JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IMAGE/VIDEO TRANSCODING WITH SPATIAL RESOLUTION REDUCTION

Shashank Sekhar[1], Dr. Sumit Bhattachargeee[2]

[1]Research Scholar, Sunrise University, Alwar, Rajasthan
[2]Research Supervisor, Sunrise University, Alwar, Rajasthan

Abstract

Optimal trade-off among distortion, rate, and/or complexity is one of the primary design concerns for image and video coding/transcoding due to the lossy nature of image/video compression and the costly bandwidth and compute resources in a multimedia system. To investigate the best RD performance of a video codec compatible to the latest video coding standard H.264 and to design computationally efficient down-sampling algorithms with high visual fidelity in the discrete Cosine transform (DCT) domain, this thesis investigates the application of rate distortion (RD) optimization approaches to image and video coding/transcoding. We have also presented a conceptual framework for image/video transcoding with spatial resolution reduction, i.e. to down-sample compressed images/video with an arbitrary ratio in the DCT domain, by examining the trade-off between distortion and complexity. To begin, we construct a set of down-sampling techniques that are modeled after a linear transform with double-sided matrix multiplication (LTDS) in the DCT domain.

Keywords: Video compression, spatial resolution adaptation, temporal resolution adaptation, perceptual video compression, CNN-based super-resolution.

INTRODUCTION

More and more people are using multimedia services like teleconferencing, video on demand, and distance learning because to the widespread availability of these tools and the exponential growth of the Internet. Video format conversion is often needed in these applications due to the need to accommodate a wide variety of channel capacities and end-user interface capabilities. One of the most important tools for accomplishing this hard process is transcoding. A transcoder is a piece of hardware or software that takes one compressed video bit-stream and produces another bit-stream with various bit-rates, spatial resolutions, compression standards, etc. Transcoding introduces significant difficulties for video watermarking technology in the area of copyright protection due to the complexity of the conversion processes it conducts. Then, unscrupulous individuals may transcode the video to produce a copyright-free sequence of videos with the same quality as the originals. In light of the aforementioned circumstances, it is imperative that an efficient video watermarking algorithm take into account the robustness of the watermark against video transcoding; however, almost all video watermarking techniques proposed in literature fail to adequately protect against copyright infringement because the embedded watermark is vulnerable to transcoding.

There have been various recent proposals in the literature for robust watermarking techniques that take transcoding resistance into account. Obtaining resilience against conversion of spatial resolution from High-Definition Television (HDTV) to Quarter Video Graphics Array is shown in Lee et a proposita's real-time video watermarking robust against transcoding (QVGA). In order to meet real-time constraints, this approach is applied directly to MPEG-2 video bit-streams, however this leaves the system open to attack when converting to other video compression standards with lower bit rates. Though the synchronization information and watermark sequence are embedded in the algorithm proposed by Chen et al. to protect against frame attacks, the re-synchronization mechanism of this scheme is not powerful enough to compensate for the frame rate reduction caused by some aggressive transcoding. In, the watermarking energy of a quantization-based video watermarking strategy in the DWT domain is customized for the human visual system (HVS). While the watermarks in this scheme are resistant to individual assaults from within the realm of signal processing, they are vulnerable to combined attacks from within the realm of typical transcoding activities. Using the Harris-Affine interest point detector, Ling et al. offer a video watermarking technique that is mostly resilient to geometrical distortions. This scheme's watermark resilience is highly dependent on precise identification of interest points, and aggressive transcoding often leads to erroneous detection of numerous interest points, lowering the scheme's performance.

LITERATURE REVIEW

Dost, S., Saud, (2022) There are primarily three distinct types of these methods: full reference (FR), reduced reference (RR), and no reference (NR) (NR). In RR techniques, we don't need to supply the original picture or video as a reference, but rather, we need to provide certain aspects (i.e., texture, edges, etc.) of the original image or video for quality evaluation. Research on RR-based quality assessment has gained traction in recent years for use in contexts as diverse as social media, video games, and streaming media. In this study, we give a review and categorization of recent studies that use RR to evaluate the quality of moving pictures. We have also included a summary of the many databases used in the area of evaluating the quality of 2D and 3D images and videos. Professionals and academics might benefit from reading this publication since it provides updated information on the development of RR-based image and video quality evaluation. This paper's examination and categorization of current methods for assessing the quality of multimedia content will be helpful for anybody interested in learning more about this topic. Additionally, it will aid the reader in choosing useful quality evaluation methodologies and criteria for their specific purposes.

Li-Heng Chen, (2022) In order to minimize the rate-distortion objective, we use a second neural network that is concurrently trained with the compression model to predict resizing factors for various inputs. Based on our findings, we propose that a "compression friendly" downsampled representation may be promptly found in the encoding phase by using an auxiliary network and differentiable picture warping. Extensive experiments on current deep image compression models demonstrate that our novel resizing parameter estimation methodology may deliver a Bjntegaard-Delta rate (BD-rate) increase of roughly 10% compared to state-of-the-art perceptual quality engines. Furthermore, we conducted a subjective quality analysis that demonstrates our novel method successfully compresses photos without sacrificing visual quality. The code used to generate this paper's results has been released under the MIT License at https://github.com/treammm/ResizeCompression in the hopes that it will be utilized to further future, repeatable studies in this area.

Vignesh V Menon (2022) The current industry standard for providing users with the best possible video quality over the internet is HTTP Adaptive Streaming (HAS). Multiple versions of the same video are encoded at different bitrates, resolutions, framerates, and encoding schemes in HAS. The goals of this research are to I provide fast and compression-efficient multi-bitrate, multi-resolution representations, (ii) provide fast and compression-efficient multi-codec representations, (iii) improve the encoding efficiency of Video on Demand (VoD) streaming using content-adaptive encoding optimizations, and (iv) provide encoding schemes with optimizations per-title for live streaming applications to reduce storage or delivery costs and/or increase quality of experience.

Dr. Y.l. Ajay kumar (2022) At some point during encoding, the ViSTRA system dynamically resamples the input video spatially and temporally based on a quantization decision option, and the decoder recreates the whole decision video. Temporal up sampling is accomplished by the repeating of frames, whereas spatial decision up sampling is achieved using a super-decision version of a Convolutional Neural Network (CNN). ViSTRA is currently an integral part of the High Efficiency Video Coding (HEVC) trendy software (HM 16.14). Results from a global trial show significant enhancement, with a BDcharge increase of 15 based on PSNR and a mean MOS difference of 0. five based on subjective visible best assessment.

Afonso, Mariana et al (2019) We present a spatio-temporal resolution adaptation (ViSTRA) video compression system, which dynamically resamples the input video in both space and time during encoding based on a quantization-resolution choice and then reconstructs the full resolution video at the decoder. In order to increase the spatial resolution, a convolutional neural network super-resolution model is used, while frame repetition is used to increase the temporal resolution. Reference program for high-efficiency video coding, which now includes ViSTRA (HM 16.14). Significant advances in BD-rate of 15% based on PSNR and an average MOS difference of 0.5 based on subjective visual quality assessments were seen in experiments that were confirmed through an international challenge.

RESEARCH METHODOLOGY

LTDS is modeled as a multi-layer neural network, which allows us to solve the optimization issue (6.12). Next, the network is trained using an algorithm for structural learning with forgetting.

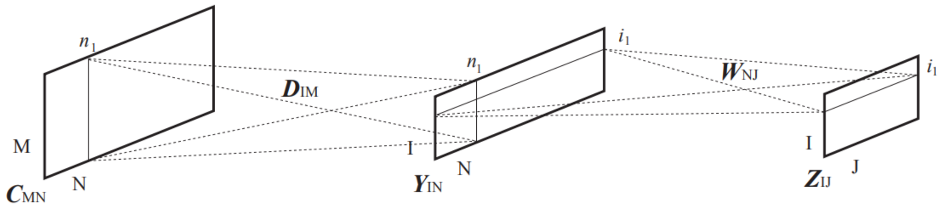


Figure 1: A three-layer network for implementing the linear transform of (6.8).

A Multiple-Layer Neural Network Structure

Figure 1 illustrates how an LTDS may be constructed as a three-layer neural network. Input layer C MN, hidden layer Y IN, and output layer Z IJ are all names taken from the multiple layer perceptron (MLP). The linear transform's matrix multiplication action is then mimicked by carefully establishing connections between units in the top and bottom layers.

The input layer's and the hidden layer's interconnections are shown in the left panel of Figure 2. Particularly, these ties are set up in accordance with the following three principles, i.e.

- Connections are established from units in a given column of the input layer to units in the same column of the hidden layer. Note that the input layer and the hidden layer have the same number of columns.
- Units in a given column of the input layer are fully connected to units in the same column of the hidden layer.
- Valid connections between any two columns share the same weight matrix, i.e., DIM.

Consequently, the output of the hidden layer is computed as $Y_{IN} = D_{IM} \cdot C_{MN}$ by a move data from the input layer to the hidden layer in a forward direction.

Likewise, the right panel of Figure 2 illustrates the links between the concealed layer and the output layer. Following the same criteria as before, links are made between rows and their respective columns.

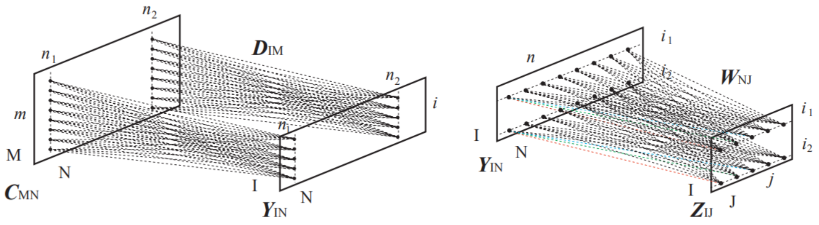


Figure 2: Illustration of selective connections in a three-layer network structure

for simulating the computation of LTDS. The left panel shows connections between the input layer C MN and the hidden layer Y IN. The right panel demonstrates connections between the hidden layer Y IN and the output layer Z IJ.

weight matrix is WNJ. Then, forwarding computation from the hidden layer to the output layer leads to $Z_{IJ} = Y_{IN} \cdot W_{NJ}$. Overall, the LTDS is implemented by a forwarding computation in the network structure as $Z_{IJ} = D_{IM} \cdot C_{MN} \cdot W_{NJ}$.

Training with Structural Learning with Forgetting

The initial motivation for the creation of SLF was the quest to simplify the framework of multilayer neural networks. SLF's fundamental notion is to clear out the complexity of a network by systematically thinning out unnecessary nodes via a decaying process. This research uses SLF to lessen the number of nodes in a three-layer network, which speeds up calculation for LTDS. There are two distinct phases to the learning process: general forgetting and more targeted forgetting.

Forgetful learning is refined to get rid of as many unnecessary first connections as feasible. By substituting rg = rf into the objective function of (6.12), i.e., the learning objective function is produced.,

$$J_f = ||\boldsymbol{D}_{\mathrm{IM}} \boldsymbol{C}_{\mathrm{MN}} \boldsymbol{W}_{\mathrm{NJ}} - \boldsymbol{V}_{\mathrm{IJ}}||^2 + \lambda \cdot r_{\mathrm{f}}.$$

Efficient Down-sampling Algorithm Design

The optimization issue in (6.12), using the 3-layer structure and the structural learning with forgetting method, is solved in the following way:

1. Generate a training set based on a given spatial-domain down-sampling method which down-samples an M×N image to a resolution of I×J. Choose several M ×N DCT images, $\{\boldsymbol{C}_{\mathrm{MN},i}, i = 1, \cdots, 5\}.$ Apply the pre-selected down sampling method discussed in Section 6.5.3 to obtain down-sampling references $\{\boldsymbol{V}_{\mathrm{IJ},i}, i = 1, \cdots, 5\}.$ The training set is $\{(\boldsymbol{C}_{\mathrm{MN},i}, \boldsymbol{V}_{\mathrm{IJ},i}), i = 1, \cdots, 5\}.$

2. Learning with forgetting. Construct the 3-layer structure with DIM and WNJ. Find a skeleton structure using the learning with forgetting algorithm.

3. Learning with selective forgetting. Refine DIM and WNJ with the learning with selective forgetting algorithm.

4. Combination of arithmetic operations to further reduce the computation cost.

As a consequence of running the aforesaid algorithm, we get a down-sampling technique based on LTDS that, for a set of parameters including, d0, and w0, minimizes the combined cost of visual quality and complexity. Changing any one of many factors will produce a procedure of varying complexity and visual quality. Values for these variables are left up to the discretion of the end user, who must take into account their own standards for acceptable quality and manageable complexity.

DATA ANALYSIS

The Learning Algorithm Has Converged. Solving (6.12) with SLF does not ensure global convergence. However, it is possible to demonstrate, for a certain set of training data, that the learning with forgetting method will converge to a minimum of (6.13) $(\boldsymbol{C}_{\mathrm{MN},i}, \boldsymbol{V}_{\mathrm{IJ},i}).$ Consider the Hessian matrix corresponding to WNJ.

$$\boldsymbol{G}_{\mathrm{NJ}\times\mathrm{NJ}}(\boldsymbol{W}) = \begin{pmatrix} \frac{\partial J_f}{\partial w_{11}\partial w_{11}} & \frac{\partial J_f}{\partial w_{11}\partial w_{12}} & \cdots & \frac{\partial J_f}{\partial w_{11}\partial w_{\mathrm{NJ}}} \\ \vdots & \cdot & \ddots & \vdots \\ \frac{\partial J_f}{\partial w_{NJ}\partial w_{11}} & \frac{\partial J_f}{\partial w_{NJ}\partial w_{12}} & \cdots & \frac{\partial J_f}{\partial w_{NJ}\partial w_{\mathrm{NJ}}} \end{pmatrix}$$

By some derivation, we have

$$\boldsymbol{G}_{\mathrm{NJ}\times\mathrm{NJ}}(\boldsymbol{W}) = \begin{pmatrix} [\mathcal{G}_{\mathrm{JJ}}]_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & [\mathcal{G}_{\mathrm{JJ}}]_{\mathrm{N}} \end{pmatrix}$$

with matrixes GJJ lying on the diagonal and $\mathcal{G}_{\mathrm{JJ}} = (\Delta \boldsymbol{Z}^{\mathrm{t}})_{\mathrm{JI}} \cdot \Delta \boldsymbol{Z}_{\mathrm{IJ}}.$ Apparently, GJJ is positive semi-definite. Therefore, the Hessian matrix $\boldsymbol{G}_{\mathrm{NJ}\times\mathrm{NJ}}(\boldsymbol{W})$ is positive semi-definite. Similarly, we can show that the Hessian matrix corresponding to DIM is,

$$\boldsymbol{H}_{\mathrm{IM}\times\mathrm{IM}}(\boldsymbol{D}) = \begin{pmatrix} [\mathcal{H}_{\mathrm{MM}}]_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & [\mathcal{H}_{\mathrm{MM}}]_{\mathrm{I}} \end{pmatrix}$$

With $\mathcal{H}_{\mathrm{MM}} = \boldsymbol{C}_{\mathrm{MN}} \cdot \boldsymbol{W}_{\mathrm{NJ}} \cdot (\boldsymbol{W}^{\mathrm{t}})_{\mathrm{JN}} \cdot (\boldsymbol{C}^{\mathrm{t}})_{\mathrm{NM}}.$ $\mathcal{H}_{\mathrm{MM}}$ has a positive ambiguity. This means that the Hessian matrix H IMIM(D) is positive semi-definite. So, Jf is a convex function in D and W, as will be shown. The learning-with-forgetting method, as demonstrated, is a gradient-descending one (6.14). Finally, we draw a conclusion on the convergence of structural learning and forgetting for reducing (6.13).

Superiority in the Eyes. Generally speaking, (6.12) demonstrates that the optimal visual quality for a down-sampling procedure achieved is limited solely by the down-sampling method chosen in advance in the spatial domain. Extensive research on spatial-domain down-sampling4 has been done, as described above. Consequently, it makes sense to utilize a technique that produces the highest possible visual quality reference picture. In particular, we choose the appropriate spatial-domain approach by analyzing low-pass filtering and interpolation layouts. There is often a compromise between aliasing, low-frequency components, and ringing in the design of low-pass filters used for down-sampling. While the anti-aliasing and preservation of low-frequency components of an image are well-served by a filter with a sharp transition band, the ringing along intensity borders in the filtered picture is a drawback of this design choice. The Butterworth filter is a common low-pass filter option because it strikes a good balance between the three criteria. This research employs a low-pass filtering procedure prior to down-sampling, and the Butterworth filter is used for this purpose. Two 1D Butterworth filters' frequency response functions guide our selection of LM1 and R1N.

$$|H(f)| = \sqrt{\frac{1}{1+(f/f_c)^{2L}}},$$

The cutoff frequency, fc, and the transition band order, L, are both defined as follows. The widely-used cubic B-spline interpolation was selected as the interpolation method of choice since it produces a derivative of continuous second order.

It would seem that the benefits of anti-aliasing, ringing avoidance, and low-frequency component preservation would be passed on to an LTDS generated in Section 6.5.3 using the aforementioned spatial-domain approach. We'll display examples of the final photographs with aliasing and ringing effects turned on so you can judge the visual quality for yourself.

Ratio of Down sampling. Because DCT is a block-based transform, the size of the picture and the block size of the DCT transform are the two criteria that determine the maximum viable ratio for a down-sampling approach in the DCT domain. Here we will look at the DCT transformation of an MN picture to an SS image. Every conceivable combination of down sampling ratios for vertical data is represented by a set of $\boldsymbol{r}_v = \{\frac{i}{\mathrm{M_S}}, i = 1, \cdots, \mathrm{M_S}\}$, while $\boldsymbol{r}_h = \{\frac{j}{\mathrm{N_S}}, j = 1, \cdots, \mathrm{N_S}\}$ includes all possible ratios for horizontally down-sampling, where $\mathrm{M_S} = \frac{\mathrm{M}}{\mathrm{S}}$ and $\mathrm{N_S} = \frac{\mathrm{N}}{\mathrm{S}}$ are the numbers of DCT blocks along the height and the width, respectively.

In case that the vertical scaling ratio and the horizontal scaling ratio are required to be the same 5, the set for all possible ratios is $\boldsymbol{r} = \{\frac{i}{G_{cd}}, i = 1, \cdots, G_{cd}\},$ where the gcd of MS and NS is the largest common divisor. With the LTDS-based approach that has been developed, r may be any ratio.

Furthermore, the suggested technique allows for a mix of any down-sampling ratios, both vertically (rh rh) and horizontally (rv rv). This allows for some wiggle room in supporting a ratio of r / r without introducing any glaring visual distortion by permitting a tiny variation between the horizontal scaling ratio and the vertical scaling ratio. More precisely, the suggested approach does the down-sampling for any ratio r. horizontally by

$r_h = \frac{\mathrm{floor}(r\cdot \mathrm{N_S})}{\mathrm{N_S}}$ and vertically by $r_v = \frac{\mathrm{floor}(r\cdot \mathrm{M_S})}{\mathrm{M_S}}$. When the difference between rh and rv is very tiny, the resulting distortion to the image's proportion is hardly perceptible. And because we can be loose with rh and rv, it's easy to tweak the final picture to fit any screen resolution. To illustrate, take a photo with the original dimensions of 480 by 720 pixels and compare it to the resolution of a common mobile device, which is 240 by 320 pixels and a DCT of 8 by 8. For a full-screen presentation, the suggested approach will scale the picture to an aspect ratio of rv = 2: 1 and rh = 2.25: 1. However, in order to show with a full screen resolution, a technique based on DCT coefficient modification must either remove 80 columns from the original picture or pad 24 blank rows to the image.

Figure 3: Five images used for building up the training set.

EXPERIMENTAL RESULTS

To accommodate users with varying priorities for visual quality over complexity, the suggested design algorithm has been implemented and used to develop a set of down-sampling algorithms in the DCT domain. When comparing the visual quality of various LTDSs, it is useful to have some kind of standard to use as a comparison tool, thus we choose for a spatial-domain approach that uses 10th order Butterworth low-pass filtering and cubic B-spline interpolation to create such pictures. Following this, we evaluate the resulting LTDSs against existing DCT-domain techniques, contrasting them with down-sampling ratios of 2:1 and 3:2.

The effectiveness of the suggested approach for down-sampling with a ratio of 2:1 is shown in Table 1 for three LTDSs. Users' preferences are used to choose which of five photos will serve as the basis for the experiments that determine the best LTDS. $\{\boldsymbol{C}_{256\times 256,i}, i = 1,\cdots,5\}$ as shown in Figure 3. A reference set $\{\boldsymbol{V}_{128\times 128,i}, i = 1,\cdots,5\}$ is constructed using the chosen spatial-domain down-sampling approach. To practice solving (6.13) we first initialize all connections with evenly distributed random integers from [0.5, 0.5]. What we've learned is $\alpha = 1\times 10^{-6},\ \rho = 0.5,\ \lambda = 0.1,$

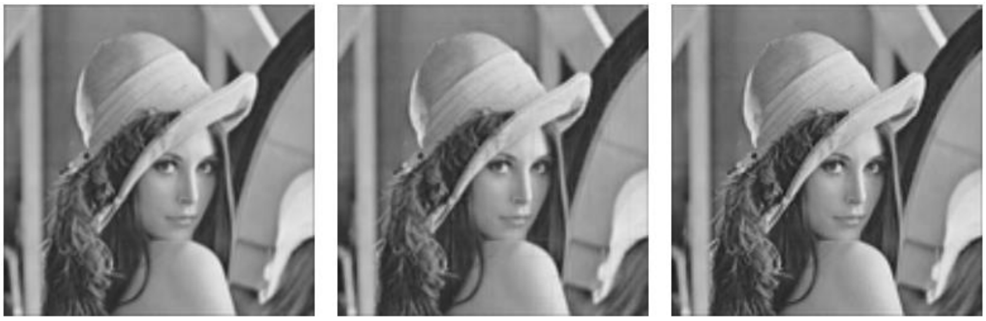
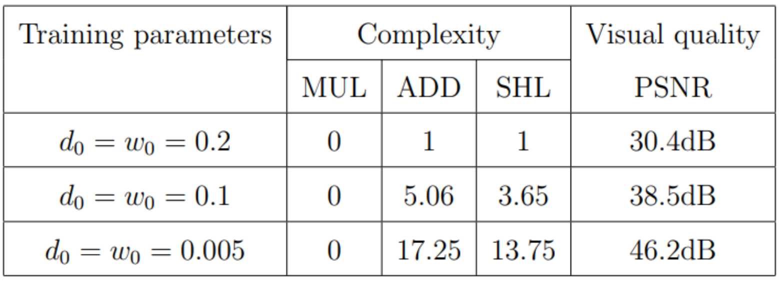


Figure 4: Comparison of visual quality for down sampling “Lena”

Table 1: Results from three different sets of training settings for the proposed approach of down sampling DCT pictures by a factor of 2:1 using LTDS. Reference pictures acquired with the down sampling technique of choice are used to determine PSNR in the spatial domain.

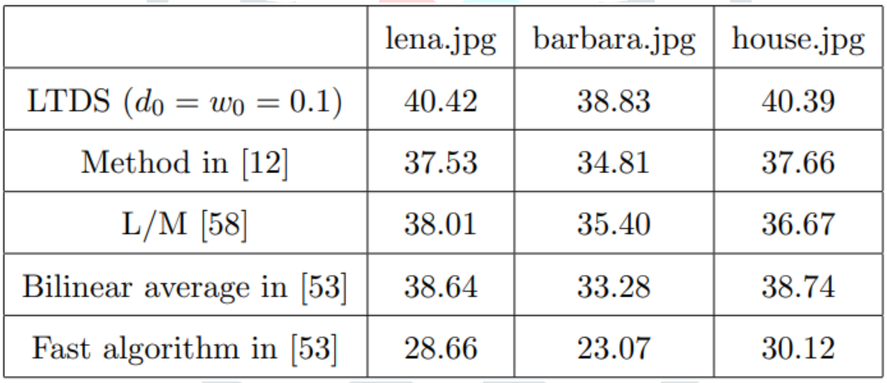


Training parameters	Complexity			Visual quality
	MUL	ADD	SHL	PSNR
$d_0 = w_0 = 0.2$	0	1	1	30.4dB
$d_0 = w_0 = 0.1$	0	5.06	3.65	38.5dB
$d_0 = w_0 = 0.005$	0	17.25	13.75	46.2dB

η = 0.02. Distortion and complexity trade-offs vary depending on the thresholds d0 and w0.

The LTDS with parameters d0 = w0 = 0.1 is preferable for down-sampling since it outperforms competing algorithms in terms of both quality and complexity. Our LTDS for d0 = w0 = 0.1 is compared to the four methods presented in, and for a ratio of 2:1. In, a technique is shown for down-sampling by a factor of 2, again using DCT coefficient modification; in, a technique is presented that is similar in spirit to the former, but it is expanded to enable a wider range of down-sampling ratios. Two methods are included in the work presented in that are optimized for 2:1 down sampling using 8 8 DCT. The first method is effectively an LTDS using bilinear interpolation, where a fresh sample is created by averaging every 22 blocks. Another is a quick approximation approach for the bilinear interpolation technique.

Table 2: Image quality by PSNR for various DCT-domain methods measured against the spatial-domain reference down-sampling method.



	lena.jpg	barbara.jpg	house.jpg
LTDS ($d_0 = w_0 = 0.1$)	40.42	38.83	40.39
Method in [12]	37.53	34.81	37.66
L/M [58]	38.01	35.40	36.67
Bilinear average in [53]	38.64	33.28	38.74
Fast algorithm in [53]	28.66	23.07	30.12

A variety of learning parameters are used to produce an image, and the subjective criteria of aliasing and ringing in the table are used to evaluate the image's visual quality. The absence of standardized reference pictures precludes the use of the PSNR metric. When comparing and contrasting different LTDSs, the reference pictures acquired using the chosen gold-standard approach provide a fair and reliable benchmark. However, due to the optimization of, these references cannot be used to compare our LTDSs with other techniques. Table 2 demonstrates that the produced LTDS (d0 = w0 = 0.1) has a 3–4dB PSNR increase over other approaches, although the down-sampled pictures don't seem noticeably different.

CONCLUSION

This chapter is inspired by the need to transcode pictures and videos; therefore it examines the trade-off between distortion and complexity when down-sampling images and videos in the DCT domain. A down sampling design approach in the discrete cosine transform (DCT) domain has been suggested. The suggested design framework does not rely on a specific spatial-domain technique. Therefore, if future research demonstrates that another spatial-domain technique provides a higher level of visual quality, we would be willing to embrace that method as the standard. Since minimizing computational complexity is a primary motivation for creating DCT-domain down-sampling techniques, we demonstrated that the proposed design framework produces LTDSs that are more effective than other DCT-domain methods in the literature. It strikes an appropriate balance between computational complexity and visual quality.

REFERENCE

1. Dost, S., Saud, F., Shabbir, M. et al. Reduced reference image and video quality assessments: review of methods. J Image Video Proc. 2022, 1 (2022). https://doi.org/10.1186/s13640-021-00578-y

2. Li-Heng Chen, (2022), "Estimating the Resize Parameter in End-to-end Learned Image Compression," arXiv:2204.12022v1 [eess.IV] 26 Apr 2022

3. Vignesh V Menon (2022), "Video Coding Enhancements for HTTP Adaptive Streaming," MM '22, October 10–14, 2022, Lisboa, Portugal © 2022 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-9203-7/22/10. https://doi.org/10.1145/3503161.3548753

4. Dr. Y.L. AJAY KUMAR (2022), "Implementation of Video Compression Based on Spatial-Temporal Resolution Adaptation," Journal of Engineering Sciences Vol 13 Issue 07,July/2022, ISSN:0377-9254

5. Afonso, Mariana et al. "Video Compression Based on Spatio-Temporal Resolution Adaptation." IEEE Transactions on Circuits and Systems for Video Technology 29 (2019): 275-280.

6. M.G. Martini, B. Villarini, F. Fiorucci, A reduced-reference perceptual image and video quality metric based on edge preservation. EURASIP J. Adv. Signal Process. 2012(1), 66 (2012)

7. Z. Wang, A.C. Bovik, Reduced-and no-reference image quality assessment. IEEE Signal Process. Mag. 28(6), 29–40 (2011)

8. L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: A feature similarity index for image quality assessment. IEEE Trans. Image Process. 20(8), 2378–2386 (2011)

9. W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Trans. Image Process. 23(2), 684–695 (2013)

10. H. Liu, C. Li, D. Zhang, Y. Zhou, S. Du, Enhanced image no-reference quality assessment based on colour space distribution. IET Image Process. 14(5), 807–817 (2020)