Low-Complexity Transcoding of JPEG Images with Near-Optimal Quality Using a Predictive Quality Factor and Scaling Parameters

Stéphane Coulombe, Senior Member, IEEE, and Steven Pigeon

***Abstract*—A common transcoding operation consists of reducing the file size of a JPEG image to meet bandwidth or device constraints. This can be achieved by reducing its quality factor (QF) or reducing its resolution, or both. In this paper, using the Structural SIMilarity (SSIM) index as the quality metric, we present a system capable of estimating the QF and scaling parameters to achieve optimal quality while meeting a device's constraints. We then propose a novel low-complexity JPEG transcoding system which delivers near-optimal quality. The system is capable of predicting the best combination of QF and scaling parameters for a wide range of device constraints and viewing conditions. Although its computational complexity is an order of magnitude smaller than the system providing optimal quality, the proposed system yields quality results very similar to those of the optimal system.**

***Index Terms*—JPEG, image transcoding, image adaptation, resolution reduction, file size reduction, image quality, SSIM.**

I. Introduction

The heterogeneous nature of mobile terminals and multimedia applications renders transcoding inevitable [1]. While Multimedia Messaging Services (MMS) require server-side adaptation to ensure interoperability between terminals [2], other applications, such as mobile browsing, will require adaptation of both page layout and media content in order to maximize the user experience (the best compromise between quality and data access time) [3]. The image-related interoperability issues most frequently encountered do not involve image formats, as the majority of the traffic involves JPEG and GIF images, but rather a resolution or file size exceeding the capabilities of the receiving terminal. For instance, in MMS, the limited memory of some mobile devices requires individual images to be under a certain file size or resolution in order to be received and displayed. More specifically, MMS v1.3 defines several content classes in its conformance document with strict maximum file sizes and resolutions [4]: Image Basic (30kB, 160x120), Image Rich (100kB, 640x480), Video Rich (100kB, 640x480), Content Rich (600kB, 1600x1200), etc. In the case of browsing, large file sizes may result in the user waiting an unacceptable length of time to access and display content. Therefore, file size reduction may help reducing latency and increasing the user experience.

Changing an image's resolution, or *scaling*, to meet a terminal's capabilities is a problem with well-known solutions. However, optimizing image quality against file size constraints remains a challenge, as there are no well-established relationships between the quality factor (QF), perceived quality, and the compressed file size. Using arbitrary scaling as an additional means of achieving precise file size reduction, rather than merely resolution adaptation, makes the problem all the more challenging.

Several studies have investigated the problem of file size (or bit rate) reduction for visual content [5]–[11]. Their results show that reduction can be achieved through adaptation of the quantization parameters (either using adaptive prediction based on the n previous frames [12], [13] or two-pass estimation [14] for better rate control) rather than through scaling. When scaling is considered, only a $2:1$ reduction is used, mostly because of the relatively simple compressed domain solutions. For most studies, since they were carried out in the context of low bit rate video, this makes sense as resolution is often limited to a number of predefined formats, several of which are linked by a $2:1$ scaling ratio. However, even in the context of still-picture coding, scaling as an adaptation strategy is not considered. For instance, Ridge [6], who provides

Stéphane Coulombe and Steven Pigeon are with the Department of Software and Information Technology Engineering, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, QC, H3C 1K3 (e-mail: {stephane.coulombe, steven.pigeon}@etsmtl.ca). This work was funded by Vantrix Corporation and by the Natural Sciences and Engineering Research Council of Canada under the Collaborative Research and Development Program (NSERC-CRD 326637-05).

excellent methods for scaling and then reducing the file size of JPEG images, does not consider estimating scaling and quality reduction in combination. We believe this to be a major shortcoming, because the best strategy for maximizing the user experience may well be to scale down the picture and compress it with a higher QF, rather than simply re-compressing it with a lower QF.

In previous work, we first presented an accurate and low-cost method to estimate the resulting compressed file size of a JPEG image subject to scaling and QF changes [15]. We noted that, for a given image, potentially many different scaling and QF combination lead to approximatively the same compressed file size, raising the question as to which combination will maximize the user experience, that is, offer the best perceived transcoded quality, especially considering that scaling could be used to hide some artifacts resulting from a coarser QF. This question was further explored in subsequent work [16], in which we proposed a system where QF and scaling are optimized jointly in order to maximize perceived quality, as measured by the Structural SIMilarity index (SSIM) proposed by Wang *et al.* [17]. Our results showed that, unlike PSNR, using SSIM leads to more subtle trade-offs between quality factor and scaling. In particular, we observed that the system proposed in [16] balances the loss of detail due to scaling and the blocking artifacts introduced by a low quality factor. Indeed, it will select solutions with smaller resolution as the maximum permissible relative size becomes smaller rather than meeting the constraints with very low quality factors leading to conspicuous blocking artifacts.

In this paper, we further investigate methods of combining QF and scaling parameters in JPEG transcoding to meet the terminal's resolution and file size constraints, while at the same time maximizing perceived quality as measured by the SSIM quality metric. We propose two systems: first, the optimal quality JPEG transcoding system (OQJT), capable of providing exact solutions; and second, the near-optimal quality JPEG transcoding system (NOQJT), which yields near-optimal quality using prediction algorithms.

The paper is organized as follows. We first state the transcoding problem in section II. In section III, we show that many QF and scaling combinations yield files of approximately the same size. The two proposed systems are presented in section IV. They are compared in section V in terms of their resulting quality, computational complexity and failure rate. A fully worked out transcoding example is presented in section VI. Finally, section VII concludes the paper.

II. The Transcoding Problem Statement

We now formally define the JPEG image transcoding problem, as well as the notation used in this paper. Let I be a JPEG compressed image and $\mathcal{QF}(I)$, $\mathcal{S}(I)$, $\mathcal{W}(I)$, and $\mathcal{H}(I)$ its quality factor, compressed file size, width, and height respectively. Note that we will assume that the definition of the QF complies with the Independent JPEG Group definition [18]. For a terminal or device D, let $\mathcal{S}(D)$, $\mathcal{W}(D)$, and $\mathcal{H}(D)$ be its maximum permissible compressed data size, image width, and image height respectively ($\mathcal{W}(D)$ and $\mathcal{H}(D)$ are usually larger than the device's screen resolution).

Let $0 < z_T \leq 1$ be an aspect-preserving scaling, or *zoom* factor. A JPEG transcoding operation, denoted $\mathcal{T}(I, QF_T, z_T)$, is the function that returns the compressed image resulting from the application of both the new quality factor QF_T and the scaling parameter z_T to the JPEG image I. A JPEG transcoding operation $\mathcal{T}(I, QF_T, z_T)$ is defined as *feasible* on device D if, for parameters I, QF_T, and z_T, we meet all of the following constraints:

$$\begin{aligned} \mathcal{S}\big(\mathcal{T}(I, QF_T, z_T)\big) &\leq \mathcal{S}(D) \\ z_T\, \mathcal{W}(I) &\leq \mathcal{W}(D) \\ z_T\, \mathcal{H}(I) &\leq \mathcal{H}(D) \end{aligned} \tag{1}$$

We define the following set of feasible JPEG transcoding operations for the image I on device D:

$$\mathcal{F}(I,D) = \big\{(QF_T, z_T) \mid \mathcal{T}(I, QF_T, z_T)) \text{ is feasible on } D\big\}$$

We define $s(I, QF_T, z_T)$, the relative size between the transcoded image and the original image I (which refers to the JPEG image initially received at the transcoder and not the original artifact-free image) as follows:

$$s(I, QF_T, z_T) = \frac{\mathcal{S}\big(\mathcal{T}(I, QF_T, z_T)\big)}{\mathcal{S}(I)}$$

Let $s_{max}(I, D)$ be the maximum acceptable relative size for image I given device D. From eqs. (1) and the fact that we never want to increase the original image's file size, it follows that:

$$s(I, QF_T, z_T) \leq s_{max}(I,D) = \min\left(\frac{\mathcal{S}(D)}{\mathcal{S}(I)}, 1\right) \leq 1 \tag{2}$$

Assuming that several values of QF_T and z_T lead to feasible transcodings, we are interested in finding

$(QF_T^*(I,D), z_{_T}^*(I,D))$, the transcoding parameters that maximize the chosen quality criterion. They are defined as:

$$\big(QF_T^*(I,D), z_{_T}^*(I,D)\big) = \underset{(QF_T, z_{_T}) \in \mathcal{F}(I,D)}{\arg\max} \; \mathcal{Q}\big(I, \mathcal{T}(I, QF_T, z_{_T})\big) \tag{3}$$

where $\mathcal{Q}(I,J)$ is a quality metric using the original image I and the transcoded image J. Ideally, the quality metric would be a measure of the perceived quality of the transcoded image alone (no reference image quality assessment); however, it is more convenient to use a measure of the distortion between the original and transcoded images (full reference image quality assessment).

The optimal transcoded image quality for a given input image I constrained to device D is then defined as:

$$\mathcal{Q}_D^*(I) = \mathcal{Q}\left(I, \mathcal{T}\left(I, QF_T^*(I,D), z_{_T}^*(I,D)\right)\right)$$

The transcoded image corresponding to the optimal parameters is denoted:

$$\mathcal{T}_D^*(I) = \mathcal{T}\left(I, QF_T^*(I,D), z_{_T}^*(I,D)\right)$$

The transcoding parameters $QF_T^*(I,D)$ and $z_{_T}^*(I,D)$ are not necessarily unique, and, amongst the parameters leading to equivalent resulting quality, the parameters that minimize file size may be favored.

III. Predicting Feasible Transcodings

In previous work, we presented methods to estimate the compressed file size of a JPEG image subject to a scaling z and a modification of its QF [15]. One form for this predictor is the following:

$$\hat{\mathcal{S}}\big(\mathcal{T}(I, QF_T, z_{_T})\big) = \mathcal{S}(I)\; M_{\widetilde{QF}_I, \widetilde{QF}_T, \tilde{z}_T}$$

where $\hat{\mathcal{S}}(\mathcal{T}(I, QF_T, z_{_T}))$ is the predicted compressed file size of the transcoded image obtained by applying quality factor QF_T and scaling parameter $z_{_T}$ to the image I. M is a 3-D array, the indices of which are the *quantized* original quality factor $QF_I = \mathcal{QF}(I)$, the desired output quality factor QF_T, and desired scaling $z_{_T}$. In our notation, the tilde ($\sim$) denotes quantized values. Suitable quantization allows the array to be searched efficiently while preventing context dilution [15]. According to this scheme, $M_{\widetilde{QF}_I, \widetilde{QF}_T, \tilde{z}_T}$ represents the *relative* size prediction (the ratio of output to input) for the various values of $\widetilde{QF}_I$, $\widetilde{QF}_T$, and $\tilde{z}_T$. It should be clear that for two different images I and J with the same original quality factor, even if $M_{\widetilde{QF}_I, \widetilde{QF}_T, \tilde{z}_T} \approx M_{\widetilde{QF}_J, \widetilde{QF}_T, \tilde{z}_T}$, the final file size $\mathcal{S}\big(\mathcal{T}(I, QF_T, z_{_T})\big)$ can be radically different from $\mathcal{S}\big(\mathcal{T}(J, QF_T, z_{_T})\big)$ since $\mathcal{S}(I)$ and $\mathcal{S}(J)$ may be independent.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	0.03	0.04	0.05	0.07	0.08	0.10	0.12	0.15	0.17	0.20
20	0.03	0.05	0.07	0.09	0.12	0.15	0.19	0.22	0.26	0.32
30	0.04	0.05	0.08	0.11	0.15	0.19	0.24	0.29	0.34	0.41
40	0.04	0.06	0.09	0.13	0.17	0.22	0.28	0.34	0.40	0.50
50	0.04	0.06	0.10	0.14	0.19	0.25	0.32	0.39	0.46	0.54
60	0.04	0.07	0.11	0.16	0.22	0.28	0.36	0.44	0.53	0.71
70	0.04	0.08	0.13	0.18	0.25	0.33	0.42	0.52	0.63	0.85
80	0.05	0.09	0.15	0.22	0.31	0.41	0.52	0.65	0.78	0.95
90	0.06	0.12	0.21	0.31	0.44	0.59	0.75	0.93	1.12	1.12
100	0.10	0.24	0.47	0.75	1.05	1.46	1.89	2.34	2.86	2.22

Table I
THE SUB-ARRAY $M_{\widetilde{80}, \widetilde{QF}_T, \tilde{z}_T}$, OPTIMIZED FROM THE IMAGE TRAINING SET DESCRIBED IN [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

An example of a sub-array of M optimized over a large image training corpus, $M_{\widetilde{80}, \widetilde{QF}_T, \tilde{z}_T}$, is shown in Table I, with corresponding expected relative absolute error shown in Table II. Throughout this paper, we present the case of $\widetilde{QF}_I = 80$, because it is the most useful, as the majority of JPEG images on the Web are compressed using a QF close to 80. Note that the quantization scheme is not fixed by this algorithm, and we selected the matrix to be 10×10 for illustration purposes. The transcodings were generated using the ImageMagick command line tools, version 6.2.4 [19]. The scaling was performed using the Blackman filter, chosen for its spectral properties [20]. The use of a different filter may lead to different numerical values which are different from those in Table I. This is acceptable as long as the same transcoding tool and filtering parameters are used for training and then operating the system.

Taking the example in section VI, we are looking for solutions with a relative file size of 0.7 or less. Note that a portion of Table I is grayed, showing *nonfeasible* solutions, either because they yield a relative file greater than 0.7, or a resolution exceeding the terminal capabilities (more than 90%, in the example). Examining Table I, we note that various combinations of QF_T and $z_{_T}$ lead to similar size predictions. For instance, $QF_T = 90$ and $z_{_T} = 50\%$ give a relative size prediction of 0.44, which is the same as that of $QF_T = 60$ and $z_{_T} = 80\%$. The best quality must lie at the boundary of the grayed area, since choosing lower QFs in a column or lower scaling in a row can only further reduce quality.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
10	112.9	69.63	48.51	36.74	28.96	24.75	21.36	18.90	17.22	15.70
20	92.75	52.81	35.78	26.65	20.53	17.52	14.93	12.97	11.63	10.23
30	82.23	44.89	30.09	22.07	16.77	14.22	11.90	10.22	8.92	7.55
40	75.74	40.34	26.84	19.52	14.64	12.32	10.15	8.57	7.27	6.45
50	70.74	36.99	24.49	17.70	13.11	10.96	8.88	7.36	6.04	6.32
60	66.28	34.14	22.48	16.19	11.82	9.84	7.81	6.36	5.00	2.40
70	60.75	30.69	20.14	14.46	10.42	8.57	6.61	5.30	4.05	2.40
80	54.08	26.83	17.56	12.65	8.97	7.33	5.55	4.50	3.53	2.42
90	44.44	21.69	14.64	10.83	7.89	6.72	5.72	5.22	4.89	2.88
100	28.84	18.59	16.62	16.17	15.39	15.01	14.70	14.06	13.90	8.39

Table II
THE EXPECTED RELATIVE ABSOLUTE ERROR $\mathrm{E}[|S(I_{out}) - \hat{S}(I_{out})|/S(I_{out})] \times 100\%$ FOR MATRIX $M_{\widetilde{80}}$. THE SHADED REGION CORRESPONDS TO EXPECTED RELATIVE ABSOLUTE ERRORS OF 10% OR LESS.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
10	24.5	25.3	26.0	26.7	27.5	29.1	30.6	32.2	34.6	37.9
20	24.9	25.8	27.1	28.1	29.0	31.2	32.9	34.6	36.9	39.8
30	25.2	26.3	27.8	29.0	29.9	32.4	33.9	35.5	37.4	40.3
40	25.4	26.7	28.5	29.8	30.6	33.1	34.3	35.7	37.0	44.0
50	25.6	27.0	29.0	30.4	31.1	33.7	34.6	35.6	36.2	46.4
60	25.8	27.4	29.6	31.1	31.7	34.3	34.8	35.5	35.2	24.9
70	26.0	27.9	30.5	32.3	32.6	35.2	35.4	35.7	34.8	29.9
80	26.5	28.8	32.1	34.3	34.6	37.6	37.8	38.8	39.2	33.1
90	27.4	30.8	36.2	40.9	43.6	51.0	57.4	65.9	75.7	43.8
100	32.4	54.4	95.6	148	202	272	348	413	500	238

Table III
THE STANDARD ERROR $\left(\mathrm{E}[(s(I_{out}) - \hat{s}(I_{out}))^2] - \mathrm{E}^2[s(I_{out}) - \hat{s}(I_{out})]\right)^{\frac{1}{2}} \times 1000$ FOR MATRIX $M_{\widetilde{80}}$.

IV. QUALITY-AWARE TRANSCODING SYSTEMS

We saw in the previous section that many QF and scaling combinations could lead to similar file sizes. We now need to find the combination that maximizes quality. This requires that we define a quality metric $\mathcal{Q}(I, J)$ to solve eq. (3). Many objective quality measures can be used. In [21], the authors state that JND, SSIM, IFC, and VIF perform much better than the rest of the algorithms (such as the widely used PSNR); VIF being the best in this class. For convenience and without loss of generality, we will use the Structural SIMilarity (SSIM) index proposed by Wang *et al.* to train the quality prediction [17]. More specifically, we are computing the Mean Structural SIMilarity (MSSIM) index to obtain the overall quality measure of an entire image. Since the original image I and the transcoded image J may differ in resolution after adaptation, we will need to scale them to a common resolution before estimating the quality of the resulting image. We propose to scale to a specific resolution based on the viewing conditions—therefore largely determined by the device D—of the transcoded image.

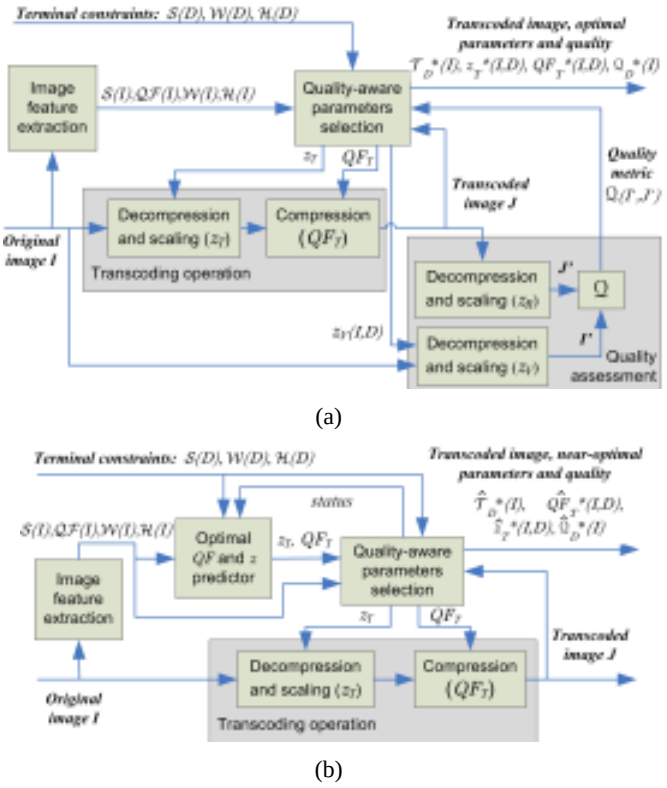

Figure 1. Proposed quality-aware image transcoding systems: (a) Optimal quality JPEG transcoding (OQJT) system (b) Proposed near-optimal quality JPEG transcoding (NOQJT) system.

We propose two quality-aware transcoding systems, shown in Fig. 1. System (a) is designed to provide the optimal solution in terms of quality for a given image while system (b) uses prediction in order to provide a near-optimal solution for a given image with a significant reduction in transcoding computations. Note that the two systems are different from the system presented in previous work [16], as will be made clear in the following pages.

A. Optimal Quality JPEG Transcoding (OQJT) System

In the system illustrated in Fig. 1(a), the set of transcoding parameters leading to optimal quality for a given image are determined. In this system, a quality-aware parameter selection module iterates through various parameters (QF_T and z_T, constrained by image features and terminal characteristics) which are provided to a transcoding engine. The transcoding engine performs transcoding operations based on these parameters. The original and transcoded images are then scaled to a common resolution in order to measure the quality of the transcoded image. These operations

are performed in a quality assessment engine. The quality metric is then provided to the quality-aware parameters selection module for determination of the optimal quality value given a certain parameter resolution (10×10 in our case, i.e. 10 values of QF_T by 10 values of $z_{_T}$). The quality-aware parameter selection module ultimately returns, at the system output, the optimal transcoding parameters $QF_T^*(I,D)$ and $z_{_T}^*(I,D)$, the optimal quality $\mathcal{Q}_D^*(I)$, and the corresponding transcoded image $\mathcal{T}_D^*(I)$. Without any further knowledge, a naïve algorithm would test all possible combinations of Q_T and $z_{_T}$. For instance, for a resolution of 10×10, 100 transcodings would have to be performed. This number is excessive and can be reduced significantly by exploiting several observations. This will be discussed further in section V-C.

In Fig. 1a), we note that both the transcoded and original images are scaled prior to quality evaluation. We define the quality metric comparing, for a viewing condition parameter $z_{_V}$, the original image I and its transcoded version J (using transcoding parameter $z = z_{_T}$) as:

$$\mathcal{Q}_{z_{_V}}(I,J) = MSSIM\left(R(I,z_{_V}), R\left(J, \frac{z_{_V}}{z_{_T}}\right)\right)$$

where $R(I,z)$ is an operator which decompresses I and scales it using scaling factor z. According to this definition and to Fig. 1a), for the image resolutions to be equal, we must have:

$$z_{_V} = z_{_T}\, z_{_R}$$

where $z_{_V} \leq 1$, since we never want to increase the resolution of the original image when comparing quality, and where $z_{_T} \leq 1$ is necessary to meet the terminal constraints. We consider four cases of interest:

Case 1: $z_{_V} = 1$. We compare the images at the resolution of the original image with $z_{_R} = 1/z_{_T}$.

Case 2: $z_{_T} < z_{_V} < 1$. We compare the images at a resolution between that of the original and that of the transcoded image, with $z_{_R} = z_{_V}/z_{_T} > 1$.

Case 3: $z_{_V} < z_{_T} \leq 1$. We compare the images at a resolution smaller than that of the transcoded image, with $z_{_R} = z_{_V}/z_{_T} < 1$.

Case 4: $z_{_V} = z_{_T} < 1$. We compare the images at the resolution of the transcoded image, therefore $z_{_R} = 1$.

The viewing conditions, controlled by parameter $z_{_V}$, play a major role in the user's appreciation of the transcoded results. If the image might be transferred later to another, more capable device (e.g. a PC), the resolution of the original image must be considered for comparison, leading to *Case 1*. *Case 2* would be used when the image is viewed at a resolution between the transcoded resolution and the original resolution: for example, the device's screen resolution, by zooming into the image, or the maximum resolution supported by the device, possibly only accessible by using pan and zoom. *Case 3* would be used when the image is viewed at a resolution smaller than the transcoded resolution: for example, the device's screen resolution, by zooming out of an image transcoded to meet the maximum resolution supported by the device. *Case 4* is not a case of interest, as it would find extremely small images (e.g. 1×1 pixels) acceptable as long as they are similar to the original image scaled at such an extreme resolution.

Table IV shows the distribution of the average MSSIM values $MSSIM_{z_{_V},\widetilde{QF}_{_I},\widetilde{QF}_{_T},\tilde{z}_{_T}}$, for $\widetilde{QF}_{_I} = 80$, computed for *Case 1* over the large image database assembled in [15]. Table VI shows average MSSIM values for *Cases 2* and *3* (combined), where the viewing conditions correspond to a maximum zoom of 90% of the size of the original picture. Table VIII shows average MSSIM values for a maximum zoom of 40%. The full details pertaining to the computation of these tables are presented in previous work [16]. It is not surprising to see that the MSSIM increases with an increase of QF and z, except when $z = 100\%$ where optimal quality is achieved for $QF_T = 80$ (i.e. the same QF as the input image), since no transcoding is required.

Tables V, VII, and IX show the variance for each case. The variance is small enough in the tables to affirm that the optimal quality solutions cannot be too far from those obtained by using the tables of average MSSIM values.

B. Predicting the Optimal Transcoding Parameters

The system described in the previous subsection requires many transcoding operations per image in order to determine the optimal solution using an exact quality criterion that is evaluated at each tentative transcoding. We are seeking a far more computing-efficient transcoding system providing near-optimal quality performance.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	0.12	0.20	0.27	0.33	0.38	0.42	0.46	0.49	0.52	0.55
20	0.14	0.24	0.33	0.40	0.47	0.52	0.56	0.59	0.63	0.66
30	0.15	0.27	0.36	0.44	0.51	0.56	0.61	0.65	0.68	0.73
40	0.16	0.28	0.38	0.47	0.54	0.59	0.64	0.68	0.71	0.77
50	0.17	0.29	0.40	0.49	0.56	0.62	0.67	0.71	0.74	0.79
60	0.17	0.31	0.42	0.51	0.58	0.64	0.69	0.74	0.77	0.86
70	0.18	0.32	0.43	0.53	0.60	0.66	0.71	0.76	0.79	0.92
80	0.19	0.34	0.46	0.56	0.64	0.71	0.76	0.80	0.83	1.00
90	0.21	0.37	0.50	0.61	0.70	0.76	0.81	0.85	0.87	0.98
100	0.23	0.42	0.57	0.69	0.78	0.83	0.88	0.91	0.93	0.99

Table IV
THE SUB-ARRAY $MSSIM_{z_V,\widetilde{80},\widetilde{QF}_T,\tilde{z}_T}$ COMPUTED FOR *Case 1* ($z_V = 100\%$) USING THE IMAGE TRAINING SET FROM [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	59	70	81	91	100	107	113	119	124	131
20	62	73	84	92	98	103	106	109	111	113
30	64	77	87	94	98	101	102	103	103	103
40	64	78	89	96	100	101	102	101	101	100
50	67	81	91	96	98	97	96	93	91	94
60	68	83	92	97	97	95	92	88	84	52
70	69	85	96	100	99	97	93	89	85	57
80	73	90	99	100	95	89	82	74	69	10
90	78	96	104	101	92	83	73	65	60	13
100	87	105	109	101	88	75	63	53	45	9

Table V
THE STANDARD DEVIATION $\times 1000$ OF SUB-ARRAY $MSSIM_{z_V,\widetilde{80},\widetilde{QF}_T,\tilde{z}_T}$ COMPUTED FOR *Case 1* ($z_V = 100\%$) USING THE IMAGE TRAINING SET FROM [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	0.13	0.22	0.30	0.36	0.42	0.46	0.50	0.53	0.55	0.59
20	0.15	0.27	0.36	0.44	0.51	0.56	0.60	0.64	0.66	0.70
30	0.17	0.29	0.40	0.49	0.56	0.61	0.66	0.69	0.72	0.76
40	0.17	0.31	0.42	0.51	0.58	0.64	0.68	0.72	0.75	0.80
50	0.18	0.32	0.44	0.53	0.61	0.67	0.71	0.75	0.78	0.82
60	0.19	0.34	0.46	0.56	0.63	0.69	0.74	0.78	0.81	0.88
70	0.20	0.35	0.48	0.57	0.65	0.71	0.76	0.80	0.83	0.93
80	0.21	0.38	0.51	0.61	0.69	0.75	0.80	0.84	0.87	1.00
90	0.23	0.41	0.55	0.66	0.75	0.81	0.85	0.88	0.91	0.98
100	0.25	0.46	0.62	0.74	0.83	0.88	0.92	0.94	0.96	0.99

Table VI
THE SUB-ARRAY $MSSIM_{z_V,\widetilde{80},\widetilde{QF}_T,\tilde{z}_T}$ COMPUTED FOR *Cases 2* AND *3* (WITH $z_V = 90\%$) USING THE TRAINING SET FROM [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	59	71	84	95	106	113	120	126	130	137
20	62	74	85	94	102	106	110	113	115	117
30	64	77	88	95	100	102	104	105	105	105
40	64	79	90	96	100	102	104	103	102	101
50	67	81	91	96	97	97	96	93	90	94
60	69	83	92	95	95	93	91	87	83	54
70	69	85	96	98	97	95	92	87	83	58
80	74	89	97	96	90	84	77	71	64	9
90	79	95	100	95	84	74	66	60	53	13
100	87	102	101	89	73	60	48	41	37	9

Table VII
THE STANDARD DEVIATION $\times 1000$ OF SUB-ARRAY $MSSIM_{z_V,\widetilde{80},\widetilde{QF}_T,\tilde{z}_T}$ COMPUTED FOR *Cases 2* AND *3* (WITH $z_V = 90\%$) USING THE TRAINING SET FROM [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	0.25	0.43	0.55	0.62	0.69	0.73	0.76	0.79	0.80	0.82
20	0.30	0.52	0.65	0.73	0.79	0.82	0.85	0.87	0.88	0.89
30	0.33	0.56	0.69	0.77	0.83	0.86	0.89	0.90	0.91	0.92
40	0.35	0.58	0.72	0.80	0.85	0.88	0.90	0.92	0.92	0.94
50	0.36	0.61	0.74	0.82	0.87	0.90	0.92	0.93	0.94	0.95
60	0.38	0.63	0.76	0.84	0.89	0.92	0.93	0.94	0.95	0.96
70	0.39	0.65	0.78	0.86	0.90	0.93	0.94	0.95	0.95	0.97
80	0.42	0.68	0.81	0.89	0.93	0.95	0.96	0.96	0.97	1.00
90	0.45	0.72	0.85	0.92	0.95	0.96	0.97	0.97	0.98	0.99
100	0.49	0.78	0.91	0.97	0.98	0.98	0.99	0.99	0.99	1.00

Table VIII
THE SUB-ARRAY $MSSIM_{z_V,\widetilde{80},\widetilde{QF}_T,\tilde{z}_T}$ COMPUTED FOR *Cases 2* AND *3* (WITH $z_V = 40\%$) USING THE TRAINING SET FROM [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

	Scaling, $\tilde{z}_T$, %									
$\widetilde{QF}_T$	10	20	30	40	50	60	70	80	90	100
10	74	93	110	120	125	126	126	125	123	121
20	77	89	98	103	103	101	98	95	91	88
30	79	88	92	93	91	87	83	79	76	72
40	81	89	92	91	88	85	81	76	74	68
50	82	87	86	82	76	72	68	62	60	56
60	84	86	82	77	70	65	59	54	51	36
70	86	88	85	78	71	67	62	58	56	43
80	89	84	75	64	53	48	41	37	33	3
90	94	81	67	53	40	35	31	28	25	6
100	99	72	45	29	19	20	17	17	15	4

Table IX
THE STANDARD DEVIATION $\times 1000$ OF SUB-ARRAY $MSSIM_{z_V,\widetilde{80},\widetilde{QF}_T,\tilde{z}_T}$ COMPUTED FOR *Cases 2* AND *3* (WITH $z_V = 40\%$) USING THE TRAINING SET FROM [15]. SHADING CORRESPONDS TO THE EXAMPLE IN SECTION VI.

Since we have shown that we can obtain a good prediction of the relative file size of an image subject to a change of quality factor and scaling parameters (see Table I with the relative absolute error presented in Table II and the standard error in Table III) as well as the corresponding predicted MSSIM values (as shown in Tables IV-IX) it seems reasonable to think that the same machine learning techniques can be used to predict the optimal set of parameters for given maximum relative file size $s_{max}(I, D)$ and viewing condition $z_v(I, D)$ where $z_v(I, D)$ is the viewing condition z_v given image I and device D. We have $z_v(I, D) \leq z_{max}(I, D)$, where

$z_{max}(I,D)$, the maximum scaling factor, is defined as:

$$z_{max}(I,D) = \min\left(\frac{\mathcal{W}(D)}{\mathcal{W}(I)}, \frac{\mathcal{H}(D)}{\mathcal{H}(I)}, 1\right) \quad (4)$$

We propose to set the viewing conditions to the maximum resolution supported by the terminal or the original size of the image, whichever is the smallest. Therefore, we propose $z_{_V}(I,D) = z_{max}(I,D)$.

Using the image training set T described in [15], we can compute, for every combination of $s_{max}(I,D)$ and $z_{_V}(I,D)$, the average optimal quality factor and scaling values obtained by using the optimal quality system in Fig. 1a). Such values are computed as:

$$\big(\overline{QF}^*_T, \bar{z}^*_T\big) = |\mathsf{T}_{QF_I}|^{-1} \sum_{J \in \mathsf{T}_{QF_I}} \underset{(QF,z)}{\arg\max} \ \mathcal{Q}_{z_{_V}}\big(J, \mathcal{T}(J,QF,z)\big) \quad (5)$$

subject to the simultaneous constraints

$$z \le z_V(I,D)$$
$$s\big(\mathcal{T}(J,QF,z)\big) \le s_{max}(I,D)$$

where $\mathsf{T}_{QF_I} = \{J \in \mathsf{T} \mid QF_J = QF_I\}$, the subset of all images in the training set with the same original QF as I.

It is important to note that the values used for computing eq. (5) are optimal values obtained from the OQJT system (Fig. 1a)) for each image in the training set. Therefore, for each image in the training set and set of constraints (relative file size and viewing condition), the QF_T and $z_{_T}$ parameters leading to the highest quality value are determined. Tables X and XI show the values of $\overline{QF}^*_T$ and $\bar{z}^*_T$ as a function of s_{max} and $z_{_V}$, for our training set. They correspond to the centroids of the optimal solutions in the space of the QF and z parameters. Table XII shows the predicted MSSIM score expressed as a function of s_{max} and $z_{_V}$. Tables X, XI, and XII can be computed for any given density for s_{max}, but only a few are shown here; the variable increments of s_{max} in the tables will be used with the examples presented in Section VI.

C. Proposed near-optimal quality JPEG transcoding (NOQJT) system

We propose to use the system illustrated in Fig. 1b). The system exploits the results described in the previous subsection to estimate the optimal set of transcoding parameters in an optimal QF and z prediction module. The system works as follows:

Viewing condition, $z_{_V}$, %

s_{max}	10	20	30	40	50	60	70	80	90	100
0.05	73.7	42.2	27.1	24.5	23.7	23.4	23.4	23.4	23.4	23.4
0.10	94.2	74.7	48.1	32.7	27.4	25.6	25.1	24.8	24.7	24.7
0.15	98.6	87.9	71.8	50.9	37.9	33.8	31.6	30.4	30.0	29.9
0.20	99.2	91.4	83.3	68.4	51.0	42.4	38.4	36.7	35.3	34.8
0.30	99.2	98.2	90.0	83.5	73.1	61.8	52.0	47.1	44.2	42.5
0.40	99.2	99.7	92.2	89.7	82.2	74.4	65.3	57.1	52.3	48.7
0.50	99.2	99.8	96.6	90.6	89.3	83.7	76.3	68.0	60.9	51.8
0.60	99.2	99.8	99.2	92.7	90.0	88.5	83.0	76.5	69.6	59.7
0.70	99.2	99.8	99.9	95.5	90.6	90.0	87.4	81.2	76.7	67.1
0.80	99.2	99.8	99.9	98.3	92.5	90.2	89.9	86.4	81.4	65.5
0.90	99.2	99.8	99.9	99.6	95.4	90.8	90.1	89.0	83.0	70.8
1.00	99.2	99.8	99.9	99.9	97.8	92.5	90.3	89.9	88.1	79.8

Table X
AVERAGE OPTIMAL QUALITY FACTOR $\overline{QF}^*_T$ FOR $\mathcal{QF}(I) = 80$, AS A FUNCTION OF s_{max} AND $z_{_V}$ USING THE IMAGE TRAINING SET FROM [15].

Viewing condition, $z_{_V}$, %

s_{max}	10	20	30	40	50	60	70	80	90	100
0.05	10.0	18.2	22.7	24.4	25.2	25.4	25.4	25.4	25.4	25.4
0.10	10.0	19.9	29.0	35.4	39.5	41.5	42.2	42.5	42.6	42.7
0.15	10.0	20.0	29.9	39.0	45.7	48.8	51.6	53.4	54.0	54.2
0.20	10.0	20.0	29.9	39.5	48.5	53.4	56.7	59.0	61.2	62.2
0.30	10.0	20.0	29.9	39.9	49.6	57.6	63.9	67.9	71.1	73.2
0.40	10.0	20.0	29.8	39.9	49.5	58.6	67.1	73.4	77.5	80.8
0.50	10.0	20.0	29.8	39.5	49.8	56.7	65.7	74.3	80.9	88.5
0.60	10.0	20.0	29.8	38.6	49.9	57.4	66.7	73.9	81.8	90.6
0.70	10.0	20.0	29.9	38.1	49.5	59.8	64.3	77.6	82.6	90.2
0.80	10.0	20.0	30.0	38.2	48.1	59.7	68.5	73.6	85.1	97.2
0.90	10.0	20.0	30.0	38.9	46.6	59.0	69.7	73.7	85.6	99.9
1.00	10.0	20.0	30.0	39.6	46.1	57.0	69.3	78.8	81.8	100

Table XI
AVERAGE OPTIMAL SCALING $\bar{z}^*_T$ (IN %), FOR $\mathcal{QF}(I) = 80$, AS A FUNCTION OF s_{max} AND $z_{_V}$ USING THE IMAGE TRAINING SET FROM [15].

Viewing condition, $z_{_V}$, %

s_{max}	10	20	30	40	50	60	70	80	90	100
0.05	0.15	0.20	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
0.10	0.24	0.36	0.42	0.44	0.45	0.45	0.45	0.45	0.45	0.45
0.15	0.23	0.36	0.44	0.48	0.50	0.51	0.51	0.51	0.51	0.51
0.20	0.25	0.41	0.52	0.58	0.61	0.61	0.62	0.62	0.62	0.62
0.30	0.25	0.45	0.55	0.63	0.67	0.69	0.70	0.70	0.71	0.71
0.40	0.25	0.46	0.56	0.66	0.70	0.73	0.75	0.76	0.76	0.76
0.50	0.25	0.46	0.59	0.66	0.74	0.76	0.78	0.79	0.80	0.81
0.60	0.25	0.46	0.61	0.67	0.75	0.79	0.81	0.81	0.82	0.84
0.70	0.25	0.46	0.62	0.69	0.75	0.81	0.82	0.84	0.84	0.86
0.80	0.25	0.46	0.62	0.71	0.75	0.81	0.85	0.85	0.86	0.89
0.90	0.25	0.46	0.62	0.73	0.76	0.81	0.85	0.86	0.87	0.94
1.00	0.25	0.46	0.62	0.74	0.78	0.81	0.85	0.88	0.89	0.99

Table XII
AVERAGE MSSIM OF OPTIMAL SOLUTIONS OBTAINED FROM SYSTEM OQJT (USED AS $\hat{\mathcal{Q}}^*_D$) USING THE IMAGE TRAINING SET FROM [15].

Step 1: For given image I and device D, compute $s_{max} = s_{max}(I,D)$ using eq. (2) and $z_{_V} = z_{_V}(I,D) = z_{max}(I,D)$ using eq. (4).

Step 2: Obtain $\overline{QF}^*_T(s_{max},z_{_V})$ and $\bar{z}^*_T(s_{max},z_{_V})$ from

Tables X and XI respectively—let us assume that $\mathcal{QF}(I) = 80$ to match the tables. Set $QF_T = \overline{QF}^*_T(s_{max},z_{_V})$ and $z_{_T} = \bar{z}^*_T(s_{max}, z_{_V})$. Note that we take the nearest smaller values in the tables if the desired s_{max} and $z_{_V}$ are not present.

Step 3: Transcode the image with the quality factor and scaling parameters QF_T and $z_{_T}$ respectively.

Step 4: If the file size of the transcoded image is too large (i.e. if $s(I, QF_T, z_{_T}) > s_{max}$), identify the quality factor and scaling parameters (QF_T and $z_{_T}$) corresponding to the next smaller value of s_{max} in the Tables X and XI and go to step 3 (actually try a smaller value of s_{max} from the tables than previously tried until a different set of QF_T and $z_{_T}$ is obtained). Otherwise go to step 5.

Step 5: Return near-optimal parameters $\widehat{QF}^*_T(I, D) = QF_T$ and $\hat{z}^*_T(I, D) = z_{_T}$, the transcoded image using these parameters $\hat{\mathcal{T}}^*_D(I) = \mathcal{T}(I, QF_T, z_{_T})$, and predicted transcoded image quality $\hat{\mathcal{Q}}^*_D(I) = \mathcal{Q}_{z_{_V}}(I, \hat{\mathcal{T}}^*_D(I))$ (using Table XII).

The NOQJT system differs from the OQJT systems in one major way: the optimal transcoding parameters are predicted rather than searched iteratively. However, in both systems, the transcoded image is always validated in order to ensure that it meets the transcoding constraints.

The two systems presented in this paper differ from those previously presented. For example, the first system we presented was capable of predicting the transcoded file size only, relying on quantization to speed up computations, as well as to minimize memory usage [15]. The first of the systems presented in this paper, the OQJT system, is capable of yielding the optimal decision for a given picture, but at a greater cost as it explores the parameter space without prediction. The system presented in [16] predicts the optimal transcoding parameters QF^*_T and $z^*_{_T}$ that maximize quality (as predicted by MSSIM or PSNR). This system uses a version of eq. (5) where the constraints are also predicted using the method presented in [15]. The parameters search is iterative in the set of predicted feasible solutions. We also showed that while the system behaves differently depending on whether PSNR or MSSIM is used as a quality metric, it takes reasonable decisions in both cases. The system in [16] (as in [15]) makes extensive use of parameter quantization to reduce memory requirements. Finally, the second system we propose in this paper, the NOQJT system, uses prediction to directly estimate the optimal parameters $\overline{QF}^*_T$ and $\bar{z}^*_T$ while maximizing the perceived quality at the same time as satisfying the constraints of maximal file size (s_{max}) and viewing conditions ($z_{_V}$). In the next section, section V, we show that the performance of the proposed NOQJT system is very good; we get near-optimal MSSIM values (i.e. close to that of the OQJT system) with significantly reduced complexity.

V. Simulation Results

In this section, we compare the OQJT system with the NOQJT system with respect to quality (using SSIM), computation complexity, and failure rate. Again, we concentrate on the case where $\mathcal{QF}(I) = 80$. The tables will be shown for s_{max} from 0.1 to 1.0 by steps of 0.1. However, in our simulations, we used tables with s_{max} from 0.05 to 0.1 by steps of 0.025, and from 0.1 to 1.0 by steps of 0.05. Even for $\overline{QF}^*_T$ and $\bar{z}^*_T$ tables (like those of Tables X and XI), we used the same resolution.

A. Quality using SSIM

For each set of constraints s_{max} and $z_{_V}$, we transcoded each image of the test set described in [15] to meet those constraints. The average MSSIM values ($\overline{MSSIM}^{\text{OQJT}}_{s_{max},z_{_V}}$) obtained using the OQJT system are provided in Table XIII. For the same set of constraints and for the same test images, we transcoded the images with the proposed NOQJT system. The average MSSIM values obtained are provided in Table XIV. Note that, for convenience, the values of QF and z used for transcoding were rounded to the nearest value corresponding to the resolution of our transcoded image database (we used a parameters resolution of 10×10, as in Table I). The average error (in %) and the variance of the average error between the two systems are presented in Table XV and Table XVI. The average error is relatively small throughout the table, but increases as the value of s_{max} decreases (i.e. as we move further away from the image's initial file size). Since, for any given test image, the NOQJT solution's MSSIM can at best equal that of the OQJT solution, the average absolute MSSIM error over each image of the test set equals the error between the average MSSIM of both systems. Indeed, we have:

	Viewing condition, z_V, %									
s_{max}	10	20	30	40	50	60	70	80	90	100
0.1	0.24	0.37	0.42	0.44	0.45	0.45	0.46	0.46	0.46	0.46
0.2	0.25	0.42	0.53	0.58	0.61	0.62	0.62	0.63	0.63	0.63
0.3	0.25	0.45	0.56	0.63	0.68	0.70	0.71	0.71	0.71	0.71
0.4	0.25	0.46	0.57	0.66	0.71	0.74	0.75	0.76	0.76	0.77
0.5	0.25	0.46	0.59	0.67	0.75	0.76	0.78	0.79	0.80	0.81
0.6	0.25	0.46	0.62	0.67	0.75	0.79	0.81	0.82	0.83	0.84
0.7	0.25	0.46	0.62	0.69	0.75	0.81	0.82	0.84	0.85	0.86
0.8	0.25	0.46	0.62	0.71	0.76	0.81	0.85	0.86	0.87	0.89
0.9	0.25	0.46	0.62	0.73	0.77	0.81	0.85	0.87	0.87	0.94
1.0	0.25	0.46	0.62	0.74	0.78	0.81	0.85	0.88	0.89	1.00

Table XIII
AVERAGE MSSIM ($\overline{MSSIM}^{\mathrm{OQJT}}_{s_{max},z_V}$) OF OPTIMAL SOLUTIONS OBTAINED WITH THE OQJT SYSTEM USING THE IMAGE TEST SET FROM [15]).

	Viewing condition, z_V, %									
s_{max}	10	20	30	40	50	60	70	80	90	100
0.1	0.23	0.33	0.35	0.36	0.38	0.37	0.37	0.37	0.37	0.37
0.2	0.25	0.41	0.52	0.56	0.56	0.60	0.56	0.57	0.57	0.57
0.3	0.25	0.44	0.55	0.63	0.67	0.68	0.69	0.68	0.68	0.67
0.4	0.25	0.46	0.56	0.66	0.70	0.73	0.73	0.75	0.74	0.75
0.5	0.25	0.46	0.58	0.66	0.74	0.76	0.77	0.79	0.79	0.80
0.6	0.25	0.46	0.61	0.67	0.75	0.78	0.80	0.81	0.82	0.84
0.7	0.25	0.46	0.62	0.68	0.75	0.80	0.82	0.84	0.84	0.86
0.8	0.25	0.46	0.62	0.70	0.75	0.81	0.84	0.85	0.86	0.89
0.9	0.25	0.46	0.62	0.72	0.76	0.81	0.85	0.86	0.87	0.94
1.0	0.25	0.46	0.62	0.73	0.77	0.81	0.85	0.88	0.88	0.99

Table XIV
AVERAGE MSSIM ($\overline{MSSIM}^{\mathrm{NOQJT}}_{s_{max},z_V}$) OF SOLUTIONS OBTAINED WITH THE NOQJT SYSTEM USING THE IMAGE TEST SET FROM [15].

$$\begin{aligned}\frac{1}{|\mathsf{T}|}\sum_{I\in\mathsf{T}}\left|\mathcal{Q}^{\mathrm{OQJT}}_{s_{max},z_V}(I)-\mathcal{Q}^{\mathrm{NOQJT}(I)}_{s_{max},z_V}\right| \\ =\frac{1}{|\mathsf{T}|}\sum_{I\in\mathsf{T}}\mathcal{Q}^{\mathrm{OQJT}}_{s_{max},z_V}(I)-\frac{1}{|\mathsf{T}|}\sum_{I\in\mathsf{T}}\mathcal{Q}^{\mathrm{NOQJT}}_{s_{max},z_V}(I) \\ =\overline{MSSIM}^{\mathrm{OQJT}}_{s_{max},z_V}-\overline{MSSIM}^{\mathrm{NOQJT}}_{s_{max},z_V}\end{aligned} \quad (6)$$

where T represents the image test set, $\mathcal{Q}_{s_{max},z_V}(I)=\mathcal{Q}_{z_V}(I,\mathcal{T}(I,QF^*_{s_{max},z_V}(I),z^*_{s_{max},z_V}(I)))$ the quality of the optimal solution given I, s_{max} and z_V (for a given method).

Not only the expected resulting quality from the two systems are closely matched (and accordingly the expected difference small), the respective rankings of solutions from both systems are highly correlated. Analysing the correlation between solution orderings from best to worst given the constraints in both systems shows a very high rank correllation. The average Spearman ρ rank correlation coefficient is found to be $\rho \approx 0.999661$, a result of high significance [22]. A Student t-distribution test estimates the probability that both rankings are the same as being essentially 1.

We also note that the system is capable of accounting for blocking artifacts, favoring smaller pictures with higher QFs having better MSSIM, over larger pictures, more crudely compressed pictures, as revealed in Table XI. For example, taking Table XI, with $z_V = 80\%$, we see that the optimal scaling z_T goes from 78.8% to 42.5% as the maximum allowable relative file size goes from 1.0 to 0.1.

B. The Effects of Quantization

In our simulation, for each image from the training set, the resulting MSSIM and file size were obtained for a quantized parameter grid, where parameters are quantized against a 10×10 grid. For a given original QF, solving eq. (3) using the quantized tables means that $\overline{QF}^*_T$ and $\bar{z}^*_T$ are approximated by $\widetilde{\overline{QF}}^*_T$ and $\tilde{\bar{z}}^*_T$, obtained using quantized quality and scaling factors—for example, 76.2 would become 80. Since prediction is only available through the quantized parameters, $\widetilde{\overline{QF}}^*_T$ and $\tilde{\bar{z}}^*_T$

	Viewing condition, z_V, %									
s_{max}	10	20	30	40	50	60	70	80	90	100
0.1	1.33	3.00	3.45	4.70	5.51	5.85	5.93	5.94	5.95	5.95
0.2	0.03	0.71	2.11	1.83	2.34	3.94	4.41	4.63	4.80	4.84
0.3	0.02	0.01	0.15	1.94	1.88	1.40	3.97	4.26	4.46	2.70
0.4	0.02	0.00	1.11	0.06	1.10	1.99	0.86	2.09	3.61	2.80
0.5	0.02	0.00	0.03	0.23	0.04	0.76	0.72	2.94	2.14	2.90
0.6	0.02	0.00	0.03	0.93	0.06	0.35	0.61	1.42	2.24	3.15
0.7	0.02	0.00	0.02	0.46	0.22	0.07	1.33	0.35	0.80	3.16
0.8	0.02	0.00	0.01	0.46	0.65	0.07	0.05	0.51	2.61	5.66
0.9	0.02	0.00	0.01	0.33	0.78	0.21	0.05	1.24	3.28	1.13
1.0	0.02	0.00	0.00	0.19	1.06	0.54	0.09	0.05	0.71	0.00

Table XV
AVERAGE ERROR $\times 100$ BETWEEN THE AVERAGE MSSIMS OF THE OQJT AND NOQJT SYSTEMS (SAME AS THE AVERAGE ABSOLUTE MSSIM ERROR OVER EACH IMAGE) USING THE IMAGE TEST SET FROM [15].

	Viewing condition, z_V, %									
s_{max}	10	20	30	40	50	60	70	80	90	100
0.1	1.62	3.53	4.05	4.86	4.92	5.09	5.19	5.20	5.23	5.23
0.2	0.49	1.67	2.60	2.91	3.38	4.13	3.31	3.44	3.61	3.67
0.3	0.43	0.33	0.81	2.70	3.05	2.48	3.38	3.65	3.85	4.26
0.4	0.42	0.10	2.29	0.55	2.35	3.33	2.19	2.44	2.55	3.48
0.5	0.42	0.10	0.27	1.00	0.56	1.75	2.69	3.58	2.61	3.46
0.6	0.42	0.10	0.32	1.99	0.64	0.93	1.40	1.99	3.55	3.43
0.7	0.42	0.10	0.28	1.35	1.02	0.66	2.48	1.22	1.73	5.24
0.8	0.42	0.10	0.24	1.40	1.60	0.63	0.51	1.33	1.91	5.41
0.9	0.42	0.10	0.24	1.27	1.67	1.05	0.58	2.04	1.91	3.99
1.0	0.42	0.10	0.08	1.03	2.09	1.44	0.68	0.58	1.59	0.05

Table XVI
AVERAGE STANDARD DEVIATION OF ERROR $\times 100$ BETWEEN THE AVERAGE MSSIMS OF THE OQJT AND NOQJT SYSTEMS (SAME AS THE AVERAGE ABSOLUTE MSSIM ERROR OVER EACH IMAGE) USING THE IMAGE TEST SET FROM [15].

are necessarily rounded to the nearest available solution. Our simulations show that if truncation is used, the size and quality predictions are necessarily pessimistic and fewer retries are made as file size prediction is systematically undershot. If rounding is used, simulations show that more accurate, yet possibly optimistic, quality prediction is achieved, although this may result in more retries as the prediction may overshoot file size. However, it is preferable to use rounding rather than truncation, as rounding enhances user experience, even at the cost of a few retries.

C. Computational Complexity

Finding the set of parameters leading to optimal quality in the OQJT system can be compared to the search for optimal motion vectors performed in video coding [23]. We have a grid size of $N \times N$ points on which a quality metric needs to be optimized assuming that the function is convex or near-convex. If we set the MSSIM of an image to -1 (which is the worst possible score) when the device constraints are not met, we have such a situation; that is, low QF and z values lead to small MSSIM values, while large QF and z values may not meet the device constraints. Therefore the optimal solutions lie somewhere in between. An exhaustive search method evaluates $N \times N$ values (i.e. N^2 transcoding operations in our case). The most efficient methods (excluding predictive methods such as PMVFAST and EPZS that use information from surrounding motion blocks, which does not apply to our case) have a complexity of the order of $\log(N)$. For instance, the Three Step Search algorithm evaluates $1 + 3 \times 8 = 25$ points [24], [25]. If the number of points to be examined can be reduced by excluding solutions that exceed the device's resolution, the number of steps in the search algorithm will nonetheless increase with parameter resolution, while a finer grid will require a deeper search.

By contrast, the proposed NOQJT system requires, on average, fewer than two transcoding operations per image, as shown in Table XVII. Interestingly, the complexity of the NOQJT system is likely to grow very slowly as the parameters resolution increases (as, on average, very few transcodings will be performed), while the complexity of the OQJT system will increase as the parameter resolutions increases. This means an important speedup for the NOQJT system, which will be at least 10 times as fast as the OQJT system. The maximum number of transcoding operations (i.e. in the worst case scenario) depends on the grid size. Grid size and other parameters from the NOQJT framework can be adapted for different application scenarios, providing various compromises between optimality of the visual quality and computational complexity (e.g. we could increase the average visual quality with an increase of average complexity by performing a *ceil* on the $\overline{QF}_T^*$ in Table X instead of rounding).

	Viewing condition, z_V, %									
s_{max}	10	20	30	40	50	60	70	80	90	100
0.1	1.16	1.57	1.78	1.90	1.77	1.85	1.85	1.85	1.85	1.85
0.2	1.02	1.03	1.10	1.36	1.57	1.20	1.79	1.79	1.79	1.79
0.3	1.00	1.25	1.03	1.04	1.09	1.43	1.16	1.74	1.74	1.99
0.4	1.00	1.03	1.00	1.04	1.03	1.08	1.85	1.14	1.88	1.46
0.5	1.00	1.01	1.36	1.01	1.13	1.04	1.79	1.03	1.15	1.17
0.6	1.00	1.00	1.24	1.00	1.02	1.77	1.10	1.04	1.06	1.08
0.7	1.00	1.00	1.08	1.62	1.01	1.08	1.05	1.08	1.15	1.07
0.8	1.00	1.00	1.02	1.94	1.00	1.02	1.30	1.16	1.03	1.01
0.9	1.00	1.00	1.02	1.62	1.79	1.01	1.05	1.04	1.01	1.02
1.0	1.00	1.00	1.00	1.36	1.80	1.00	1.02	1.19	1.19	1.01

Table XVII
AVERAGE NUMBER OF TRANSCODING OPERATIONS PER IMAGE WITH THE PROPOSED NOQJT SYSTEM USING THE IMAGE TEST SET FROM [15].

	Viewing condition, z_V, %									
s_{max}	10	20	30	40	50	60	70	80	90	100
0.1	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
0.2–1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table XVIII
TRANSCODING FAILURE RATE (IN %) WITH THE PROPOSED NOQJT SYSTEM USING THE IMAGE TEST SET FROM [15].

D. Failure Rate

An important aspect of a transcoding system to study is the failure rate, i.e. how often the system cannot find a solution to the constraints. We observe from Table XVIII that the system only fails for cases where $s_{max} = 0.1$, which is the smallest value shown in the table. We have this problem because the parameter set used in our simulations is limited to scaling factors of 10% or more. In practice, this could be easily solved by reducing the scaling values until the constraints are met, leading to no failure at all. However, it can be argued that, in such a situation, the right thing to do is to fail since the returned image may be of no use as its resolution could be too small. For instance, it may not make much sense to reduce a 640x480 image to below 64x48 to fit a very small file size constraint.

VI. A TRANSCODING EXAMPLE

Consider a device with $\mathcal{S}(D) = 30500$, $\mathcal{W}(D) = 640$, $\mathcal{H}(D) = 480$, and an image, Lena, with $\mathcal{S}(I) = 43266$,

$\mathcal{W}(I) = 512$, $\mathcal{H}(I) = 512$ and $\mathcal{QF}(I) = 80$.

Step 1: We compute $s_{max} = \min\left({}^{30500}\!/_{43266}, 1\right) \approx 0.7$ and $z_{_V} = z_{max}(I, D) = \min\left(\frac{640}{512}, \frac{480}{512}, 1\right) \approx 90\%$.

Step 2: Using Table X, we find that $QF_{_T} = \overline{QF}^*_{_T}(0.7, 0.9) = 76.7$, which we quantize to 80. Using Table XI, we find that $z_{_T} = \bar{z}^*_T(0.7, 0.9) = 82.6\%$, which we quantize to 80%.

Step 3: Transcode the image with the quality factor and scaling parameters $QF_{_T}$ and $z_{_T}$ respectively. After transcoding, we obtain an image with $s(I, 80, 80\%) = 0.68$ (see Table XIX).

Step 4: Since the transcoded image meets the target file size (0.68< 0.70), we go to step 5.

Step 5: Return near-optimal parameters $\widehat{QF}^*_{_T}(I, D) = 80$ and $\hat{z}^*_T(I, D) = 80\%$, the transcoded image using the parameters $\hat{\mathcal{T}}^*_D(I) = \mathcal{T}(I, 80, 80\%)$, and predicted transcoded image quality $\hat{\mathcal{Q}}^*_D(I) = \mathcal{Q}_{z_{_V}}(I, \hat{\mathcal{T}}^*_D(I)) = 0.84$ from Table XII (with $z_{_V} = 0.9$) while the true image quality from Table XX is 0.86 (a $\sim 2\%$ error).

Looking at Tables XIX and XX, the actual results from transcodings on the Lena image, the OQJT system would select, after several transcoding iterations, $QF^*_{_T}(I, D) = 80$ and $z^*_{_T}(I, D) = 80\%$ leading to a relative file size of 0.68 and a quality of $\mathcal{Q}^*_D(I) = 0.86$. Therefore, for Lena under these constraints, we obtained an image with optimal quality with NOQJT with a single transcoding operation.

Let us consider a second, more extreme, example, the results of which are shown in Fig. 2. Let us keep the viewing conditions $z_{_V} = 90\%$, but set the maximum relative filesize to $s_{max} = 0.2$. Reapplying the procedure, we find that $\overline{QF}^*_{_T} = 35.3 \approx 40$ and $\bar{z}^*_T = 61.2\% \approx 60\%$. We get $s(I, 40, 60\%) = 0.22 \geqslant 0.2$, which is not an acceptable solution, and the algorithm retries different parameters. Reducing s_{max} to 0.15, the tables yield $QF_{_T} = 30.0$, $z_{_T} = 54.0\% \approx 50\%$. The new relative size is $s(I, 30, 50) = 0.13$, which is now an acceptable solution. Therefore, $\widehat{QF}^*_{_T} = 30$, $\hat{z}^*_T = 50\%$ and $\hat{\mathcal{Q}}^*_D = 0.51$. After transcoding we find that $\mathcal{Q}_D = 0.58$. The OQJT system finds $QF^*_T = 50$ and $z_{_T} = 50\%$ yielding an image quality of 0.63, which is close to the 0.58 resulting from the NOQJT system. However, we can observe that, as shown in figures II and VII, the system becomes less precise as the target relative file sizes become smaller. Still, we managed to obtain a good solution in 2 transcodings only.

	Scaling, $z_{_T}$, %									
$QF_{_T}$	10	20	30	40	50	60	70	80	90	100
10	0.01	0.02	0.03	0.05	0.07	0.09	0.11	0.13	0.15	0.18
20	0.01	0.03	0.05	0.07	0.10	0.14	0.17	0.21	0.25	0.29
30	0.02	0.04	0.07	0.10	0.13	0.18	0.23	0.28	0.33	0.39
40	0.02	0.04	0.08	0.12	0.16	0.22	0.27	0.33	0.40	0.48
50	0.02	0.05	0.09	0.13	0.18	0.25	0.31	0.39	0.46	0.51
60	0.02	0.05	0.10	0.15	0.21	0.28	0.36	0.44	0.54	0.73
70	0.03	0.06	0.11	0.18	0.25	0.34	0.43	0.53	0.65	0.89
80	0.03	0.08	0.14	0.22	0.31	0.42	0.54	0.68	0.82	1.00
90	0.04	0.11	0.20	0.31	0.44	0.62	0.79	0.99	1.20	1.18
100	0.08	0.24	0.47	0.79	1.15	1.63	2.10	2.63	3.23	2.48

Table XIX
ACTUAL RELATIVE FILE SIZES FOR LENA, $QF_I = 80$, AS A FUNCTION OF QF_T AND $z_{_T}$.

Using the same size contraints, let us compare with a simple algorithm that first scales to the viewing conditions then lowers QF_T until the target size is met. For the first example, we have $z_{_V} = z_{_T} = 90\%$ and $s_{max} = 0.7$. Using Table XIX, we can observe that the simple algorithm finds $QF_T = 70$, resulting in a MSSIM of 0.85 instead of 0.86 as found by the OQJT and NOQJT algorithms. It also managed to perform this with a single trancoding (but it is higher in general). For the second example, we have $z_{_V} = z_{_T} = 90\%$ and $s_{max} = 0.2$. After 7 transcodings, the algorithm will find the feasible solution $z_{_T} = 90\%$ and $QF_T = 10$ leading to $s(I, 10, 90\%) = 0.15$ and an MSSIM of 0.53 (while we have 0.58 for NOQJT and 0.63 for OQJT). Not only this simple algorithm yields an inferior quality to compared with NOQJT but the computational complexity is significantly higher. This shows how resilient the NOQJT system is, and how it compares favorably with both the OQJT system and a naïve approach wherein only the QF is adapted after an initial scaling to the viewing conditions. The NOQJT system can provide visual results close to those of the OQJT system with impressive improvement in computational complexity.

VII. CONCLUSIONS

In this paper, we analyzed the impact of various combinations of QF and scaling parameter values on the quality of transcoded images. Using SSIM, we showed how quality varies with quality factor QF_T and scaling $z_{_T}$ for various viewing conditions. We also proposed two quality-aware transcoding systems:

(a)

(b)

(c)

(d)

Figure 2. Transcoded Lena (details) showing solutions from the example in section VI, with a maximum relative file size of 0.2 and viewing condition z_V=90% (color enhanced and scaled for display purposes): (a) original image, with $QF_I = 80$; (b) solution from the NOQJT system with $\widehat{QF}_T^* = 30$ and $\hat{z}_T^* = 50\%$; (c) optimal solution using the OQJT system, with $QF_T^* = 50$ and $z_T^* = 50\%$; (d) naïve solution with $QF_T = 10$ and $z_T = 90\%$.

an optimal quality JPEG transcoding (OQJT) system and a near-optimal quality JPEG transcoding (NOQJT) system. We compared the two systems with respect to quality, computation complexity, and failure rate. The NOQJT system yields very similar quality as OQJT with a complexity up to 25 times smaller than that of the

	Scaling, z_T, %									
QF_T	10	20	30	40	50	60	70	80	90	100
10	0.15	0.27	0.33	0.39	0.44	0.47	0.50	0.52	0.53	0.56
20	0.19	0.31	0.41	0.47	0.53	0.57	0.60	0.63	0.65	0.67
30	0.21	0.35	0.44	0.52	0.58	0.62	0.65	0.68	0.71	0.74
40	0.23	0.37	0.47	0.55	0.61	0.65	0.69	0.72	0.75	0.79
50	0.24	0.39	0.50	0.57	0.63	0.68	0.72	0.75	0.78	0.79
60	0.25	0.40	0.51	0.60	0.65	0.70	0.74	0.78	0.81	0.88
70	0.26	0.42	0.54	0.62	0.68	0.73	0.78	0.82	0.85	0.96
80	0.28	0.45	0.57	0.65	0.72	0.77	0.82	0.86	0.88	1.00
90	0.30	0.49	0.61	0.70	0.77	0.83	0.88	0.91	0.93	0.99
100	0.31	0.53	0.68	0.80	0.88	0.93	0.96	0.97	0.98	0.99

Table XX
ACTUAL MSSIM FOR LENA, WITH $QF_I = 80$, AS A FUNCTION OF QF_T AND z_T, UNDER VIEWING CONDITIONS $z_V = 90\%$

OQJT system and performs an average between 1 and 2 transcoding operations per image. We have shown that the failure rate can be made to be arbitrarily close to zero. The proposed framework can be adapted to various applications scenarios.

REFERENCES

[1] S. Chandra and C. Schlatter Ellis, "JPEG compression metric as a quality aware transcoding," in *Unix Symposium on Internet Technologies and Systems*, 1999.

[2] S. Coulombe and G. Grassel, "Multimedia adaptation for the multimedia messaging service," *IEEE Communications Magazine*, vol. 42, no. 7, pp. 120–126, July 2004.

[3] R. Han, P. Bhagwat, R. LaMaire, T. Mummert, V. Perret, and J. Rubas, "Dynamic adaptation in an image transcoding proxy for mobile web browsing," *IEEE Personnal Communications*, vol. 5, no. 6, pp. 8–17, December 1998.

[4] Open Mobile Alliance, "MMS conformance document," *Candidate version 1.3, OMA-TS-MMS-CONF-V1_3-20080128-C*, Jan. 2008.

[5] Z. Lei and N.D. Georganas, "Accurate bit allocation and rate control for DCT domain video transcoding," in *IEEE CCECE 2002. Canadian Conference on Electrical and Computer Engineering*, 2002, vol. 2, pp. 968–973.

[6] J. Ridge, "Efficient transform-domain size and resolution reduction of images," *Signal Processing: Image Communication*, vol. 18, no. 8, pp. 621–639, Sept. 2003.

[7] V. Ratnakar and V. Ivashin, "File size bounded JPEG transcoder," May 2001, US Patent 6,233,359.

[8] S.-F. Chang and D. G. Messerschmitt, "Manipulation and compositing of MC-DCT compressed video," *IEEE Journal of Selected Areas in Communications*, vol. 13, no. 1, 1994.

[9] Y. Wang, J.-G. Kim, S.-F. Chang, and H.-M. Kim, "Utility-based video adaptation for universal media access (UMA) and content-based utility function prediction for real-time video transcoding," *IEEE Transactions on Multimedia*, vol. 13, no. 9, pp. 213–220, Sept. 2007.

[10] P. Yin, M. Wu, and B. Liu, "Video transcoding by reducing spatial resolution," *Procs. IEEE Int. Conference on Image Processing*, vol. 1, pp. I–972–I–975, 2000.

[11] T. Shanableh and M. Ghanbari, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats," *IEEE Transactions on Multimedia*, vol. 2, no. 2, pp. 101–110, June 2000.

[12] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits and Systems For Video Technology*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[13] A. Vetro, P. Yin, and H. Sun, "Reduced spatio-temporal transcoding using an intra refresh technique," *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, vol. 4, pp. IV–723–IV–726, 2002.

[14] P. Yin and J. Boyce, "A new rate control scheme for H.264 video coding," *Procs. IEEE Int. Conference on Image Processing*, pp. I–449–I–452, 2004.

[15] S. Pigeon and S. Coulombe, "Computationally efficient algorithms for predicting the file size of JPEG images subject to changes of quality factor and scaling," in *Proceedings of the 24th Queen's Biennial Symposium on Communications*, Queen's University, Kingston, Canada, June 2008, pp. 378–382.

[16] S. Coulombe and S. Pigeon, "Quality-aware selection of quality factor and scaling parameters in jpeg image transcoding," in *IEEE Symposium on Computational Intelligence for Multimedia Signal and Vision Processing*, Mar. 2009, pp. 68–74.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[18] T. Lane, P. Gladstone, L. Ortiz, J. Boucher, L. Crocker, J. Minguillon, G. Phillips, D. Rossi, and G. Weijers, "The independent JPEG group software release 6b," 1998.

[19] "ImageMagick command-line tools," http://www.imagemagick.org/script/index.php.

[20] R. B. Blackman and J.W. Tukey, "Particular pairs of windows," in *The Measurement of Power Spectra, from the Point of View of Communications Engineering*. 1959, pp. 95–100, Dover.

[21] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[22] C. Spearman, "The proof and measurements of association between two things," *Amer. J. Psycho.*, vol. 15, pp. 72–101, 1904.

[23] Y. Wang, Y.-Q. Zhang, and J. Ostermann, *Video Processing and Communications*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001.

[24] R. Li, B. Zeng, and M. Liou, "A new three-step search algorithm for block-motion estimation," *IEEE Trans. Circuits and Systems For Video Technology*, vol. 4, no. 4, pp. 438–442, Aug. 1994.

[25] H.-M. Hang, Y.-M. Chou, and S.-C. Cheng, "Motion estimation for video coding standards," *Journal of VLSI Signal Processing*, vol. 17, pp. 113–136, 1997.