No-reference prediction of quality metrics for H.264-compressed infrared sequences for unmanned aerial vehicle applications

Hossain, Kabir; Mantel, Claire; Forchhammer, Søren

Published in:
Journal of Electronic Imaging

Link to article, DOI:
10.1117/1.jei.28.4.043012

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
No-reference prediction of quality metrics for H.264-compressed infrared sequences for unmanned aerial vehicle applications

Kabir Hossain
Claire Mantel
Søren Forchhammer
No-reference prediction of quality metrics for H.264-compressed infrared sequences for unmanned aerial vehicle applications

Kabir Hossain, Claire Mantel, and Søren Forchhammer
Technical University of Denmark, Departments of Photonics Engineering, Kongens Lyngby, Denmark

Abstract. We propose a no-reference (NR) method for estimating the scores of metrics assessing the quality of infrared (IR) sequences compressed with H.264 for low-complexity unmanned aerial vehicle (UAV) applications. The scenario studied is to estimate the quality on an on-ground computer to avoid performing the processing onboard, due to the computational and memory limitations of the onboard hardware. For low complexity and fast feedback, a bitstream-based (BB) approach was chosen. The original IR sequences are captured by UAV, and then BB and pixel-based (PB) features are computed. Thereafter, a feature selection process is applied and the selected features are mapped using support vector regression, to predict the quality scores of full reference metrics. The method is evaluated for the NR prediction of four image and one video quality metrics. A set of five UAV- and three ground-IR sequences are used for evaluation. The proposed NR method consistently achieves robust results for the different objective metrics tested (Spearman rank-order correlation coefficients ranging from 0.91 to 0.99). A comparison with estimations based on features from three NR models from the literature proves to be in favor of the proposed method. © 2019 SPIE and IS&T [DOI: 10.1117/1.JEI.28.4.043012]

Keywords: infrared image quality assessment; H.264/AVC; dynamic range reduction operator; machine learning; no-reference; bitstream.

1 Introduction

The progress of low-form factor long-wave infrared (LWIR) sensors has made it possible to embed them on unmanned aerial vehicles (UAVs). Possible applications range from wild-life monitoring to inspection of energy installations for maintenance operations. In the case of district heating systems, a leak of warm water from the pipes can occur for many reasons, e.g., due to unprofessional installation, lack of maintenance, or end of service life. To avoid diminished performance due to the leakages, a technical inspection must be conducted to identify the leakage area in the distribution system. As the LWIR sensor captures heat variations, potential defects, such as leakages in the pipes of water or poor insulation, will appear as local hotspots. A fast and simple way to perform the inspection is to use a UAV, or drone, which allows inspection in a much cheaper way than with a helicopter and much faster than from the ground. After the image acquisition, the IR sequences obtained will be processed to detect the leakages, either by computer vision algorithms or by human inspection. Moreover, the IR images may also be used for navigating the drone. Therefore, the quality of the obtained IR sequences should be good enough for further processing and needs to be monitored and controlled. This estimation of quality for IR images is the focus of this paper.

Image quality assessment or video quality assessment (IQA/VQA) methods aim at evaluating the perceived characteristics of a visual input. Among those methods, the algorithmic ones, also called objective metrics, can be divided in three categories: full reference (FR), reduced reference (RR), and no reference (NR). For FR metrics, both the original signal and the distorted signals are used for quality estimation. In the case of RR methods, some features reflecting the original signal are used along the distorted version of the signal to evaluate the quality. In contrast to both FR and RR methods, NR methods do not require accessing the original signal but estimate the quality using only the distorted signal. This paper presents an NR framework that allows predicting FR quality metrics for IR images transmitted to the ground. The targeted application is the inspection of district heating installations with a UAV. The global setup and on-ground processing of this work are shown in Figs. 1 and 2, respectively.

In a low-complexity setup on the encoder side as above, the quality cannot be estimated on board due to hardware limitations. Whereas, based on the received coded sequences, the decoder side (on ground) has the resources to process and evaluate the quality (on-line) of the received image sequences or video stream in an NR manner. The on-ground computer can then send feedback to the UAV if the quality of IR sequences decreases significantly at any point. In this case, the feedback time should be low to allow for a rapid change in settings of the acquisition on the UAV.

The proposed method is illustrated in Fig. 3. As depicted, the proposed model consists of three main steps. Most IR sensors output 14 or 16 bits when acquiring radiometric temperature information; however, most image/video pipelines have a bit-depth of 8 bits. Therefore, the first step consists of reducing the dynamic range of the video sequence on board the UAV using simple operators and thereafter compressing (with H.264) and transmitting the sequence via ordinary pipelines, i.e., 8 bits per pixel, to the ground. In the second step, features characterizing the quality of the

*Address all correspondence to Kabir Hossain, E-mail: kabir@fotonik.dtu.dk
compressed sequences are generated using two different approaches. Two different approaches are considered. The first approach is fast and of relatively low complexity: it consists of computing features based on information obtained directly on the compressed H.264 bitstream format and is called bitstream-based (BB) approach. In the second approach, called pixel-based (PB), the features are based on the decoded sequence, leading to a longer computation time. Among the computed features, the most relevant ones are then selected. Finally, the selected features are mapped to the objective scores of FR quality metrics using support vector regression (SVR).

The main contribution of this work is the design of an NR framework to predict the quality grades of FR metrics for IR sequences. The method is adaptable to various systems and metrics, i.e., given training data, any image or video quality metric can be predicted on the on-ground computer. Furthermore, using only the bitstream as input without necessitating access to the original signal allows both saving computational power on the drone and adding flexibility to the proposed framework. Furthermore, three dynamic range reduction (DRR) operators are considered here [histogram equalization (HE), linear scaling (LS), and plateau histogram equalization (PE)], in order to explore how DRR affects the accuracy of quality prediction and the robustness of the framework across DRR operators. These three DRR operators were chosen for their really low complexity, to be applied on the UAV.

In the experimental section of this work, we apply the framework presented to predict the following FR quality metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), universal quality index (UQI), noise quality measure (NQM), and video quality metric (VQM). It should be noted that the sequences received by the ground computer are intended to be used for UAV navigation as well as the automatic detection of, e.g., leakages in underground pipes. Therefore, adaptability of the NR framework, which is presented here, to various metrics is very relevant. The quality metrics predicted in this paper are designed for visual content but are here applied to IR.
sequences. Even though there are obvious differences between IR and visible images, e.g., differences in texture and contrast, as detailed in Ref. 7, their content is also correlated in terms of object shapes and boundaries. One assumption of this work is indeed that those metrics can, to a certain extent, also evaluate the quality of IR sequences, captured by UAV. However, it should be noted that the framework can be extended to any quality metric, including prospective quality metrics designed specifically for IR content and training to predict performance of subjective tests given subjective data.

The remainder of this paper is organized as follows: the related prior work is introduced in Sec. 2. The proposed method is then detailed in Sec. 3. In Sec. 4, experimental results are presented: the performance of NR estimation of FR objective metrics is evaluated along with the influence of DRR operators on quality metrics prediction. Finally, conclusions are provided in Sec. 5.

2 Background and Prior Work

This paper presents an NR framework to predict the FR image quality estimations of IR sequences for low-complexity applications, such as UAV. Although VQA has been researched extensively9 to the best of our knowledge, very little research has been conducted to evaluate the quality of IR sequences.9

The difference between thermal IR and visible images originates in their two different bandwidths (7.5 to 14 μm for LWIR versus 380 to 750 nm for visible). As for visible images, models have been researched to study the statistical regularities present in IR images either by studying wavelet decomposition8 or natural scene statistics represented by mean-subtracted contrast normalized coefficients.10 Both cases show that although the parameters of the models are different, the models designed for visible images can be used to characterize IR images. The second study8 is particularly relevant in the considered scenario as the authors map those natural scene statistics (NSS) values of IR to subjective quality grades of images degraded by blur or additive white Gaussian noise using SVR [with an obtained accuracy of at most 0.825 in Spearman rank-order correlation coefficients (SROCCs)]. However, the considered IR images are either 8 bits or linearly scaled, when nowadays most IR sensors have a bit-depth of 14 to 16 bits.

As different DRR operators are evaluated to explore how the choice of a DRR operator influences the performance of the NR prediction of FR metrics, the literature study first focuses on DRR operators for IR images (Sec. 2.1). Then the second domain presented here is VQA, for which few studies focusing on IR were found. Owing to the lack of studies investigating IR images specifically, the prior work on VQA also includes NR quality prediction of visible sequences (Sec. 2.2).

2.1 Infrared Imaging

This subsection presents the state of the art for DRR and NR quality predictions of IR sequences. In this paper, three simple global DRR operators are used to reduce the original dynamic range of the sequences (14 to 16 bits) to the traditional 8 bits. In addition to the very simple LS, the two other operators considered are histogram-based: HE and PE.9 The main reason for choosing these three DRR operators is that their computational simplicity allows them to run on the low-complexity computing unit embedded on the UAV. There exist many DRR operators, e.g., Refs. 20-24 which outperform these three operators. These DRR operators have a common feature of operating locally on the images, which increases their computational complexity. One approach to reducing the dynamic range is to decompose the image into base and detail components and then reduce the dynamic range of the base component. The decomposition can, for instance, be done using local edge-preserving filters, as in Ref. 25 or a guided filter, as in Refs. 20 and 21. In addition, in Ref. 22, the authors defined a plausibility mask to avoid noise amplification in the DRR process. Another approach is to express the DRR problem as the minimization of a function in the gradient domain. Qiao and Ng presented a method to reduce the dynamic range of high-dynamic-range visible image by using localized gamma correction. The idea was adopted from a paper focused on IR image data, but applied to the visible domain.

The benefit of quality evaluation of IR images is valid for many applications whether it is medical imaging or target detection. Indeed, in Ref. 26, the authors use IR images to detect and localize closed vessels and state that “noise” affects the performance of the detection process. Therefore, having a quality control mechanism would be beneficial. Another example is the paper from Zhang et al. focusing on enhancing IR images for target detection. As mentioned by the authors, quality enhancement leads to enhanced performance of target detection. In this paper, the enhancement indices used to evaluate the performance are the target contrast with respect to the background, the average gradient, and the entropy of the image.

As stated before, few publications investigating the quality of IR images exist. In Ref. 27, the authors estimate the amount of additive nonuniformity (NU) in IR images based on their spatial structure. They focus on distortion-specific estimation of one type of noise and not global quality metric. In our preliminary work, the NR prediction of quality metrics of H.264-compressed IR sequences for UAV application was presented. This paper shall extend the preliminary research to more DRR operators, i.e., HE, LS, and PE to explore how DRR operators impact the performance of the method. Furthermore, the proposed method is here compared with features from other NR methods.

2.2 Visible Imaging

The framework presented here, i.e., performing NR BB quality predictions of IR sequences, is motivated by similar frameworks designed for visible sequences. One assumption is that the IR model requires a different parameterization. The idea is that the use of machine learning (ML) will allow adapting to IR imaging techniques designed for visible imaging. Therefore, in this subsection, several NR BB and PB quality assessment methods for visible sequences are presented for video compression defects.

Segard et al. presented an NR PB VQA method by analyzing individual frames, the intraframes (I-frame). They considered H.264- and MPEG-2-coded sequences and also presented an algorithm estimating the PSNR. Their video quality prediction applies ML and the results correlate well with subjective mean opinion scores (MOSs). Dakkar et al. proposed an NR IQA method using particle swarm...
optimization and SVR. At first, the score vector is constructed by using several state-of-the-art NR-IQA metrics, e.g., blind/reference-less image spatial quality evaluator (BRISQUE) and natural image quality evaluator (NIQE). Then, SVR-based fusion strategy is applied to predict MOS and difference MOS (DMOS).

In contrast to the PB approaches, Shahid et al. presented an NR BB video quality estimation using a least-square support vector machine. They extracted features from the H.264/AVC-coded bitstream of visible videos, such as the percentage of intra- and inter-coded macroblock, skipped macroblock, and so on. The model is also tested for its ability to predict subjective MOS, PEVQ, PSNR, SSIM, and MSSIM using low-resolution common intermediate format (CIF) and quarter CIF visible video sequences. Two NR and RR quality estimation techniques, using both PB and BB information, for H.264/AVC-encoded video sequences are presented by Shahid et al.

In that paper, the majority of the features are computed from bitstream and the rest of them are estimated from the pixels. Thereafter, the authors employ a least absolute shrinkage and selection operator regression model to select features and estimate MOS. Mocanu et al. presented an NR VQA method using different ML algorithms to predict subjective scores. Their list of features includes bitrate and percentage of intra (I), predicted (P), bidirectional (B) frames lost. Then statistical analysis is done on the performance of various ML methods (e.g., linear regression, decision trees for regression, random neural networks, and deep belief networks).

### 3 Proposed Method

The setup for the proposed framework is shown in Figs. 1 and 3. In this section, the details of the feature generation and mapping steps are presented. The features generated are of two types: the BB features, computed directly from the H.264-compressed video stream, and the PB features, which are computed on the decoded videos. The BB approach, being faster and computationally less expensive, is the main target of this paper.

The H.264 video compression standard allows for three types of coding for each frame: I, P, or B. The I-frames are compressed in a stand-alone manner, i.e., based only on the content of the considered frame. P and B frames are compressed using prediction (i.e., motion compensation) from content of other frames. Our target application requires low latency and low complexity. In addition, the frame rate of the IR sequences (9 Hz) is lower than the standard frame rate for visual sequences, which mean the motion vectors are larger between consecutive frames. We, therefore, chose to compress the sequences using I-frames only.

The compression of I-frames in H.264 is performed by applying a spatial integer block transform to image blocks of 4 x 4 or 8 x 8 pixels (similar in concept to the discrete cosine transform) along with a quantization, set by choosing a quantization parameter (QP), which is used to reduce the amount of information in a controlled manner. The QP determines the trade-off between the quality and the bitrate of the compressed sequence. This section is organized as follow: first, the feature generation (both for BB and PB approaches) is described in Sec. 3.1, then the feature selection process is presented in Sec. 3.2, before describing the ML algorithm used in Sec. 3.3.

#### 3.1 Feature Generation by Analysis of H.264 Intra-Frame

As presented in Sec. 3.1, two approaches are considered for the feature generation step. In the BB approach, the features are computed from the H.264-compressed stream, whereas for the PB approach the H.264 stream is first decompressed. Although both approaches are considered here, the BB is prioritized due to its faster speed and computationally less expensive characteristics.

In the BB approach, two sets of features are provided to the ML mapping: one set based on the QP extracted from the bitstream and the other based on an estimation of PSNR computed from transformed coefficients and the QP. Although this estimation of PSNR is not our final goal, it is used to build features provided to the ML mapping. Many papers estimated PSNR without a reference signal (i.e., in an NR manner). Here, the procedure described in Ref. 28 is used to obtain an initial estimate that is used to define features for predicting the objective scores for all metrics and not just PSNR. More details are given in Sec. 3.1 about this initial PSNR estimation to derive features from I-frames of H.264 video sequences in a BB and NR manner.

We calculate a total of 18 BB features (Table 1), based on QP and estimated PSNR. As both QP and estimated PSNR are computed on a frame-by-frame level, they need to be temporally pooled into a value for each sequence. The chosen pooling schemes are mean, standard deviation, maximum differences, maximum differences in consecutive I-frames, and the gradient of a line fitted to QP or estimated PSNR. Further pooling methods consist in dividing the QP and estimated PSNR values into two clusters, using k-means clustering, the high and low clusters. The motivation for this segmentation is that the high and the low clusters are not equally important for video quality.

Therefore, four features are also calculated using the mean of high and low clusters and the weighted mean in each of the clusters.

Moreover, as mentioned above, we optionally use some additional PB features, which are measured from decoded IR sequences (see Table 1). The PB features consist of IQA (adapted from the BRISQUE metric as described in Ref. 28) and perceptual spatial information (SI) and perceptual temporal information (TI) features.

#### 3.2 Feature Selection

The feature selection procedure is used to increase the accuracy of the prediction of metrics. For high-dimensional data, feature selection is very useful since SVR is nonsparse. The feature selection is applied before training with SVR. In short, the main idea of feature selection is to rank the features (including some additional noise features) and then eliminate the lower-ranking features, which are not ranked above the average of the noise features. A description of the feature selection process can be found in Ref. 28.

We have evaluated our method for three cases corresponding to three types of input features. The first consists of only the BB features, the second consists of the BB features supplemented with SI/TI features, and the third consists of the BB features supplemented with IQA features. It should be noted that, the frame rate of the IR sequence is 9 Hz, which means the motion between consecutive images is larger than for a standard video frame rate (25 Hz). Furthermore, as the images are compressed using I-frames only, the role of
Table 1 Number of features calculated for ML.

<table>
<thead>
<tr>
<th>Features by pooling</th>
<th>No. of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB features PSNR</td>
<td>Estimated</td>
</tr>
<tr>
<td></td>
<td>(a) Standard deviation (SD)</td>
</tr>
<tr>
<td></td>
<td>(b) Mean</td>
</tr>
<tr>
<td></td>
<td>(c) Maximum differences of est. PSNR</td>
</tr>
<tr>
<td></td>
<td>(d) Maximum differences of consecutive I-frames</td>
</tr>
<tr>
<td></td>
<td>(e) The gradient of a line fitted to est. PSNR</td>
</tr>
<tr>
<td></td>
<td>(f) Mean of SI</td>
</tr>
<tr>
<td></td>
<td>(g) Mean of high cluster</td>
</tr>
<tr>
<td></td>
<td>(h) Weighted mean of high cluster</td>
</tr>
<tr>
<td></td>
<td>(i) Weighted mean of low cluster</td>
</tr>
<tr>
<td>QP</td>
<td>Same as above</td>
</tr>
<tr>
<td>IQA</td>
<td>(a) SD and (b) mean</td>
</tr>
<tr>
<td>PB features SI/TI</td>
<td>(a) Maximum element of SI</td>
</tr>
<tr>
<td></td>
<td>(b) Minimum element of SI</td>
</tr>
<tr>
<td></td>
<td>(c) Mean of SI</td>
</tr>
<tr>
<td></td>
<td>(d) SD of SI</td>
</tr>
<tr>
<td></td>
<td>(e) Maximum element of TI</td>
</tr>
<tr>
<td></td>
<td>(f) Mean of TI</td>
</tr>
<tr>
<td></td>
<td>(g) SD of TI</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
</tr>
</tbody>
</table>

TI features to evaluate the quality of the compression is negligible. Overall, the feature selection procedure significantly improved our results, as shown in Table 1 and Sec. 4.

3.3 Support Vector Regression

In this paper, the ML algorithm that is used to map the features to objective quality measures is SVR. A brief description of SVR can be found in Ref. 2.

The SVR equation can be written as

\[ y_n - f(x) \leq \epsilon, \]  

where \( y_n \) is the output, \( f(x) \) is the function of feature vector, \( n \) is the number of output, and \( x \) is the corresponding input feature vector. The parameter \( \epsilon \) in the SVR formulation affects the smoothness of the SVR response. The function of feature vector \( x \) can be calculated by the following equation:

\[ f(x) = \sum_{i=1}^{N} (\alpha_i^t - \alpha_i) K(x_i, x) + \beta, \]  

subject to \( (\alpha_i^t - \alpha_i) \in [0, C] \), where \( N \) is the number of feature vectors, \((\alpha_i^t - \alpha_i)\) are so called support vectors and \( \beta \) is an offset, \( K(x_i, x) \) is the kernel function, and \( x_i \) denotes feature vectors. Here, \( C \) is the cost that controls the trade-off range between the training error and the flatness of the solution. In our work, we adapt the implementation from Ref. 28, wherein more details are described. The radial basis function, which is used as a kernel function in this work, is given as

\[ K(x_i, x) = e^{-\gamma ||x_i - x||^2}, \]

where \( \gamma \), in the radial basis function, controls the standard deviation of the Gaussian function. The three parameters, i.e., \( \gamma \), \( \epsilon \), and \( C \), in the SVR formulation are crucial because inaccurately choosing these parameters can lead to underfitting or overfitting the model. Therefore, when training the model, for each metric, the optimal values are searched (see Sec. 4). For more details on SVR, readers are referred to Refs. 1 and 10.

4 Performance Evaluation

In this section, the performance of the proposed method is evaluated. First, the setup is described in Sec. 4.1, and then the results are analyzed as to the method itself in Sec. 4.2, compared with other methods in Sec. 4.3, and evaluated across the two DRR operators in Sec. 4.4.

4.1 Experiment Description

The process of the performance evaluation simulates the setup illustrated in Figs. 2 and 3. For the experiment, a total of eight IR sequences were used: five acquired by a UAV and three acquired on ground. The five UAV sequences were obtained by night flights inspecting the areas with district heating pipelines, searching for potential leakages. These sequences were provided by Drone Systems (Denmark). The UAV sequences, originally available in 16 bits, are first converted from 16 to 8 bits. Then the compression, feature generation, and quality metric estimation are carried out on the 8-bit sequences.

The UAV sequences are converted into 8 bits (as shown in Fig. 3) using three low-complexity DRR operators: LS, HE, and PE (with a plateau of 1000) as part of...
the experiment. It could be noted here that all three were
tested but only results for two DRR operators are reported
as LS is very simple and it gives inferior results. A total of
three ground sequences are also used: one is from the INO
database and the other two were acquired by the authors.

The eight sequences (five by UAV and three on ground),
converted to 8-bit format, were compressed by the
H.264 encoder software (version: 2389), which is a H.264/AVC-
compliant video encoder. The compression was performed
using I-frames only for low-delay and low-complexity cod-
ing. Each sequence was encoded at seven different bitrates,
thus creating a total of 35 coded sequences from the five
UAV sequences and 21 from the three ground sequences.

As described in Table 2, the sequences were encoded with
different bitrates, depending on the DRR operator and reso-
lutions considered. The bitrates were chosen so that the
compressed sequences are similar in terms of PSNR.

The proposed method is applied to predict the scores
of four image and one video FR metrics: PSNR, SSIM, UQI,
NQM, and VQM. The chosen image metrics are
computed frame by frame and then averaged into one grade
per sequence.

For the five UAV sequences, 80% of the sequences (four
contents out of five) are used for training and the remaining
20% are used for testing (the fold being split by different
content, i.e., original sequence). For performance evaluation,
we have performed experiments including all two DRR
operators as well as for each DRR operator, separately. The
experiments are also carried out for the combined set of
all eight sequences (five UAVs and three on ground). For eight sequences, again a leave-one-out approach was chosen
for testing and the remaining seven sequences were used for
training.

4.2 Performance of the Proposed Method

The results are reported in Table 3 by SROCCs, linear
correlation coefficients (LCC), and root mean square error
(RMSE). As mentioned, DRR operators may influence the
performance of prediction; therefore, three different DRR

<table>
<thead>
<tr>
<th>Table 2</th>
<th>IR sequences and x264 encoder parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Setting</td>
</tr>
<tr>
<td>Resolution (five UAV IR)</td>
<td>640 × 512</td>
</tr>
<tr>
<td>Resolution (three ground)</td>
<td>320 × 256 (two sequences were captured by Boson camera)</td>
</tr>
<tr>
<td>Frames</td>
<td>300</td>
</tr>
<tr>
<td>FPS</td>
<td>9 and 10 (INO database)</td>
</tr>
<tr>
<td>For HE, bitrates (kbps) (five UAV)</td>
<td>200, 500, 1000, 2000, 3000, 4000, 5000</td>
</tr>
<tr>
<td>For PE, bitrates (kbps) (five UAV)</td>
<td>50, 100, 200, 500, 1000, 2000, 3000</td>
</tr>
<tr>
<td>Bitrates (kbps) (three ground)</td>
<td>50, 100, 200, 300, 400, 700, 1000</td>
</tr>
</tbody>
</table>

| Table 3 | Cross-validation results of SROCCs, LCC, and RMSE predicting FR metrics. |
|---------|-----------------|-----------------|-----------------|
|         | PSNR            | NQM             | PSNR            | NQM             |
|         | SROCC | LCC | RMSE | SROCC | LCC | RMSE |
| BB      | 0.97  | 0.98 | 1.00 | 0.98  | 0.99 | 1.59 |
| BB + IQA | 0.99  | 0.99 | 0.70 | 0.98  | 0.99 | 1.70 |
| BB + SI/TI | 0.98  | 0.97 | 1.17 | 0.99  | 0.99 | 0.98 |
| SSIM    |       |     |     |       |     |     |
| UQI     |       |     |     |       |     |     |
| BB      | 0.98  | 0.98 | 0.01 | 0.95  | 0.98 | 0.04 |
| BB + IQA | 0.97  | 0.98 | 0.02 | 0.92  | 0.97 | 0.06 |
| BB + SI/TI | 0.98  | 0.98 | 0.02 | 0.98  | 0.99 | 0.02 |
| VQM     |       |     |     |       |     |     |
| BB      | 0.96  | 0.98 | 0.04 |       |     |     |
| BB + IQA | 0.97  | 0.99 | 0.09 |       |     |     |
| BB + SI/TI | 0.98  | 0.99 | 0.07 |       |     |     |
| PSNR    |       |     |     |       |     |     |
| NQM     |       |     |     |       |     |     |
| BB      | 0.99  | 0.99 | 0.38 | 0.99  | 0.99 | 0.30 |
| BB + IQA | 0.99  | 0.99 | 0.74 | 0.99  | 0.99 | 0.62 |
| BB + SI/TI | 0.99  | 0.99 | 0.38 | 0.99  | 0.99 | 0.58 |
| SSIM    |       |     |     |       |     |     |
| UQI     |       |     |     |       |     |     |
| BB      | 0.99  | 0.97 | 0.07 | 0.97  | 0.99 | 0.02 |
| BB + IQA | 0.98  | 0.99 | 0.72 | 0.98  | 0.99 | 0.02 |
| BB + SI/TI | 0.99  | 0.99 | 0.01 | 0.99  | 0.99 | 0.01 |
| VQM     |       |     |     |       |     |     |
| BB      | 0.97  | 0.98 | 0.04 |       |     |     |
| BB + IQA | 0.96  | 0.99 | 0.03 |       |     |     |
| BB + SI/TI | 0.97  | 0.99 | 0.03 |       |     |     |
operators have been applied to convert the 16-bit images to 8-bit representation. The best results for predicting each metric are highlighted in bold. As described in Sec. 3, the cost, $C$ [Eq. (4)], $\gamma$ [Eq. (3)], and $\epsilon$ [Eq. (1)] parameters are found by grid search based on SVR results (evaluated by SROCCs) for NR quality metrics prediction. As an example, Fig. 4 shows the SROCC values as a function of gamma, cost, and epsilon parameters. The figures depict the SROCC performance for the five UAV sequences (SSIM metric). The optimal parameters are $10^{-5.7}$. $10^{2}$, and $2^{-53}$ for gamma, cost, and epsilon, respectively. Figure 4(d) shows the gamma versus cost plot, when epsilon is fixed, and the Fig. 4(b) represents the gamma versus epsilon, when cost is fixed.

The regression plots for SSIM prediction using BB features only are shown in Fig. 3. The corresponding regression plots for the four other metrics are shown in Sec. 7 (Figs. 8–11). The proposed NR solution is capable of predicting the quality metrics for H.264-compressed IR sequences with a high accuracy.

The first table, Table 3(a) presents results for images grouping the two DRR operators together. Tables 3(b) and 3(c) present separate evaluation for each of the two chosen DRR operators, i.e., HE and PE. The experiments are finally carried out for the three ground sequences combined with the five UAV sequences, to test that the proposed method works well across different scenarios [Table 3(d)].

As shown in Table 3(b), in the case of BB feature only, the SROCC values are 0.97, 0.98, 0.98, 0.95, and 0.96 for PSNR, NQM, SSIM, UQI, and VQM, respectively. Results obtained with HE gave slightly improved results in terms of SROCCs compared with PE, as shown in Tables 3(b) and 3(c). In Table 3(d), the obtained SROCC values are 0.99, 0.99, 0.99, 0.97, and 0.97 for PSNR, NQM, SSIM, UQI, and VQM, respectively. As mentioned above, the experiment was also carried out for LS, but it is not reported in the table since LS is very simple and the visual quality of image is not satisfying (Fig. 4L). The LS sequences are encoded with four different bitrates (25, 50, 100, and 200) and the corresponding SROCC values obtained are 0.98, 0.95, 0.96, 0.90, and 0.92 for PSNR, NQM, SSIM, UQI, and VQM, respectively.

In Table 3(d), the set of eight sequences (three ground and five UAV using HE) are used for this experiment. The SROCC values obtained are 0.97, 0.93, 0.95, 0.94, and 0.98 for PSNR, NQM, SSIM, UQI, and VQM, respectively. Overall as shown in Tables 3(a)–3(d), the proposed method predicts PSNR scores with high accuracy (SROCC values from 0.97 to 0.99), whereas it predicts NQM scores with comparatively lower accuracy (SROCC values from 0.93 to 0.99) for codec only (BB). The 0.93 in SROCCs for BB prediction of NQM [Tables 3(b) and 3(d)] is really a not-so-high SROCC value. It should be pointed out that in this case a clear improvement is achieved by adding PB features. For UQI, the obtained SROCC values range from 0.94 to 0.97. For SSIM, the BB version of our NR method gives SROCC values from 0.95 to 0.99. The proposed method is also applied to predict the grades from VQM and it achieves SROCC values from 0.96 to 0.98. When we consider BB features combined with PB features, the reported best SROCC values are in the range from 0.96 to 0.99 for the five metrics (PSNR, NQM, SSIM, UQI, and

### Table 3 (Continued)

<table>
<thead>
<tr>
<th>(c) Five IR sequences for PE</th>
<th>PSNR</th>
<th>NQM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SROCC</td>
<td>LCC</td>
</tr>
<tr>
<td>Five UAV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>BB + IQA</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>BB + SI/TI</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

| (d) Eight IR sequences (five UAV sequences using HE and three ground sequences) |
|----------------------------|------|-----|
|                            | PSNR | NQM |
|                            | SROCC | LCC | RMSE | SROCC | LCC | RMSE |
| BB                        | 0.97  | 0.96 | 1.38 | 0.93  | 0.94 | 2.59 |
| BB + IQA                   | 0.97  | 0.97 | 1.42 | 0.93  | 0.91 | 3.19 |
| BB + SI/TI                | 0.99  | 0.99 | 0.74 | 0.97  | 0.97 | 1.91 |

<table>
<thead>
<tr>
<th>SSIM</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SROCC</td>
<td>LCC</td>
<td>RMSE</td>
<td>SROCC</td>
<td>LCC</td>
<td>RMSE</td>
</tr>
<tr>
<td>BB</td>
<td>0.95</td>
<td>0.94</td>
<td>0.04</td>
<td>0.94</td>
<td>0.93</td>
<td>0.07</td>
</tr>
<tr>
<td>BB + IQA</td>
<td>0.96</td>
<td>0.92</td>
<td>0.04</td>
<td>0.97</td>
<td>0.94</td>
<td>0.05</td>
</tr>
<tr>
<td>BB + SI/TI</td>
<td>0.97</td>
<td>0.94</td>
<td>0.03</td>
<td>0.91</td>
<td>0.91</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VQM</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SROCC</td>
<td>LCC</td>
<td>RMSE</td>
<td>SROCC</td>
<td>LCC</td>
<td>RMSE</td>
</tr>
<tr>
<td>BB</td>
<td>0.98</td>
<td>0.99</td>
<td>0.04</td>
<td>0.97</td>
<td>0.96</td>
<td>0.04</td>
</tr>
<tr>
<td>BB + IQA</td>
<td>0.96</td>
<td>0.99</td>
<td>0.04</td>
<td>0.98</td>
<td>0.99</td>
<td>0.02</td>
</tr>
<tr>
<td>BB + SI/TI</td>
<td>0.97</td>
<td>0.99</td>
<td>0.02</td>
<td>0.98</td>
<td>0.99</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Both LCC and RMSE values are coherent with the computed SROCCs. Therefore, the proposed method obtains satisfying results for the prediction of all tested metrics.

Table 4 presents the features selected per category for each quality metric. The number in the table represents the number of features selected by the feature selection procedure and the letters besides that represent the corresponding features (described in Table 1). For BB, in most of the cases, the feature selection chose mean of high and low cluster, weighted mean of high cluster, and so on, from QP and estimated PSNR, respectively. Overall, in most cases, the feature selection procedure chose 5 to 13 bitstream (BB) features (out of 18), which thereafter are provided to the ML algorithm (i.e., SVR). Furthermore, the experiments are carried out for PB features, i.e., SI/TI and IQA features.
and IQA are selected. The process selects 1 to 3 SI features (out of 4), 1 to 2 TI features (out of 3), and 1 or 2 IQA features (out of 2) provided to the SVR.

### 4.3 Comparative Performance Evaluation

The method proposed here was compared with estimation based on features from three NR VQA models: two video methods, i.e., Søgaard et al. and Video-BLIINDS, and one image method, Curvelet-QA, for the five UAV IR sequences. In all three cases, the MATLAB code provided by the authors was used to compute the corresponding features. For practical reasons (code unavailable for CurveletQA and in an incompatible language for Video-BLIINDS), the mapping from features to objective scores was performed using our MATLAB SVR analysis. For the CurveletQA, the image features are turned into video features by averaging over time. The training and testing schemes for these NR methods have been conducted with the same cross-validation approach as used in this paper.

The results of the comparative evaluation are shown in Table 5 using SROCCs. Overall, the proposed NR BB and PB (BB + SI/TI) prediction methods perform better than using the features of the three methods from the literature. For example, the proposed method (BB + SI/TI) outperforms the application of Søgaard’s features in terms of SROCCs by up to 0.14 for HE and 0.29 for PE.

### 4.4 Performance for Different Dynamic Range Reduction Methods

The results have been calculated for three DRR operators. The results using DRR operators are presented in Tables 3b and 3c and mentioned in Sec. 4.2. The main idea here is to explore how the DRR operator affects the prediction of the metrics and the robustness across the choice of DRR operator (which highly influences the image rendering as shown in Fig. 6). As mentioned above, three DRR operators, HE, LS, and PE, are used to convert the 16-bit images into an 8-bit format. Figure 6 shows that the local contrast is the highest with the HE DRR operator and much lower for the LS.

We have calculated the average SROCC and LCC values over all metrics, as shown in Fig. 7. The best average SROCC values are 0.986 and 0.986, and the best average LCC values are 0.990 and 0.990 for HE and PE, respectively. These results show that our method is both well performing and robust across the two DRR operators used for our UAV sequences.

We have also calculated NR estimation using BB PSNR for three DRR operators (here the result of LS is not reported since LS is very simple), which are used as an intermediate result for feature generation for all VQAs (see Sec. 6). For performance evaluation of BB PSNR, the SROCCs between the initial estimated PSNR and true PSNR are calculated. The performance in terms of SROCCs are 0.99 and 0.98 and

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Cases</th>
<th>QP (9)</th>
<th>Est. PSNR (9)</th>
<th>SI (4)</th>
<th>TI (3)</th>
<th>IQA (2)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>BB</td>
<td>6 (a, b, e–h)</td>
<td>3 (d, e, h)</td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>BB + IQA</td>
<td>5 (a, b, f–h)</td>
<td>4 (d–g)</td>
<td>1(a)</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>BB + SI/TI</td>
<td>4 (a, b, f)</td>
<td>3 (e–g)</td>
<td>1(a)</td>
<td>1(g)</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>NQM</td>
<td>BB</td>
<td>3 (a, b, e)</td>
<td>4 (b, d, e, g)</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>BB + IQA</td>
<td>1 (b)</td>
<td>4 (b–d, g)</td>
<td>2 (a, b)</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>BB + SI/TI</td>
<td>3 (a, b, e)</td>
<td>3 (a, b, d)</td>
<td>3 (b–d)</td>
<td>2 (f, g)</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>SSIM</td>
<td>BB</td>
<td>8 (a–h)</td>
<td>5 (a, d–g)</td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>BB + IQA</td>
<td>4 (a, f–h)</td>
<td>4 (d–g)</td>
<td>1(a)</td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>BB + SI/TI</td>
<td>6 (a, c, d, f–h)</td>
<td>3 (e, g, h)</td>
<td>1(a)</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>UQI</td>
<td>BB</td>
<td>4 (a, f–h)</td>
<td>1 (a)</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>BB + IQA</td>
<td>5 (a, d, f–h)</td>
<td>2 (e, h)</td>
<td>2 (a, b)</td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>BB + SI/TI</td>
<td>5 (a, d, f–h)</td>
<td>1 (a)</td>
<td>3 (a, c, d)</td>
<td>1(e)</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>VQM</td>
<td>BB</td>
<td>5 (a, b, f–h)</td>
<td>4 (d–g)</td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>BB + IQA</td>
<td>6 (a, c, d, f–h)</td>
<td>4 (e–h)</td>
<td>2 (a, b)</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>BB + SI/TI</td>
<td>7 (a–d, f–h)</td>
<td>6 (a, d–h)</td>
<td>2 (a, b)</td>
<td>1(g)</td>
<td></td>
<td>16</td>
</tr>
</tbody>
</table>

---

**Table 4** Features selected per category for PE and HE DRR operators (the number beside the name of each metric indicates the number of potential features before selection).
Table 5  Performance (SROCC) comparison with other NR features (five UAV sequences). The best SROCC scores appear in bold.

<table>
<thead>
<tr>
<th></th>
<th>Video-BLIINDS41</th>
<th>Søgaard et al.</th>
<th>Curvele-QA</th>
<th>Proposed method (only BB)</th>
<th>Proposed method (BB+SI/TI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HE Seq.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>0.77</td>
<td>0.96</td>
<td>0.93</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>NQM</td>
<td>0.90</td>
<td>0.88</td>
<td>0.90</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.62</td>
<td>0.89</td>
<td>0.61</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>UQI</td>
<td>0.63</td>
<td>0.85</td>
<td>0.62</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>VQM</td>
<td>0.60</td>
<td>0.86</td>
<td>0.63</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>PE Seq.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>0.91</td>
<td>0.97</td>
<td>0.93</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>NQM</td>
<td>0.92</td>
<td>0.90</td>
<td>0.90</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.57</td>
<td>0.80</td>
<td>0.61</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>UQI</td>
<td>0.60</td>
<td>0.80</td>
<td>0.59</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>VQM</td>
<td>0.75</td>
<td>0.70</td>
<td>0.66</td>
<td>0.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Fig. 6  Three DR-reduced images (PE, HE, and LS) from flight 1.

Fig. 7  Average SROCC and LCC values for five UAV sequences and eight IR sequences (five UAV using HE and three ground sequences). (a) Average SROCCs for two DRR operators, combined and individually and eight IR sequences. (b) Average LCC for two DRR operators, combined and individually and eight IR sequences.
with ML (for codec only) they are 0.99 and 0.99 for HE and PE, respectively.

5 Conclusion

The framework of an NR prediction of FR quality metrics for IR sequences compressed with H.264 was presented for low-complexity and low-delay UAV applications. After mapping from 16- to 8-bit with a DRR operator, the IR sequences are compressed with a H.264 encoder using I-frames only. Features, estimated directly from the bitstream in a BB approach or in combination with PB features, then undergo a selection step before being mapped using SVR to predict quality metrics in an NR manner.

Four image and one video quality metrics, i.e., PSNR, NQM, SSIM, UQI, and VQM, are used for the test cases and performance evaluation. The proposed method achieves average SROCC values (for BB) of 0.982 and 0.980 for the two DRR operators: HE and PE, respectively. Adding PB features to the BB approach gives a slight improvement with average SROCC values of 0.986 for both HE and PE, respectively. The results show that the proposed method performs well when compared with features from three NR VQA models from the literature. Moreover, testing on two low-complexity DRR operators, combined and individually and eight IR sequences (five UAV using HE and three ground sequences).

The proposed framework can be extended to any quality metric, including prospective quality metrics designed for IR content. Future investigation would be to extend the scope of this work by evaluating the presented method for other use cases and to predict quality for IR sequences compressed with a H.265/HEVC video codec.

6 Appendix A: Initial Peak Signal-to-Noise Ratio Estimation for Feature Generation

PSNR is commonly used as a simple measure of the quality of reconstructed images and video data. In this section, an initial BB PSNR estimation procedure is described. Later, this initial PSNR is used to build features to predict quality metrics.

There are many papers that estimate PSNR without reference signal for H.264-compressed videos. We adapt the procedure described in Ref. [25] and assume that the distribution of the original transform coefficients is known. The local mean square error (MSE) of the k′th coefficient will be denoted by \( \xi^2_{k,j} \) for each macroblock (MB) in the coded stream representing 16 × 16 pixels. The input image is divided into 16 × 16 blocks, called MB. The overall DCT sub-band for the whole frame can be measured by simply averaging over all quantized values in a given transform point, i.e., having the same two-dimensional frequency. The DCT coefficients are divided into different frequency bands, known as DCT sub-bands. The weighted summation of each of the quantization intervals, being indexed by \( j \) (which is defined by the QP value), can be calculated as

\[
\xi^2_k \approx \sum_j P_k(X_j) \frac{\int_{a_j}^{b_j} f_x(x)(X_j - x)^2 dx}{\int_{a_j}^{b_j} f_x(x) dx}, \tag{4}
\]

where \( f_x(x) \) is the original distribution of the coefficient at band \( k \) and \( P_k(X_j) \) denote the percentage of the transform coefficients in band \( k \) inside the \( j \)th quantization interval, which is bounded by \( a_j \) and \( b_j \). A single Cauchy distribution has been used for each coefficient \( k \) to model the original distributions of the transform coefficients. The percentage of \( P_0 \) coefficients smaller than \( \alpha \times \xi \) in the reconstructed coefficients, where \( \xi \) is the quantization step and \( \alpha \) is a constant value, is used to balance the peak value for different \( \xi \) empirically. It is used to estimate the shape parameter \( \gamma_k \) of each Cauchy distribution at band \( k \), which is calculated using \( P_0 \) as

\[
\gamma_k = \frac{\alpha \times \xi}{\tan \frac{\alpha \times \xi}{2}}. \tag{5}
\]

Using Eqs. (4) and (5), the sub-band MSE is calculated by summing the contributions, for quantization intervals \( \hat{X_j} \neq 0 \) and \( \hat{X_j} = 0 \), respectively. If \( \hat{X_j} = 0 \), as shown in Ref. [25], the contribution is calculated as

\[
\xi^2_{k,j} = P_k(\hat{X}_j) \left( \alpha \times \xi \times \gamma_k - \gamma_k^2 \arctan \frac{\alpha \times \xi}{\gamma_k} \right), \tag{6}
\]

where \( P_k(\hat{X}_j) \) is the ratio of reconstructed coefficients in position \( k \) inside the \( j \)th quantization interval belonging to \( \hat{X}_j \). If \( \hat{X}_j \neq 0 \), then

\[
\xi^2_{k,j} = P_k(\hat{X}_j) \times \xi \times \gamma_k \left( + P_k(\hat{X}_j)(\hat{X}_j^2 + \gamma_k^2) \arctan \frac{\hat{X}_j}{\gamma_k} \right. \\
- P_k(\hat{X}_j)(\hat{X}_j^2 + \gamma_k^2) \arctan \frac{\hat{X}_j - (1 - \alpha) \times \xi}{\gamma_k} \right. \\
+ P_k(\hat{X}_j) \gamma_k \hat{X}_j \log \frac{|\hat{X}_j - (1 - \alpha) \times \xi|^2 + \gamma_k^2}{(\hat{X}_j + \alpha \times \xi)^2 + \gamma_k^2}. \tag{7}
\]

Finally, the MSE is calculated by summing over all the quantization intervals and coefficient positions, yielding \( \text{MSE}_{\text{est}} \) that is used to calculate the estimated PSNR

\[
\text{MSE}_{\text{est}} = \frac{1}{16} \sum_{k=1}^{16} \sum_{j} \xi^2_{k,j}, \tag{8}
\]

\[
\text{PSNR}_{\text{est}} = 10 \log_{10} \frac{255^2}{\text{MSE}_{\text{est}}}. \tag{9}
\]

7 Appendix B: Results for No-Reference Prediction of Quality Metrics

In this section, the additional regression plots for NQM, PSNR, SSIM, UQI, and VQM prediction using BB features only are presented (Figs. 8-11). The plots are shown here for two DRR operators, combined and individually and eight IR sequences (five UAV using HE and three ground sequences).
Fig. 8 Plot of the NR quality estimation for two DRR operators (HE and PE): (a) NQM, (b) PSNR, (c) UQI, and (d) VQM.

Fig. 9 Plot of the NR quality estimation for PE: (a) NQM, (b) PSNR, (c) UQI, and (d) VQM.
Fig. 10 Plot of the NR quality estimation for HE: (a) NQM, (b) PSNR, (c) UQI, and (d) VQM.

Fig. 11 Plot of the NR quality estimation for eight IR sequences (five UAV using HE and three ground sequences): (a) NQM, (b) PSNR, (c) UQI, and (d) VQM.
Acknowledgments
This research was partly funded by the EUDP project “Energy efficient drone inspection system for spotting leakages,” Grant No. 64015-0072.

References

Kabir Hossain received his BE degree in computer engineering and his ME degree in electronics and computer engineering from Chonnam National University, South Korea, in 2011 and 2013, respectively. Currently, he is working toward his PhD with the Coding and Visual Communication Group, Department of Photonics Engineering of the Technical University of Denmark, Lyngby. His research interests include image and video coding, image and video quality assessments, computer vision, and image processing with machine learning and neural network.

Claire Mantel received her MSc degree and PhD in signal processing from Grenoble Polytechnic Institute, France, in 2007 and 2011, respectively. Currently, she is working as a senior researcher at the Department of Photonics Engineering of the Technical University of Denmark. Her research interests include image and video coding, visual quality assessment, and different aspects of the image processing chain for unmanned aerial vehicle-based imaging.

Soren Forchhammer received his MS degree in engineering and his PhD from the Technical University of Denmark, in 1984 and 1988, respectively. Currently, he is a professor at DTU Fotonik, Technical University of Denmark, where he has been since 1988. He is head of the Coding and Visual Communication Group at DTU Fotonik. His interests include source coding, image and video coding, distributed source coding, drone video, processing for image displays, coding for optical communications, information theory, and visual communications.