

# Quality-aware selection of quality factor and scaling parameters in JPEG image transcoding

Stéphane Coulombe and Steven Pigeon

**Abstract**—Reducing the file size of a JPEG image to meet bandwidth or terminal constraints is a common transcoding operation. The reduction can be achieved by reducing either the quality factor (QF) or the resolution, or both. In this paper, we analyze the impact of QF and scaling parameter choices on the quality of the resulting images, as measured by a quality metric such as the Structural SIMilarity index (SSIM). We propose a quality-aware transcoding system which considers the quality of transcoded images when QF and scaling are selected jointly. Its goal is to select QF and scaling parameters that maximize the user experience under a given viewing condition, as measured by the chosen quality metric.

## I. INTRODUCTION

THE heterogeneous nature of mobile terminals and multimedia applications renders transcoding inevitable [1]. Multimedia Messaging Services (MMS), for example, require server-side adaptation to ensure interoperability between terminals [2]. The most frequent image-related interoperability issues 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, the limited memory of some mobile phones requires images to be under a certain size or resolution in order to be received and displayed.

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 scaling as an additional means of achieving 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 [3]–[5]. Their results show that reduction can be achieved through adaptation of the quantization parameters, rather than through scaling. For most studies, this makes sense, since they were carried out in the context of low bit rate video, where resolution is often limited to a number of predefined formats. However, even in the context of still-picture coding, scaling as an adaptation

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, H3E 1K6 (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).

strategy is not considered. For instance, Ridge [4], who provides 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 have presented methods to estimate the compressed file size of a JPEG image subject to scaling and QF changes [6]. We note that several combinations of QF and scaling lead to the same approximate file size, raising the question of which combination would maximize user experience, estimated by an objective quality metric.

In this paper, we investigate combining QF and scaling parameters in JPEG transcoding to meet the terminal's resolution and file size constraints, while at the same time maximizing quality.

## 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  $QF(I)$ ,  $S(I)$ ,  $W(I)$ , and  $H(I)$  its quality factor, compressed file size, width, and height respectively. Note that we will assume that the QF complies with the definition put forward by the Independent JPEG Group [7]. For a terminal or device  $D$ , let  $S(D)$ ,  $W(D)$ , and  $H(D)$  be its maximum permissible compressed file size, image width, and image height respectively ( $W(D)$  and  $H(D)$  are usually larger than the device's screen size).

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

$$\begin{aligned} S(T(I, QF_{out}, z)) &\leq S(D) \\ z W(I) &\leq W(D) \\ z H(I) &\leq H(D) \end{aligned}$$

We define the following set of feasible JPEG transcoding operations for the image  $I$ :

$$\mathcal{F}(I, D) = \{(QF_{out}, z) \mid T(I, QF_{out}, z) \text{ is feasible on } D\}$$

Assuming that several values of  $QF_{out}$  and  $z$  lead to feasible transcodings for device  $D$ , we are interested in finding  $(QF_{out}^*(I, D), z^*(I, D))$ , the values of  $QF_{out}$  and  $z$  that maximize the transcoded quality of  $I$  under a certain criterion  $\mathcal{Q}$ . These are given by:

$$(QF_{out}^*(I, D), z^*(I, D)) = \arg \max_{(QF_{out}, z) \in \mathcal{F}(I, D)} \mathcal{Q}(I, T(I, QF_{out}, z)) \quad (1)$$

where  $\mathcal{Q}(I, J)$  is a quality metric using the original image  $I$  and the transcoded image  $J$ . Ideally, the resulting transcoding quality would be assessed from the transcoded image alone, however, it may be more convenient to use a measure of distortion between the original and the transcoded images.

### 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 [6]. One form for this predictor is the following:

$$\hat{S}(I, QF_{out}, z) = S(I) M_{\widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}}$$

where  $\hat{S}(I, QF_{out}, z)$  is the predicted compressed file size of the transcoded image obtained by applying quality factor  $QF_{out}$  and scaling parameter  $z$  to the image  $I$ . The array  $M$ , computed off-line, is indexed by the *quantized* original quality factor  $\widetilde{QF}_{in}$  ( $\widetilde{QF}_{in}$  is to be understood as  $\mathcal{Q}\mathcal{F}(I)$ , the original QF of  $I$ ),  $QF_{out}$ , and  $z$ . We will be using the tilde ( $\sim$ ) to denote quantized values. Suitable quantization allows the array to be searched efficiently while preventing context dilution [6]. After deciding on appropriate quantization levels  $\widetilde{QF}_{in}$ ,  $\widetilde{QF}_{out}$ , and  $\tilde{z}$ , the entry  $M_{\widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}}$  is given by:

$$M_{\widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}} = |\mathbb{T}_{\widetilde{QF}_{in}}|^{-1} \sum_{J \in \mathbb{T}_{\widetilde{QF}_{in}}} s(J, \widetilde{QF}_{out}, \tilde{z}) \quad (2)$$

where  $\mathbb{T}_{\widetilde{QF}_{in}}$  is the subset of all images  $J$  in the training set  $\mathbb{T}$  (a large corpus of representative images gathered in [6]) such that  $QF_{in}(J)$  quantizes to  $\widetilde{QF}_{in}$ , and where  $|\mathbb{X}|$  denotes the cardinality of set  $\mathbb{X}$ . The function  $s(J, QF_{out}, z)$  is defined as

$$s(J, QF_{out}, z) = \frac{S(T(J, QF_{out}, z))}{S(J)}$$

that is, the relative size between the transcoded image  $J$ , to which both the quality factor  $QF_{out}$  and scaling  $z$  are applied, and the original image  $J$ .

Therefore, according to this scheme,  $M_{\widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}}$  is the relative size prediction (output versus input) for the various values of  $\widetilde{QF}_{in}$ ,  $\widetilde{QF}_{out}$ , and  $\tilde{z}$ . An example of such a sub-array,  $M_{\widetilde{80}, \widetilde{QF}_{out}, \tilde{z}}$ , thus obtained is shown in Table I. We present an example with  $\widetilde{QF}_{in} = 80$ , because it is the most useful, as the majority of images found on the Web have been compressed using a  $QF$  close to 80. Note that the quantization scheme is not fixed by this algorithm, and

$\widetilde{QF}_{out}$	Scaling, $\tilde{z}$									
	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  $M_{\widetilde{80}, \widetilde{QF}_{out}, \tilde{z}}$ , OPTIMIZED FROM THE IMAGE TRAINING SET DESCRIBED IN [6]. SHADING CORRESPONDS TO THE EXAMPLES IN SECTION VII.

we selected a simple quantization so that the matrix is of dimension  $10 \times 10$  (using  $\widetilde{QF}_{out} \in \{10, 20, \dots, 100\}$  and  $\tilde{z} \in \{10\%, 20\%, \dots, 100\%\}$ ) for illustration purposes only.

Taking the examples in section VII, a portion of Table I is grayed, indicating that the solutions are *nonfeasible*, either because a relative file size is larger than 0.25 (light or dark gray) or larger than 0.50 (dark gray). Examining Table I, we note that various combinations of  $QF_{out}$  and  $z$  lead to similar size predictions. For instance,  $QF_{out} = 90$  and  $z = 50\%$  give a prediction of 0.44, which is the same as that for  $QF_{out} = 60$  and  $z = 80\%$ . The best quality transcodings necessarily lie at the boundary of the grayed area, since lower QFs in a column or lower scaling in a row can only further reduce quality. In addition, we may have to reject some solutions because they exceed the maximum resolution of the receiving terminal.

### IV. PROPOSED QUALITY-AWARE TRANSCODING SYSTEM

If many QF and scaling combinations lead to similar file sizes, then we need to select the combination that maximizes the user experience. This requires that we define the quality metric  $\mathcal{Q}(I, J)$  to solve eq. (1). Many choices are possible, as it has been shown that JND, SSIM (MS), IFC, and VIF all perform much better than the widely used PSNR, the best performer being VIF [8]. For convenience and without loss of generality, we will use the Structural SIMilarity (SSIM) index proposed by Wang *et al.* [9]. 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 both to a specific resolution based on the viewing conditions—largely determined by the device  $D$ —of the transcoded image.

Let us now propose a novel quality-aware transcoding system, as shown in Fig. 1. The system is composed of several modules. A *quality-aware parameter selection* module provides estimates of optimal target parameters  $z_T = \hat{z}^*(I, D)$  and  $QF_T = \hat{QF}_{out}^*(I, D)$  for the *transcoding operation* module based on information ( $\hat{S}$  and  $\hat{Q}$ ) provided

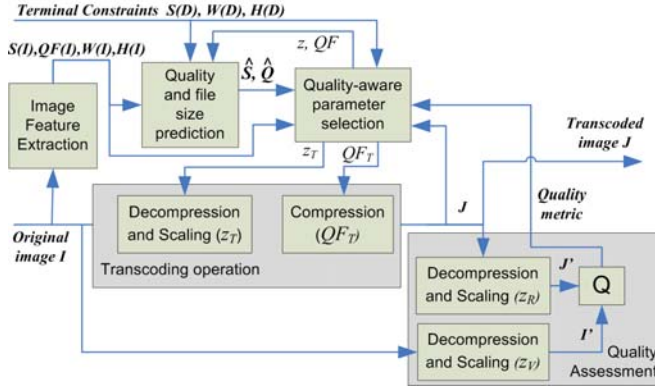


Fig. 1. Proposed quality-aware image transcoding system.

by the *file size and quality prediction* module, terminal constraints, the original image's features (obtained from the *image feature extraction* module), and the actual transcoded image. The latter is required in case the predictors do not perform well and the file size of the transcoded image exceeds the constraints. In that case, the *quality-aware parameter selection* module will select a new set of parameters until a valid transcoding is achieved.

The quality of the transcoded image may be measured in an optional *quality assessment* module to validate that the solution's quality is indeed satisfactory. Scaling operations may be used, in the quality assessment module, to scale both the original and the transcoded image to a common resolution to apply the quality metric, in our case the SSIM.

The quality-aware parameter selection module will select the best parameters for the terminal constraints:

$$(QF_{out}^*(I, D), z^*(I, D)) = \arg \max_{(QF_{out}, z) \in \mathcal{F}(I, D)} SSIM(I', J')$$

where  $I'$  and  $J'$  are the original and transcoded images, both scaled using factors  $z_V$  and  $z_R$  respectively. According to Fig. 1, 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 original image's resolution when comparing quality, and where  $z_T \leq 1$  to meet the terminal constraints. We consider three 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 image and that of the transcoded image, with  $z_R = z_V/z_T > 1$ .
- Case 3:*  $z_V = z_T < 1$ . We compare the images at the resolution of the transcoded image, therefore  $z_R = 1$ .

The viewing conditions, controlled by parameter  $z_V$  ( $z_T \leq z_V \leq 1$ ), play a major role in the user's appreciation of the transcoded results. If the image might be transferred to another, more capable device later (e.g. a PC), the resolution of the original image must be considered (*Case 1*). *Case 2* applies when the image is viewed at a resolution between that of the transcoded image and that of the original image; for example, the maximum resolution (which may be accessible only through pan and zoom) is supported by the device. *Case 3* applies to the case where the image will only be viewed on the terminal. However, in *Case 3*, one must provide for instances where the transcoded image is of degenerate size. Unless thumbnailing is desired, *Case 3* cannot completely account for image quality in the case where the transcoded image is very small, and so is to be avoided.

If the image is to be optimized for viewing on the terminal only, then we propose that the viewing condition scaling  $z_V$  be set to the maximum resolution supported by the terminal, which may exceed its screen resolution.

## V. PREDICTING THE QUALITY OF TRANSCODED IMAGES

As with the file size predictor presented in section III, we would like to have a quality predictor which can be used in the proposed quality-aware transcoding system. Therefore, using a similar approach to the one described in [6] for file size prediction, we derive a quality predictor.

We want to obtain the quality of the transcoded picture following application of a new QF  $1 \leq QF_{out} \leq 100$  and scaling factor  $0 < z \leq 1$  when the viewing parameter  $z_V$  is applied for image comparison. Let

$$Q_{z_V}(I, QF_{out}, z) = SSIM \left( R(I, z_V), R \left( T(I, QF_{out}, z), \frac{z_V}{z} \right) \right) \quad (3)$$

be the viewing condition aware quality function, where  $R(J, u)$  is an operator which decompresses  $J$  and scales it using scaling factor  $u$ ,  $QF_{out}$  is the desired output QF,  $z$  is the desired scaling factor, and  $SSIM(I, J)$  compares the two images scaled at a common resolution.

As the function  $Q_{z_V}$  is expensive to compute for all combinations of  $QF_{out}$  and  $z$ , it should rather be *predicted* from a precomputed array of possible parameter sets. Let this be precomputed into an array  $SSIM$ , the indices of which are the *quantized* viewing condition  $z_V$ , the original QF ( $QF_{in}$ ), the transcoded QF ( $QF_{out}$ ), and the scaling factor  $z$ . In our experiments, we used the quantized values  $\{10, 20, \dots, 100\}$  for  $\widetilde{QF}_{in}$  and  $\widetilde{QF}_{out}$  and  $\{0.1, 0.2, \dots, 1.0\}$  for  $\widetilde{z}$  and  $\widetilde{z}_V$ .

According to this scheme, the quality prediction for quantized input  $\widetilde{QF}_{in}$ , quantized desired output  $\widetilde{QF}_{out}$ , and

$\widetilde{QF}_{out}$	Scaling, $\tilde{z}$ , %									
	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	<b>0.77</b>
50	0.17	0.29	0.40	0.49	0.56	<b>0.62</b>	0.67	0.71	0.74	0.79
60	0.17	0.31	0.42	0.51	0.59	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	0.93
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	0.99
100	0.23	0.42	0.57	0.69	0.78	0.83	0.88	0.91	0.93	0.99

TABLE II

THE SUBARRAY  $SSIM_{\tilde{z}_V, \tilde{z}, \widetilde{QF}_{out}, \tilde{z}}$  COMPUTED FOR *Case 1* ( $z_V = 1$ ) USING THE IMAGE TRAINING SET FROM [6]. SHADING CORRESPONDS TO THE EXAMPLES IN SECTION VII.

$\widetilde{QF}_{out}$	Scaling, $\tilde{z}$ , %									
	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	0.51	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	<b>0.89</b>	0.93	0.95	0.96	0.96	0.97	1.00
90	0.45	0.72	0.85	<b>0.92</b>	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	0.98	1.00

TABLE III

THE SUBARRAY  $SSIM_{\tilde{z}_V, \tilde{z}, \widetilde{QF}_{out}, \tilde{z}}$  COMPUTED FOR *Case 2* (WITH  $z_V = 40\%$ ) USING THE TRAINING SET FROM [6]. SHADING CORRESPONDS TO THE EXAMPLES IN SECTION VII.

$\widetilde{QF}_{out}$	Scaling, $\tilde{z}$ , %									
	10	20	30	40	50	60	70	80	90	100
10	5.8	6.8	8.1	9.1	10.1	10.8	11.5	12.1	12.5	13.2
20	6.0	7.2	8.4	9.3	9.9	10.4	10.8	11.1	11.2	11.5
30	6.2	7.5	8.7	9.5	9.9	10.1	10.3	10.4	10.4	10.4
40	6.3	7.7	8.9	9.6	10.0	10.1	10.2	10.1	10.0	9.9
50	6.4	7.9	9.0	9.6	9.8	9.7	9.6	9.3	9.1	9.5
60	6.6	8.2	9.3	9.7	9.7	9.5	9.2	8.8	8.5	5.2
70	6.8	8.5	9.6	10.0	9.8	9.6	9.1	8.6	8.3	5.1
80	7.2	8.9	9.9	10.0	9.4	8.9	8.1	7.4	6.9	0.9
90	7.6	9.6	10.4	10.1	9.1	8.2	7.2	6.4	5.9	1.2
100	8.4	10.4	10.9	10.0	8.6	7.2	6.1	5.1	4.4	0.8

TABLE IV

THE STANDARD DEVIATION ( $\times 100$ ) OF SUBARRAY  $SSIM_{\tilde{z}_V, \tilde{z}, \widetilde{QF}_{out}, \tilde{z}}$  COMPUTED FOR *Case 1* ( $z_V = 1$ ) USING THE IMAGE TRAINING SET FROM [6].

$\widetilde{QF}_{out}$	Scaling, $\tilde{z}$ , %									
	10	20	30	40	50	60	70	80	90	100
10	7.4	9.3	11.0	12.0	12.5	12.6	12.6	12.5	12.3	12.1
20	7.7	8.9	9.8	10.3	10.3	10.1	9.8	9.5	9.1	8.8
30	7.9	8.8	9.2	9.3	9.1	8.7	8.3	7.9	7.6	7.2
40	8.1	8.9	9.2	9.1	8.8	8.5	8.1	7.6	7.4	6.8
50	8.2	8.7	8.6	8.2	7.6	7.2	6.8	6.2	6.0	5.6
60	8.4	8.6	8.2	7.7	7.0	6.5	5.9	5.4	5.1	3.6
70	8.6	8.8	8.5	7.8	7.1	6.7	6.2	5.8	5.6	4.3
80	8.9	8.4	7.5	6.4	5.3	4.8	4.1	3.7	3.3	0.3
90	9.4	8.1	6.7	5.3	4.0	3.5	3.1	2.8	2.5	0.6
100	9.9	7.2	4.5	2.9	1.9	2.0	1.7	1.7	1.5	0.4

TABLE V

THE STANDARD DEVIATION ( $\times 100$ ) OF SUBARRAY  $SSIM_{\tilde{z}_V, \tilde{z}, \widetilde{QF}_{out}, \tilde{z}}$  COMPUTED FOR *Case 2* (WITH  $z_V = 40\%$ ) USING THE TRAINING SET FROM [6].

quantized scalings  $\tilde{z}$  and  $\tilde{z}_V$  is given by:

$$SSIM_{\tilde{z}_V, \widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}} = |\mathbb{T}_{\widetilde{QF}_{in}}|^{-1} \sum_{J \in \mathbb{T}_{\widetilde{QF}_{in}}} Q_{\tilde{z}_V} \left( J, \widetilde{QF}_{out}, \tilde{z} \right) \quad (4)$$

where  $\mathbb{T}_{\widetilde{QF}_{in}}$  is defined as previously and  $Q_{\tilde{z}_V}$  is given by eq. (3). Accordingly,  $SSIM_{\tilde{z}_V, \widetilde{QF}_{in}}$  denotes a slice of that array, a matrix, with indices  $\widetilde{QF}_{out}$  and  $\tilde{z}$ . The predictor  $\hat{Q}_{z_V}$  for (3) now becomes:

$$\hat{Q}_{z_V}(I, QF_{out}, z) = SSIM_{\tilde{z}_V, \widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}} \quad (5)$$

Tables II and III show the distribution of the average SSIM values  $SSIM_{\tilde{z}_V, \widetilde{QF}_{in}, \widetilde{QF}_{out}, \tilde{z}}$ , for  $\widetilde{QF}_{in} = 80$ , computed for *Cases 1* and *2* over the large image database assembled in [6]. Table III shows average SSIM values for *Case 2*, where the viewing condition scaling corresponds to a maximum zoom of 40% of the size of the original picture ( $z_V = 40\%$ ). The values obtained in Tables II and III are quite reliable. Indeed, their standard deviations, presented in Tables IV and V, are quite low. Therefore, any element of these SSIM tables derived for various  $z_V$  can be used reliably as quality estimators  $\hat{Q}$  in the proposed transcoding system.

## VI. THE TRANSCODING PROCESS

In the proposed quality-aware transcoding system illustrated in Fig. 1, the parameter selection process would be as follows: the quality-aware parameter selection module selects, from all the predicted feasible transcodings, the one providing the best predicted quality for the metric in use. Feasible solutions are predicted based on terminal constraints such as the display resolution and predicted file size from Table I. Quality could be estimated from Table II (for  $z_V = 100\%$ ) or Table III (for  $z_V = 40\%$ ) (given that  $QF(I) = 80$ ), or from another pre-computed table derived from the actual  $QF(I)$  and viewing condition scaling  $z_V$ . After the optimal parameters have been estimated, they are used to perform an actual transcoding. The transcoded image is then validated. If its file size is satisfactory, then the transcoded image is returned. Otherwise, another transcoding operation is performed with a smaller transcoded file size target.

Alternatively, the file size prediction may be used to identify the  $(QF_{out}, z)$  pairs lying around the boundary (on both sides of the boundary) of feasible solutions, and evaluating, for each, the actual quality of the actually transcoded image. The transcoded image with the highest quality (while meeting the terminal's constraints) would then be returned as the system's output. This second option

involves more computation, but ensures that the best possible parameters for a given image are selected.

## VII. TRANSCODING EXAMPLES

Let us consider a first example inspired from the MMS application. Let  $D$  be a device with the remaining message length (excluding headers and message structure) of  $S(D) = 11000$  bytes, with a maximum resolution of  $\mathcal{W}(D) = 1024$ ,  $\mathcal{H}(D) = 768$ , and an image to be sent, Lena, with  $S(I) = 43266$ ,  $\mathcal{W}(I) = 512$ ,  $\mathcal{H}(I) = 512$  and  $QF(I) = 80$ . The maximum acceptable scaling factor is  $z = 100\%$  since both  $\mathcal{W}(D) \geq \mathcal{W}(I)$  and  $\mathcal{H}(D) \geq \mathcal{H}(I)$ . The maximum relative file size is  $\frac{11000}{43266} \approx 0.25$ .

Using Table I, we determine which parameters are part of the feasible set  $\mathcal{F}(I, D)$ . In Table I, the nonfeasible solutions—those with a relative size larger than 0.25—are shaded in light or dark gray. Considering *Case 1*, where the transcoded image is scaled up to be compared to the original image, and using Table II, we find that the parameters that maximize the SSIM quality metric are  $QF_{out} = 50$  and  $z = 60\%$ , yielding a predicted SSIM score of 0.62 (see the bold value in the table of optimal solutions for the various examples presented here). Now, looking at Tables VI and VII, which provide the true file sizes and SSIM values for Lena (obtained by performing actual transcodings on Lena), we find that the parameters that maximize the SSIM quality metric are  $QF_{out} = 70$  and  $z = 50\%$ , yielding an SSIM score of 0.61. The solution  $QF_{out} = 50$  and  $z = 60\%$  is a close second, yielding an SSIM score of 0.60. Therefore, the solution obtained using the predictors and the predicted quality are quite good. Using the PSNR as the quality metric (now using Table IX), we obtain a very different solution:  $QF_{out} = 10$  and  $z = 100\%$ , for a PSNR of 26.2 dB. Figure 2 shows the visual results of each solution. The PSNR solution in Figure 2(c) has the highest resolution but the smallest QF, and exhibits an excessive number of blocking artifacts. The solution in Figure 2(d) has a very low resolution but the highest QF, and is quite blurry. The SSIM solution in Figure 2(b) seems to be the best compromise and appears to provide the best perceptual quality.

Now consider a maximum relative size of 0.50 with the other constraints the same as previously. Then, we find, using Table I and Table II (where the nonfeasible solutions are shaded with dark gray), that the parameters that maximize the SSIM quality metric are  $QF_{out} = 40$  and  $z = 100\%$ , yielding a predicted SSIM score of 0.77. Now, looking at Tables VI and VII, we find that many combinations of parameters maximize the SSIM quality metric for Lena (any set of values on the diagonal from  $(QF_{out} = 90, z = 50\%)$  to  $(QF_{out} = 40, z = 100\%)$ ); all of them yielding an SSIM score of 0.71. Therefore, our predicted solution is among the optimal solutions, although our predicted SSIM is not as good as expected. Using the PSNR as the quality metric, we obtain the same solution:

$QF_{out}$	Scaling, $z$ , %									
	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 VI  
RELATIVE FILE SIZES FOR LENA,  $QF_{in} = 80$ .

$QF_{out}$	Scaling, $z$ , %									
	10	20	30	40	50	60	70	80	90	100
10	0.14	0.23	0.28	0.33	0.37	0.39	0.42	0.43	0.45	0.46
20	0.16	0.27	0.34	0.40	0.46	0.49	0.52	0.55	0.57	0.58
30	0.18	0.30	0.38	0.45	0.50	0.54	0.58	0.61	0.63	0.66
40	0.20	0.32	0.41	0.48	0.53	0.58	0.61	0.65	0.67	0.71
50	0.21	0.34	0.43	0.51	0.56	0.60	0.64	0.68	<b>0.71</b>	0.72
60	0.23	0.35	0.45	0.53	0.58	0.63	0.67	<b>0.71</b>	0.74	0.84
70	0.23	0.37	0.48	0.56	<b>0.61</b>	0.66	<b>0.71</b>	0.75	0.78	0.93
80	0.24	0.39	0.51	0.59	0.65	<b>0.71</b>	0.75	0.80	0.83	1.00
90	0.27	0.44	0.55	0.65	<b>0.71</b>	0.78	0.82	0.86	0.89	0.98
100	0.29	0.49	0.64	0.76	0.84	0.89	0.92	0.94	0.96	0.99

TABLE VII  
SSIM FOR LENA,  $QF_{in} = 80$ ,  $z_V = 100\%$ .

$QF_{out}$	Scaling, $z$ , %									
	10	20	30	40	50	60	70	80	90	100
10	0.28	0.44	0.53	0.59	0.64	0.68	0.71	0.73	0.75	0.76
20	0.33	0.53	0.63	0.69	0.76	0.80	0.82	0.84	0.85	0.85
30	0.37	0.58	0.69	0.75	0.81	0.84	0.86	0.88	0.89	0.90
40	0.40	0.61	0.72	0.78	0.84	0.87	0.89	0.91	0.92	0.93
50	0.43	0.63	0.74	0.81	0.86	0.89	0.91	0.93	0.93	0.93
60	0.45	0.65	0.77	0.83	0.88	0.91	0.92	0.94	0.95	0.96
70	0.47	0.68	0.80	0.86	0.90	0.93	0.94	0.95	0.96	0.98
80	0.49	0.71	0.83	<b>0.88</b>	0.92	0.95	0.96	0.97	0.97	1.00
90	0.54	0.77	0.87	<b>0.92</b>	0.95	0.97	0.98	0.98	0.98	0.99
100	0.58	0.83	0.94	0.98	0.99	0.99	0.99	0.99	0.99	1.00

TABLE VIII  
SSIM FOR LENA,  $QF_{in} = 80$ ,  $z_V = 40\%$ .

$QF_{out} = 40$  and  $z = 100\%$ , for a PSNR of 31.9 dB.

Let us reconsider two examples, but now with a new device  $D$  such that  $\mathcal{H}(D) = 320$  and  $\mathcal{W}(D) = 240$ , leading to a maximum allowable scaling  $z$  of 40% (since  $z_{max} = \min(\frac{320}{512}, \frac{240}{512}) \approx 47\%$ ). The viewing conditions, determined by the device, suggest comparison at  $z_V = 40\%$ . Let us first consider the case where the maximum relative file size is 0.25. Using the SSIM computed for a viewing condition of  $z_V = 40\%$ , as shown in Table III, we find that  $QF_{out} = 80$  and  $z = 40\%$  for an SSIM of 0.89. Now, looking at Tables VI and VIII for Lena, we find the same solution,  $QF_{out} = 80$  and  $z = 40\%$ , yields an SSIM score of 0.88, which is very close to the predicted value of 0.89. If we are using the PSNR instead (and Table X), we obtain the same solution, for a PSNR of 34.3 dB.

$\widetilde{QF}_{out}$	Scaling, $\bar{z}$ , %									
	10	20	30	40	50	60	70	80	90	100
10	17.3	19.3	20.6	21.6	22.1	23.0	23.5	24.0	24.4	<b>26.2</b>
20	17.8	20.1	21.6	22.7	23.2	24.4	25.1	25.8	26.4	28.7
30	18.0	20.4	22.0	23.2	23.7	25.1	25.9	26.7	27.4	30.2
40	18.1	20.6	22.2	23.5	23.9	25.5	26.3	27.3	28.1	<b>31.9</b>
50	18.2	20.7	22.4	23.7	24.1	25.7	26.7	27.7	28.6	32.5
60	18.4	20.8	22.6	23.9	24.2	26.0	27.0	28.1	29.1	33.0
70	18.4	21.0	22.7	24.1	24.4	26.3	27.3	28.6	29.7	37.3
80	18.4	21.1	22.9	24.4	24.6	26.6	27.8	29.3	30.6	54.9
90	18.6	21.3	23.2	24.7	24.9	27.1	28.3	30.1	31.6	48.0
100	18.7	21.5	23.4	25.0	25.1	27.5	28.8	30.7	32.2	51.4

TABLE IX

THE SUBARRAY  $PSNR_{\bar{z}_V, \bar{z}, \widetilde{QF}_{out}, \bar{z}}$  COMPUTED FOR *Case 1* (WITH  $z_V = 1$ ) USING THE IMAGE TRAINING SET FROM [6]. SHADING CORRESPONDS TO THE EXAMPLES IN SECTION VII.

$\widetilde{QF}_{out}$	Scaling, $\bar{z}$ , %									
	10	20	30	40	50	60	70	80	90	100
10	20.0	22.6	24.4	25.8	27.4	28.6	29.7	30.6	31.3	31.9
20	20.9	23.9	26.1	28.0	30.0	31.7	33.0	34.1	34.9	35.7
30	21.4	24.5	27.1	29.3	31.6	33.4	34.9	36.1	37.0	37.9
40	21.7	25.0	27.7	30.2	32.6	34.6	36.2	37.5	38.4	39.7
50	21.9	25.3	28.2	31.0	33.5	35.6	37.2	38.5	39.5	40.7
60	22.1	25.6	28.6	31.7	34.3	36.4	38.1	39.4	40.4	41.2
70	22.4	26.0	29.2	32.8	35.4	37.6	39.3	40.7	41.7	44.5
80	22.7	26.5	30.0	<b>34.3</b>	36.9	39.2	40.9	42.3	43.4	58.5
90	23.0	27.1	31.0	<b>37.1</b>	39.2	41.7	43.4	44.8	45.8	51.9
100	23.3	27.6	32.1	43.8	43.5	46.5	48.2	49.5	50.4	54.7

TABLE X

THE SUBARRAY  $PSNR_{\bar{z}_V, \bar{z}, \widetilde{QF}_{out}, \bar{z}}$  COMPUTED FOR *Case 2* (WITH  $z_V = 40\%$ ) USING THE TRAINING SET FROM [6]. SHADING CORRESPONDS TO THE EXAMPLES IN SECTION VII.

Finally, considering that a maximum relative file size of 0.50 with the other device constraints the same as previously, we find that the parameters that maximize the SSIM quality metric are  $QF_{out} = 90$  and  $z = 40\%$ , yielding a predicted SSIM score of 0.92. Now, looking at Tables VI and VIII, we find very same solution and predicted quality value. Using the PSNR as the quality metric, we again obtain the same solution for a PSNR of 37.1 dB.

## VIII. DISCUSSION

Using the PSNR as the quality metric, we observe that the system favors solutions with the highest allowed resolution, regardless of the QF (a trend we have observed over many examples and viewing conditions beyond those presented here). By contrast, using the SSIM as the quality metric, the system seems to arrive at a better tradeoff between the loss of detail due to scaling and the blocking artifacts introduced with a low QF. Indeed, it tends to select a solution with smaller resolution as the maximum permissible relative file size becomes smaller, thereby avoiding solutions with high resolution at the expense of an unacceptably small QF. We observed this behavior in the examples in the previous section, especially that in Figure 2. For instance, the optimal solution for a maximum relative file size of 0.25 was  $QF_{out} = 50$  and  $z = 60\%$  (not the poorer quality  $QF_{out} = 10$  and  $z = 100\%$ ), while for a maximum relative

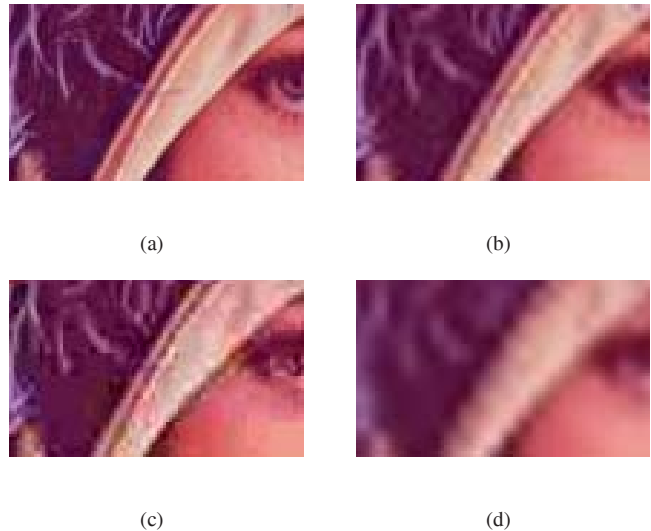


Fig. 2. Transcoded Lena (details) showing feasible solutions to the problem of a maximum relative file size of 0.25 with  $z_V=100\%$ : (a) original image; (b) optimal solution using the SSIM ( $QF = 50$  and  $z = 60\%$ ); (c) optimal solution using the PSNR ( $QF = 10$  and  $z = 100\%$ ); (d) a high-QF low-resolution solution ( $QF = 100$  and  $z = 20\%$ ).

file size of 0.50, it was  $QF_{out} = 40$  and  $z = 100\%$ , yielding a solution with maximal resolution.

In literature, researchers often claim the superiority of the SSIM over the PSNR for the objective assessment of JPEG-coded images [8]. This seems amply justified in terms of the scope of this research, as we believe it is more appropriate to use a smaller-resolution higher-QF solution than the highest possible resolution with extremely low QF (as would be favored by the PSNR).

Comparing Tables II and III, we observe that, overall, the SSIM values for  $z_V = 100$  (Table II) are smaller than those for  $z_V = 40$  (Table III). This observation also holds when the PSNR is used. This is due to the fact that, as  $z_V$  decreases, the original image is more aggressively scaled down for comparison (and therefore filtered), reducing the difference between the scaled down original image and the transcoded image. Therefore, as  $z_V$  decreases, quality values in the tables increase. However, it would be wrong to conclude that it is better to optimize for a smaller viewing condition scaling  $z_V$  because it yields higher quality values. The viewing condition scaling  $z_V$  should be fixed by the context of the transcoding problem, and from it an optimal solution identified.

For simplicity, we have assumed in our examples that  $z_V$  was set to the maximum resolution supported by the terminal, which may exceed its screen resolution—possibly being accessible to the user through a pan/zoom function. Alternatively, we may want to set  $z_V$  to the device's screen size. Interestingly, we observe that, for such case, the resolution of the optimal solution may be higher than  $z_V$ .

For instance, looking at Table III, we observe that, for a maximum relative size of either 0.25 or 0.50 (to use the same examples as in the previous section), we obtain better quality by transcoding with  $z = 50\%$  than with  $z = 40\%$ . This shows that it makes sense for an MMS mobile terminal to support higher resolution than its screen resolution for image storage in order to optimize quality, even if the images are only to be viewed on the device. Of course, the original reason why MMS mobile terminals support higher resolutions has more to do with the benefit of being able to zoom and pan with images and to display them on a more capable device such as a PC.

## IX. CONCLUSIONS

In this paper, we analyzed the impact of various combinations of QF and scaling parameter values on the quality of transcoded images. Using the SSIM, we showed how quality varies with  $QF_{out}$  and scaling for various viewing conditions. We also proposed a quality-aware transcoding system capable of predicting the best combination of QF and scaling for any given metric. Using the PSNR, solutions with maximal allowed resolution are typically selected without considering the QF. However, in general, using the SSIM metric, the systems tend to select parameters that balance output QF and scaling better than the PSNR. So, as the target file size becomes smaller the optimal resolution tends to be smaller to allow the selection of a reasonable quality factor and better overall quality. This confirms our initial hypothesis, which is that 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. We have also shown how this can be performed.

## 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] 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.
- [4] J. Ridge, "Efficient transform-domain size and resolution reduction of images," *Signal Processing: Image Communication*, vol. 18, no. 8, pp. 621–639, Sept. 2003.
- [5] V. Ratnakar and V. Ivashin, "File size bounded JPEG transcoder," May 2001, US Patent 6,233,359.
- [6] 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.
- [7] 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.
- [8] Hamid R. Sheikh, Muhammad F. Sabir, and Alan 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.
- [9] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.