PERFORMANCE ANALYSIS OF TEXTURE SIMILARITY METRICS IN HEVC INTRA PREDICTION

Karam Naser, Vincent Ricordel, Patrick Le Callet

▶ **To cite this version:**

Karam Naser, Vincent Ricordel, Patrick Le Callet. PERFORMANCE ANALYSIS OF TEXTURE SIMILARITY METRICS IN HEVC INTRA PREDICTION. Video Processing and Quality Metrics for Consumer Electronics (VPQM), Feb 2015, Chandler, Arizona, United States. hal-01150595

HAL Id: hal-01150595
https://hal.science/hal-01150595v1

Submitted on 11 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PERFORMANCE ANALYSIS OF TEXTURE SIMILARITY METRICS IN HEVC INTRA PREDICTION

Karam Naser, Vincent Ricordel, Patrick Le Callet

LUNAM University, University of Nantes, IRCCyN UMR CNRS 6597
Polytech Nantes, Rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France
karam.naser; vincent.ricordel; patrick.le-callet
@univ-nantes.fr

ABSTRACT

The visual signal is highly occupied by regions of homogeneous and repetitive patterns known as Textures. Textures have a common property that their similarity highly deviates from point by point comparison, i.e, two textures can look very similar even if they have some shift, rotation and difference in their distribution.

In the context of compression, All of the MPEG reference encoders (including HEVC), aim at minimizing the bitrate at a certain distortion level measured in terms of pixel comparison. For textures, this kind of distortion measure does not usually reflect the amount of perceived distortion. For this reason, we investigate the use of state of the art perceptual similarity metrics as a replacement for this measure. In other words, we aim at optimization the bitrate such that we minimize the perceptual distortion rather that the pixels difference. We used two metrics (Local Radius Index and Structure Texture Similarity Metric) in selecting the best intra prediction mode and block partitioning. Experimental results showed that these metrics try always to retain some structural properties of the textures. These metrics also showed a better rate-distortion performance when the distortion is measured via a distance metric based on texture features.

1. INTRODUCTION

Visual signal is a rich source of information. It can be decomposed into different classes and components. In this work, we consider it as two components, structure and texture. Structure contains the semantic meaning of the scene (edges, lines, corners, etc.) and textures fill the gap between structures. According to this, Textures range from very simple ones, like a DC block, to more complex ones. They can be also classified as regular and stochastic. Each texture type possesses different spectral, statistical and perceptual properties. For this reason, encoding them without considering their properties does not end up with the best rate-quality performance, which is the optimal goal of video compression.

In terms of visual similarity, humans are less sensitive to variations in textures, i.e, two textures can look very similar even if they have some deviations in scale, orientation and repetition. In contrast, they can look very different even if they have small average pixel difference. Therefore, assessing the texture similarity by the means of pixel comparison is avoided. Instead, the comparison can be done between structural information (encoded in frequency subband channels) and/or statistical information. this problem is an active research topic in both engineering and psychology. Details of different texture similarity metrics can be found in [1].

Video coding standards, such as HEVC, aim at minimizing the bitrate within a certain distortion level. The measured distortion is typically the mean squared error. This type of distortion, which is based on comparing pixel values, does not proportionally reflect the amount of perceived distortion (especially for texture components). For this reason, many approaches consider different weight for the distortion computed at each block according to its texture complexity (ex. [2]). Others consider replacing this measure by more appropriate ones (ex. [3]). In contrast, some approaches try to replace some textures with a synthesized ones which look visually very similar (ex. [4] [5]).

In this paper, we investigate the possibility of encoding static textures in such a way that the encoder try to minimize the bitrate while maximizing visual similarity between the reconstructed and the original signal. We borrow two similarity metrics from the texture retrieval problem. The two metrics, known as Local Radius Index (LRI) and Structure Texture Similarity Index (STSIM), are the most recent and successful ones in texture retrieval. We used them to select the best intra prediction mode and block partitioning. By doing so, we do not violate the HEVC standard, the HEVC decoder can thus be directly used to decode the resulting bitstream.

The rest of the paper is organized as follows: Section 2 gives an overview of the texture similarity metrics used

in this work. Section 3 presents the procedure carried out to evaluate each metric. In Section 4.1, the experimental results are provided and discussed with a conclusion given in Section 5.

2. OVERVIEW OF TEXTURE SIMILARITY METRICS

Texture similarity metrics exist in various forms. Some of them compare the statistics of textures in the spatial domain and others in the subband frequency domain (details can be found in [1]. In this paper, we consider LRI and STSIM as being recent and successful texture similarity metrics.

2.1. STSIM

STSIM was presented in [6] and further improved in [7]. It is based on comparing set of statistics in the subband decomposition. These statistics consist of mean, standard deviation, and horizontal and vertical auto-correlation of each subband. Beside that, it computes also the cross correlation between subbands with the same scale or subbands with the same orientation. This set of statistics provides a solid description of a given texture and thus can well characterize the similarity between two textures.

2.2. LRI

LRI is a successive to STSIM. LRI is much less computationally expensive as compared to STSIM and performs better in the context of similar texture retrieval [8]. It computes local index for each pixel in the spatial domain, beside that, it also computes the local binary pattern, standard deviation of each subband in the subband frequency domain and an intensity penalization term. Thus it is a combination between the analysis in frequency decomposition and spatial domain.

2.3. Adaptation of Similarity Metrics in HEVC

2.3.1. STSIM

One interesting property of STSIM is that it is bounded between one and zero, where one is the maximum similarity index. Embedding this metric in HEVC is straightforward. As we want to use it as a distortion metric, we consider the distortion function as:

$$D_{STSIM} = 1 - I_{STSIM} \tag{1}$$

where I_{STSIM} is the STSIM similarity index.

2.3.2. LRI

LRI, in contrast, is not bounded. It computes information divergence between the distribution of the local indexes of two textures. This divergence, as well as the one used inside the local binary pattern, can easily have infinite value when used as block based distortion measure (specially for small blocks comparison) as both of them compute the logarithm of the probability distribution. To overcome this problem, we modified both of them in the following manner:

- We use a normalized Kullback Leiber Divergence (KLD) in LRI. Since the Gibbs inequality assures that KLD value cannot accede the entropy of the first element of two compared distributions, we normalize KLD by dividing it by this entropy. The only exception is when any compared distributions have a probability value of zero when the other does not. In this case, we assume that the two distributions are different and we return the maximum value which is one.
- We normalize also the the local binary pattern term which computes the log-likelihood function between two distributions. The function is divided it by the entropy of the second distribution. The similarity function is then obtained by subtracting the normalized LBP from one. Similarly, we return the maximum value (one) when the distributions are assumed to be different.

We also eliminate the intensity penalization term in our work. This term was initially added to penalize the differences in the intensity values. Since we normalized LRI, the IP term has a greater impact on the overall metric and causes the metric to be more pixel dependent, which is far away from our goal.

3. PERFORMANCE EVALUATION

To evaluate the performance of the metrics, we implemented both STSIM and LRI and integrated them in HEVC encoder. These metrics were used inside the cost function for selecting the best intra prediction mode and block partitioning. To elaborate more, the details of HEVC intra prediction mode selection is given below:

3.1. Intra Mode Selection in HEVC reference Encoder

HEVC defines 33 possible directional prediction modes. Beside this, it also defines DC and Planar prediction modes. For each mode, the residual block can be directly encoded (transform, qunatization and entropy coding) or further partitioned into quadtree. Each time, the cost function is computed. The particular mode and partition which minimizes the cost is then selected.

In the reference encoder, HEVC considers only 3 most probable modes for the full rate-distortion optimization. These 3 modes are the ones among the 35 modes that have a minimum Sum of Absolute Transformed Difference (SATD) between the original and prediction blocks. For selecting the best mode, HEVC computes the Sum of Squared Difference (SSD) as a distortion metric in the cost function.

3.2. Replacing HEVC Distortion Metric

In this work, we replaced the SATD and SSD of HEVC reference encoder by distortion metrics evaluated with perceptual similarity metrics. We experimented both STSIM and LRI and adapted in HEVC as described in 2.3. To keep the range of SATD and SSD, the metrics were multiplied by:

- $255 \times N \times N$ for SATD
- $255^2 \times N \times N$ for SSD.

where N is the dimension of the prediction block.

4. EXPERIMENTS AND RESULTS

We have experimented the use the two metrics in HEVC for coding static textures. We used Brodazt textures downloaded from USC-SIPI dataset [9]. This contains 13 different gray scale textures (see Fig. 1) which are extensively used in textures analysis for engineering and psychophysical experiments. We used HM 9.0 [10] as a host encoder. In the following subsections, we provide the details of each experiment.

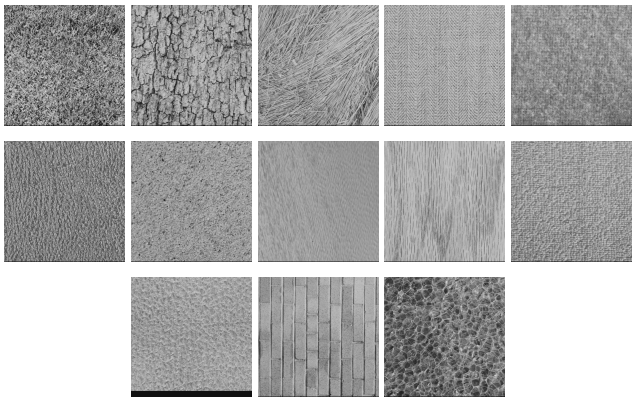

Fig. 1. Texture dataset used in this paper (from USC-SIPI dataset [9])

4.1. Quality of the Decoded Textures

To study the impact of using perceptual similarity metrics in HEVC, we encoded all textures for different QP values. First we compared the visual quality when using the default metric or the perceptual ones. For low QP values (low compression), no change has been observed. For high QP, however, significant changes can be seen. Examples of this are shown in the figures below.

The first example is shown in Fig. 2. We can see that encoding a highly structured texture with a large QP value results in loosing most of its semantics. This is because many blocks are replaced by DC values. Using either one of the metrics can retain the overall structure of the texture. One can also notice that there exists many wrong directions, but the overall quality is much more better.

Another example is shown in Fig.3. In this figure, the effect of wrong prediction direction is more clear when LRI is used. On the other hand, the right part of the texture is completely eliminated when the default metrics are used.

Fig. 4 gives an example of a extensive compression. In this case, the details of the texture are lost in all three coding configuration. But one thing to notice that there is different kind of compression artifacts when the perceptual metrics are used. We noticed that whatever hard the compression is, the perceptual metrics try to provide a structure of a textures rather than simple DC blocks. This can be particularly interesting if we consider the no reference quality of the decoded texture. This means that the original texture is somehow replaced by another texture which looks more natural than when the default metrics are used.

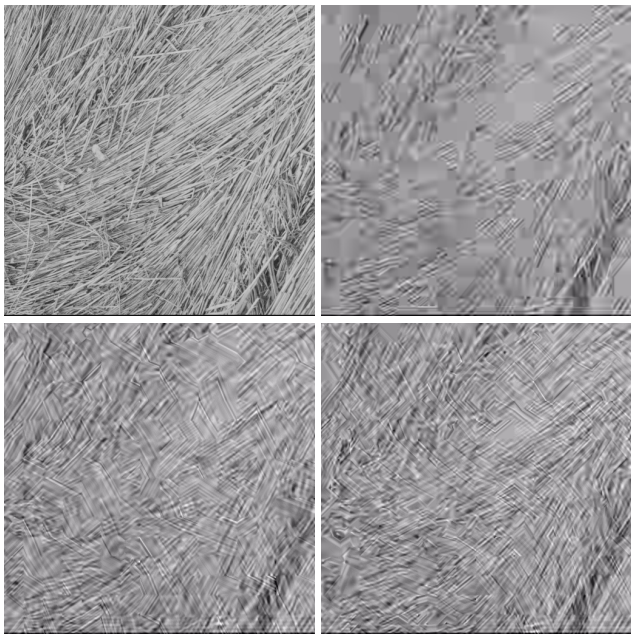

Fig. 2. Example of using LRI and STSIM inside HEVC (QP=51). Top left: original Image, top right: encoded with default HEVC metrics, down left: encoded with LRI, down right: encoded with STSIM.

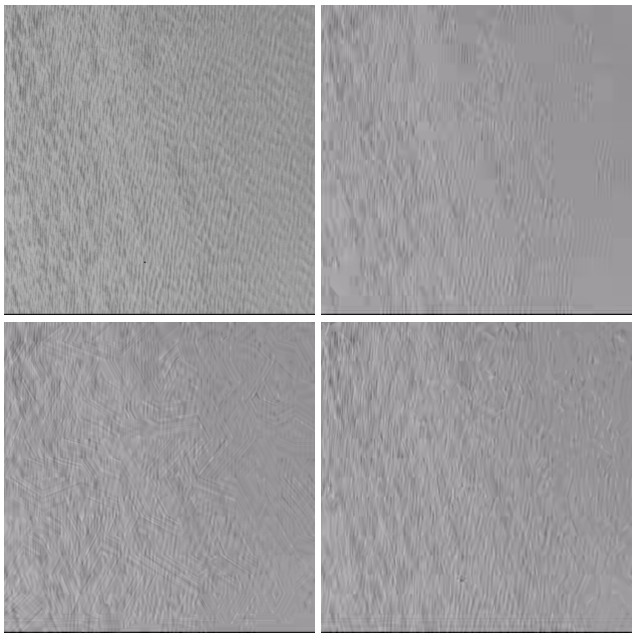

Fig. 3. Example of using LRI and STSIM inside HEVC (QP=43). Top left: original Image, top right: encoded with default HEVC metrics, down left: encoded with LRI, down right: encoded with STSIM.

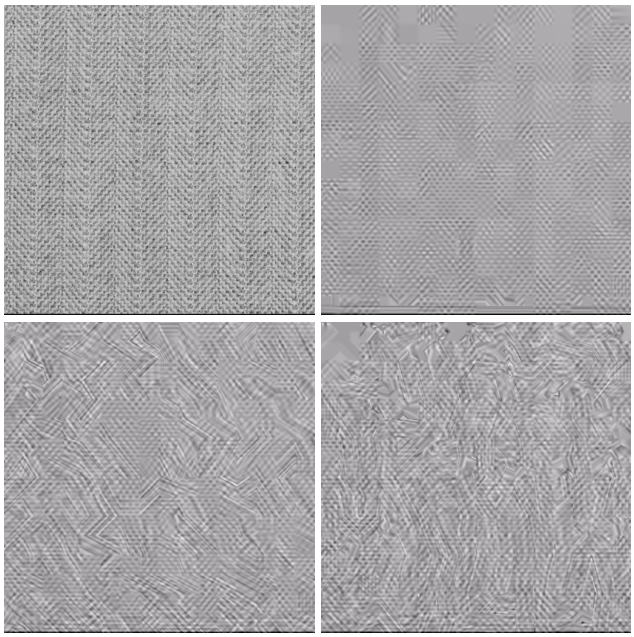

Fig. 4. Example of using LRI and STSIM inside HEVC (QP=51). Top left: original Image, top right: encoded with default HEVC metrics, down left: encoded with LRI, down right: encoded with STSIM.

4.2. Encoder Partitioning Behavior

To understand more the effects of each metric on the prediction mechanism, we measured the number of splitting depths that the encoder uses to encode each texture. We ran a simulation on the 13 textures from the used dataset with a quantization parameter (QP) taken range of [22, 27, 32, 37, 43, 47, 51]. The corresponding histograms are shown in Fig. 5. The splitting depth of zero means that the prediction block has its maximal size (64x64). Increasing the splitting depth by one corresponds to partition the block into four sub-blocks. In this figure, the histograms were scaled by the number of smallest blocks (4x4) that the corresponding splitting depth contains. This was done to have a fair comparison between splitting depths as each splitting occupies different areas of the frame. We observe from these histograms that when the default metrics are used, the encoder uses small prediction blocks for low compression (low QP) to better approximate the input signal. For high compression, it tries to approximate large prediction blocks (mostly with DC values) to have better compression. The behavior is totally different when LRI or STSIM is used. The encoder behavior does not change much as the compression changes. It uses always large block sizes to approximate the input signal and small block sizes (less than 16x16) are rarely chosen. The is because these metrics compare statistics of different distributions. For small block sizes, there is always a lack of enough statistics and usually these metrics return a high value of distortion in such a condition.

4.3. Rate Distortion Analysis

Until now, we considered only the quality of the decoded textures. In video coding, however, the comparison between two different schemes is a rate-distortion based. For this work, the rate-distortion comparison is not straightforward. This is because, up to our knowledge, no reliable quality metric has been designed designed for textures. The usual approach of comparing PSNR value does not provide a useful information as it compare pixel values, which is far away from the purpose of this work. Beside that, the type of distortion when the perceptual metrics is used is totally different and cannot be assessed by PSNR (ex. see Fig 2).

Again, we follow the same approach of borrowing a distance metric from texture retrieval problem. We avoid using LRI or STSIM as this may result in biased assessment. We used another metric [11] which is based on comparing features of textures in the frequency domain. This metric is compared to both LRI and STSIM ([8]) and provides close by results in retrieval rate. This metric compares the energy and mean of frequency subbands (computed using Gabor filters). The metric was downloaded from the authors website and used as a distortion metrics in our work.

By calculating the distance measure by this metric to

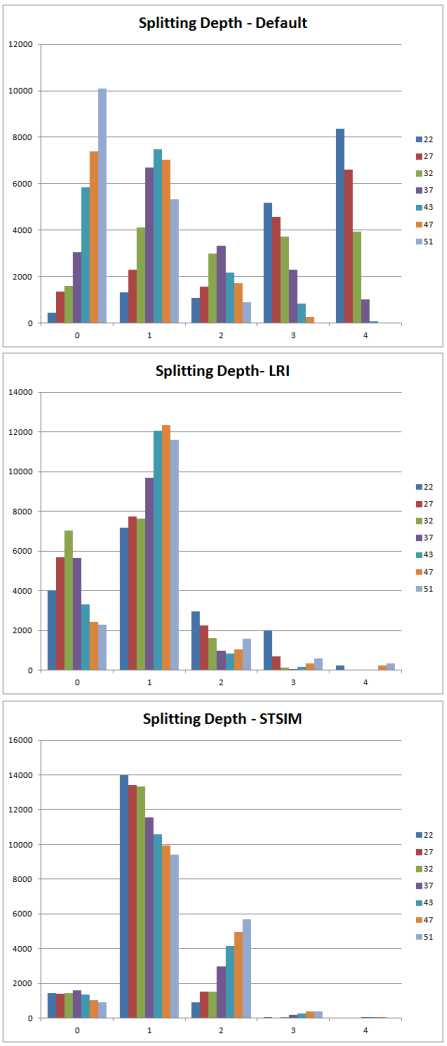

Fig. 5. Histograms of splitting depths vs QP. Each depth is scaled by the number of 4x4 block that it has

the original texture for different compression levels, we obtained the curves shown in Fig 6. What we observe from these is that in most cases, LRI and STSIM provides better score than the default metrics in the low bitrate region. For high bitrate region, no gain when using LRI or STSIM. This can prove that even for high compression, both the similarity metrics try to keep the structural information rather than replacing the texture blocks with DC blocks.

5. CONCLUSION

In this paper, we studied the effect of replacing some of the default distortion metrics in HEVC by perceptual ones. These metrics were used to select the best prediction mode and block partitioning. We used recent perceptual similarity metrics, namely LRI and STSIM which have been developed in the context of texture retrieval. The direct benefit of this approach is that it is its compatibility with HEVC standard, which means no modification to the decoder is needed.

When these metrics are used, the structural information of the textures is tried to be retained in contrast to the default metrics which tries to smooth the contents and potentially replace them by DC values. For severe compression, the decoded textures can have a noisy structure as HEVC intra prediction cannot provide anything better than parallel lines of the directional prediction. Using both metrics, wrong prediction directions might be selected. This is because these metrics are less sensitive to pixel by pixel comparison. LRI, as compared to STSIM, is much less computationally expensive. But it results in more wrong prediction directions than STSIM. The reason behind this is the approximation used when the metric is adapted to this work.

The encoder behavior is changed when these metrics are used. It always tends to use large prediction block sizes for all range of compression. This is mainly because in small blocks, there is a lack of enough statistics to compare and the metrics will return high dis-similarity values. Experimental results showed that using these metrics can improve the overall perceived quality of the decoded textures. The small details of textures are better preserved and the decoded textures look more pleasant.

In terms of texture similarity, the rate-distortion curves show that both metrics perform better than the default metric. The distortion metric that was used compares the energy and mean of subband frequency channels obtained by Gabor filter.

As a conclusion, the use of texture similarity metrics can generally give a better no-reference quality than the default one. A possible further improvement will be rate-distortion optimization and perceptual post processing to reduce the effects of wrong prediction direction. Beside that, since the metrics usually work for large block size, a hybrid mechanism can be used to reduce the complexity of their use. That is, for blocks larger than 16x16, the perceptual metrics can be used where the default metrics can be used for the rest.

6. ACKNOWLEDGMENT

his work was supported by the Marie Sktodowska-Curie under the PROVISION (PeRceptually Optimized VIdeo CompresSION) project bearing Grant Number 608231 and Call Identifier: FP7-PEOPLE-2013-ITN.

7. REFERENCES

[1] T. Pappas, D. Neuhoff, H. de Ridder, and J. Zujovic, "Image analysis: Focus on texture similarity," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2044–2057, Sept 2013.

[2] H. Yu, F. Pan, Z. Lin, and Y. Sun, "A perceptual bit allocation scheme for h. 264," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 4–pp.

[3] T.-S. Ou, Y.-H. Huang, and H. H. Chen, "Ssim-based perceptual rate control for video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 5, pp. 682–691, 2011.

[4] F. Zhang and D. R. Bull, "A parametric framework for video compression using region-based texture models," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 7, pp. 1378–1392, 2011.

[5] J. Balle, A. Stojanovic, and J.-R. Ohm, "Models for static and dynamic texture synthesis in image and video compression," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 7, pp. 1353–1365, 2011.

[6] X. Zhao, M. G. Reyes, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for retrieval applications," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 1196–1199.

[7] J. Zujovic, T. Pappas, and D. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval," *Image Processing, IEEE Transactions on*, vol. 22, no. 7, pp. 2545–2558, July 2013.

[8] Y. Zhai, D. Neuhoff, and T. Pappas, "Local radius index - a new texture similarity feature," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 1434–1438.

[9] "USC-SIPI Dataset." [Online]. Available: http://sipi.usc.edu/database/

[10] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG, "HEVC test model 9.0," Tech. Rep., 2012.

[11] B. S. Manjunath and W.-Y. Ma, "Texture features for browsing and retrieval of image data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 8, pp. 837–842, 1996.

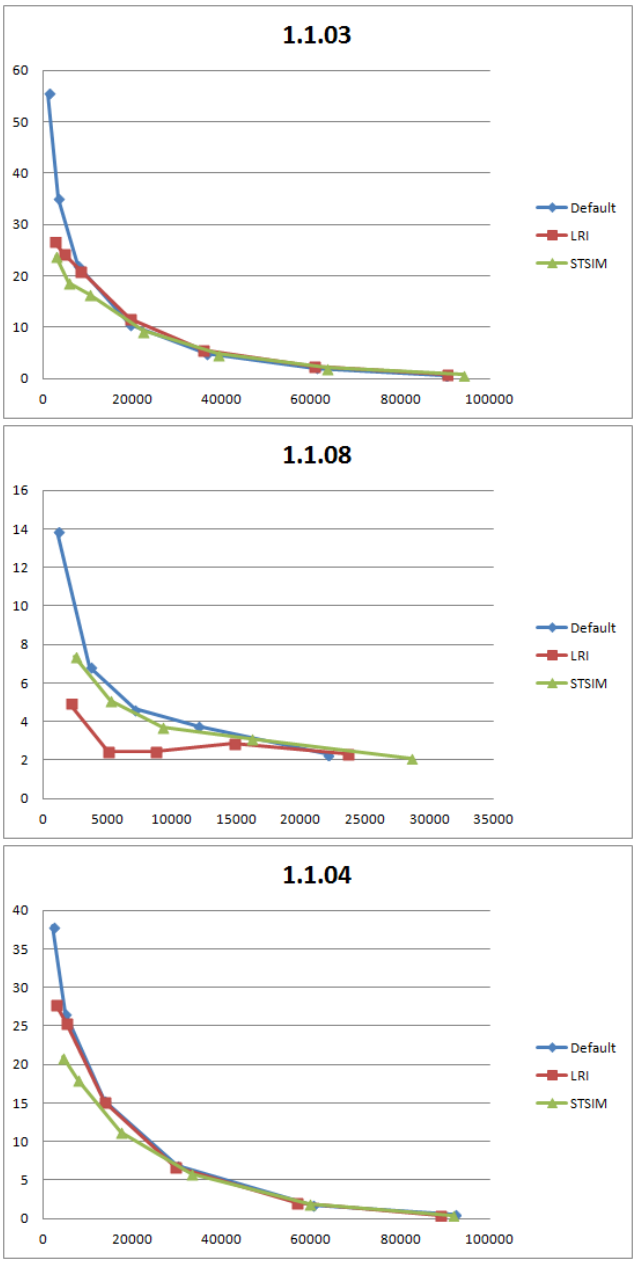

Fig. 6. Rate Distortion (Using Gabor distance metric) of the three texture shown in Fig. 2, 3 and 4 respectively. x-axes: Bytes used to encode the texture, y-axes distance to the original texture