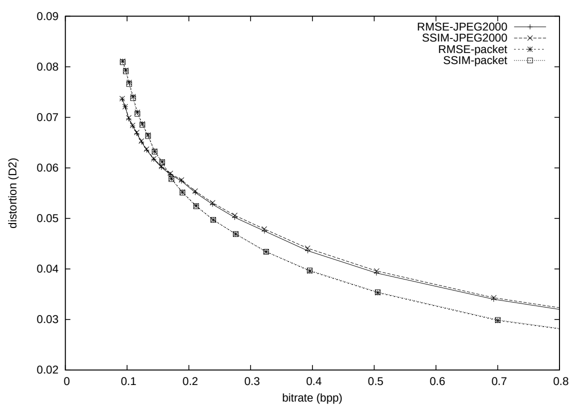


Fig. 3
D_2 RATE-DISTORTION RESULTS FOR THE GALWAY IMAGE SET.

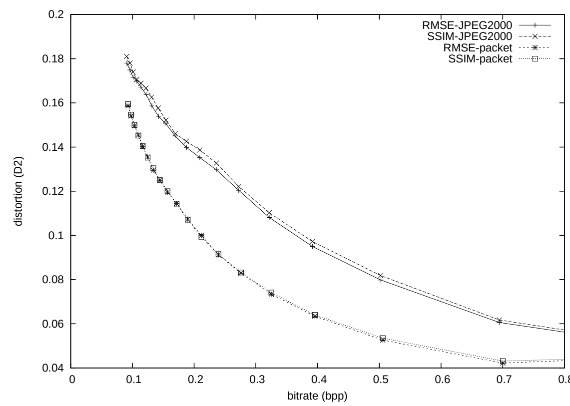


Fig. 5
D_2 RATE-DISTORTION RESULTS FOR THE PIG IMAGE SET.

4. Conclusions and Future Directions

In this paper, we examined the application of the D_2 SSIM-based metric as the prediction measure in MST-based lossy image set compression algorithms. We observed that for certain types of images, the D_2 metric can be used in place of RMSE to improve the performance of the MST lossy image set compression scheme.

This work represents our first step in this direction. The experimental results gave us some insights on how SSIM-based metrics behave in the MST compression scheme. We are currently investigating the use of a "windowed" SSIM metric that should adapt to local statistics better, as well as a number of ideas arising from this research.

References

[1] M. Adams. JasPer project. http://www.ece.uvic.ca/ mdadams/jasper/.

[2] S. Ait-Aoudia and A. Gabis. A comparison of set redundancy compression techniques. *EURASIP Journal on Applied Signal Processing*, 2006:1–13, 2006.

[3] D. Brunet, E.R. Vrscay, and Z. Wang. A class of image metrics based on structural similarity quality index. In *Image Analysis and Recognition (ICIAR) 2011, Part I*, number 6753 in Lecture Notes in Computer Science, pages 100–110, 2011.

[4] J.P. Cheiney and A. Touir. Fi-quadtree: A new data structure for content-oriented retrieval and fuzzy search. In *Proceedings of the International Symposium on Large /Spatial Databases*, pages 23–32, 1991.

[5] C.-P. Chen, C.-S. Chen, K.-L. Chung, H.-I. Lu, and G. Tang. Image set compression through minimal-cost prediction structures. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1289–1292, 2004.

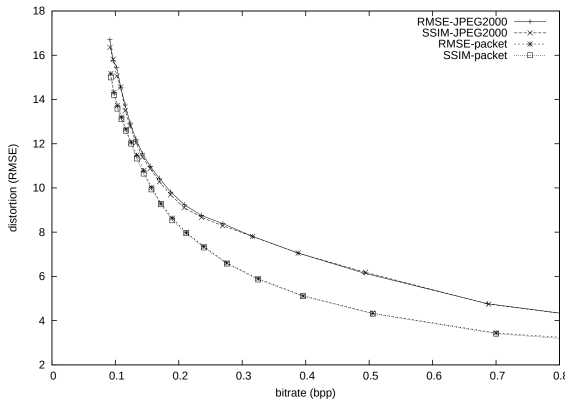


Fig. 6

RMSE RATE-DISTORTION RESULTS FOR THE JOE IMAGE SET.

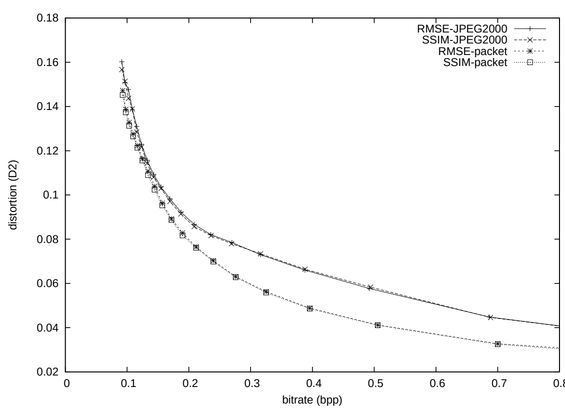


Fig. 7

D_2 RATE-DISTORTION RESULTS FOR THE JOE IMAGE SET.

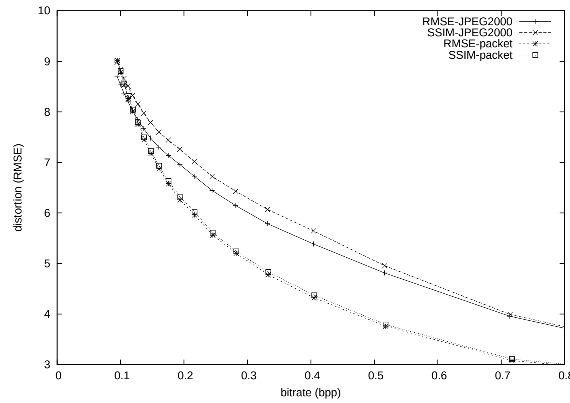


Fig. 8

RMSE RATE-DISTORTION RESULTS FOR THE GOES IMAGE SET.

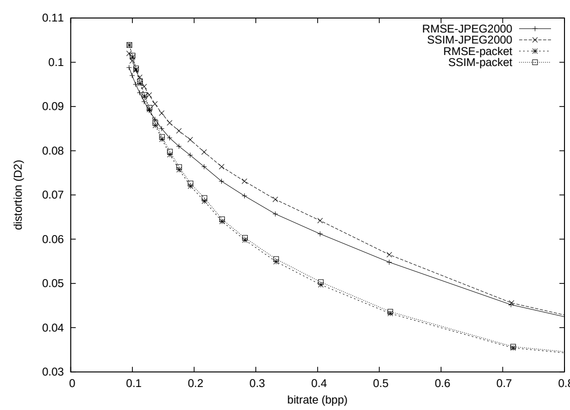


Fig. 9

D_2 RATE-DISTORTION RESULTS FOR THE GOES IMAGE SET.

[6] Galway.net. http://www.galway.net/webcam/images.shtml.

[7] B. Gergel. Automatic compression for image sets using a graph theoretical framework. Master's thesis, University of Lethbridge, March 2007.

[8] B. Gergel, H. Cheng, and X. Li. A unified framework for lossless image set compression. In *Proceedings of the Data Compression Conference*, page 448, 2006.

[9] B. Gergel, H. Cheng, C. Nielsen, and X. Li. A unified framework for image set compression. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pages 417–423, 2006.

[10] B. Gergel, A. Schmieder, X. Li, and H. Cheng. A study of prediction measures for lossy image set compression. In *Proceedings of the 2008 International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV'08)*, pages 69–74, 2008.

[11] GOES project science. http://rsd.gsfc.nasa.gov/goes.

[12] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition, 2007.

[13] JoeTourist Infosystems. http://www.JoeTourist.net.

[14] G. Jomier, M. Manouvrier, and M. Rukoz. Storage and management of similar images. *Journal of the Brazilian Computer Society*, 6(3):13–25, 2000.

[15] K. Karadimitriou. *Set redundancy, the enhanced compression model, and methods for compressing sets of similar images*. PhD thesis, Louisiana State University, 1996.

[16] K. Karadimitriou and J. M. Tyler. Min-max compression methods for medical image databases. *ACM Special Interest Group on Management of Data*, 26(1):47–52, 1997.

[17] K. Karadimitriou and J. M. Tyler. The centroid method for compressing sets of similar images. *Pattern Recognition Letters*, 19(7):585–593, 1998.

[18] F. Meyer, A. Averbuch, and J. Strömberg. Fast adaptive wavelet packet image compression. *IEEE Transactions on Image Processing*, 9(5):792–800, 2000.

[19] Y. S. Musatenko and V. N. Kurashov. Correlated image set compression system based on new fast efficient algorithm of Karhunen-Loeve transform. In C.-C. J. Kuo, S.-F. Chang, and S. Panchanathan, editors, *Proceedings of The International Society for Optical Engineering: Multimedia Storage and Archiving Systems III*, pages 518–529, October 1998.

[20] H. Nagata and H. Tanaka. Estimate of total volume of medical data in a year in Japan. In *Proceedings of the APAMI and CJKMI-KOSMI Conference*, pages 117–121, 2003.

[21] C. Nielsen and X. Li. MST for lossy compression on image sets. In *Proceedings of the Data Compression Conference*, page 463, 2006.

[22] C. Nielsen, X. Li, and K. Abma. Methods of grouping similar images for compression coding. In *Proceedings of the International Conference on Computer Vision (WCAC'05)*, pages 93–99, Las Vegas, June 20–23, 2005.

[23] M. Vassilakopoulos and Y. Manolopoulos. Dynamic inverted quadtree: a structure for pictorial databases. *Information Systems*, 20(6):483–500, 1995.

[24] M. Vassilakopoulos, Y. Manolopoulos, and N. Economou. Overlapping quadtrees for the representation of similar images. *Image and Vision Computing*, 11(5):257–262, 1993.

[25] J. Wang and H. Yan. Form image compression using template extraction and matching. In *Proceedings of Visual Information Processing*, December 2000.

[26] Z. Wang and A.C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[27] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.