https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-108

No Reference Prediction of Quality Metrics for H.264 Compressed Infrared Image Sequences for UAV Applications

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

Abstract

The framework for this research work is the acquisition of Infrared (IR) images from Unmanned Aerial Vehicles (UAV). In this paper we consider the No-Reference (NR) prediction of Full Reference Quality Metrics for Infrared (IR) video sequences which are compressed and thus distorted by an H.264 codec. The proposed method works as a Bitstream Based (BB) approach and it may thus be applied on-ground. Three different types of features are first computed: codec features (based on information extracted from the bitstream), image quality features (based on BRISQUE evaluations) and Spatial and temporal perceptual information.

Those features are then mapped, using a machine learning (ML) algorithm, the Support Vector Regression (SVR), to the quality scores of Full Reference (FR) quality metrics. The novelty of this work is to design a NR framework for the prediction of quality metrics by applying ML algorithm in the IR domain.

A set of 5 drone energy leakage image sequences and 3 ground IR image sequences are used for evaluating the performance of the proposed method. Each of the image sequences are encoded at 4 different bitrates and the prediction of the proposed method is compared with the true FR quality metrics scores of four images metrics: PSNR, NQM, SSIM and UQI and one video metric: VQM. Results show that our technique achieves a fairly reasonable performance. The improved performance obtained in SROCC and LCC is up to 0.99 and the RMSE is reduced to as little as 0.01 between the actual FR and the estimated quality scores for the H.264 coded IR sequences.

Introduction

Infrared images have a wide range of applications such as inspection of buildings or energy systems, wildlife monitoring etc. An IR camera is able to detect the locations of defects in manufactured materials and mounting it on an Unmanned Aerial Vehicle (UAV) allows performing inspections much cheaper than with a helicopter and much faster than from the ground. In order to be streamed to a base station, the acquired IR images needs to be compressed. For all of the above-mentioned applications, it is required to process IR sequences (either by visual inspection or computer vision algorithm), which implies that the image quality should be good enough for further processing. Thus quality control and estimation are very important for IR image acquisition.

Visual quality assessment methods can be divided in to three categories: Full Reference (FR), Reduced-Reference (RR) and No Reference (NR). In a FR scenario, an original undistorted video sequences is compared to a degraded version of the same video. In the RR case, partial information of the original video is available for quality estimation. The NR method does not require accessing the original signal. In our case, we use NR VQA for estimating a quality score.

Even though NR video quality estimation algorithms are an active research topic for visible sequences [1,2], not much work has been done so far on IR with similar models, especially for UAV applications. The contribution in this paper is the creation of a NR framework for the prediction of quality metrics for IR image sequences.

The remainder of this paper is organized as follows: firstly, the I-frames analysis is presented, secondly, the VQA features estimation procedures, the machine learning (ML) algorithm and feature selection procedure are briefly described and finally the results are discussed.

State of the Art

Søgaard et al. [2], presented a NR pixel based (PB) video quality assessment (VQA) method by analyzing I-frames in H.264 and MPEG-2 coded sequences. In [2], a NR PB PSNR estimation algorithm is presented for visible video sequences compressed with h.264/AVC.

In [4], an HEVC based NR PB VQA technique is proposed. This paper also uses a Machine Learning (ML) algorithm, Elastic net, to predict the subjective video quality scores using HEVC features. The experiments were carried out for visible sequences which are different from IR sequences.

In contrast of the PB approaches, M. Shahid et al. [5] presented a NR bitstream based (BB) VQA technique using least square support vector machines (LS-SVM) for predicting the quality of H.264/AVC coded visible videos. The features used include percentage of Intra; Inter coded macroblock, skipped macroblock and so on. The experiments were carried out using low resolution CIF and QCIF visible video sequences.

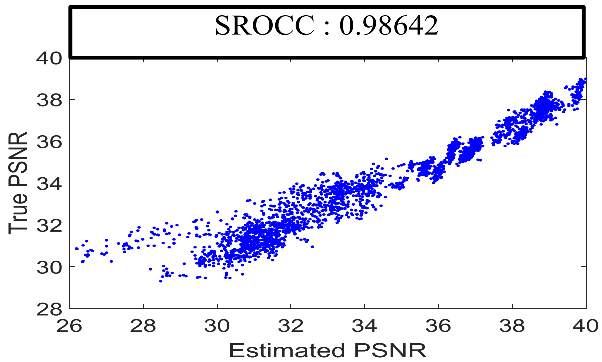
Goodall et al. [3] focus on tasking applications for uncompressed IR data. In this paper they focus on how noise and blur influence the perceived quality of IR images. Therefore they are not predicting metrics but quality grades.

In this work we aim for a BB NR estimation of quality metrics for H.264 compressed IR video sequences for low complexity UAV application.

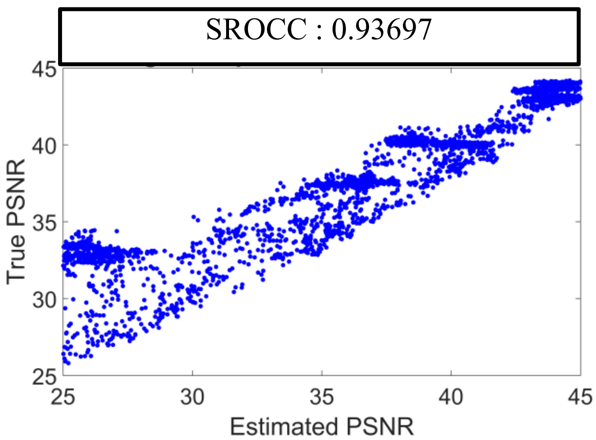
H.264 I Frame Analysis

The quantization parameter (QP) and an estimation of PSNR are used in this paper as features. It should be mentioned here that the PSNR as such is not our only goal but it is used to build the features necessary for the machine learning (ML) algorithm to predict the quality score. We have calculated different kinds of features from the QP values and estimated PSNR, corresponding to various spatial poolings. The prime objective of this paper is to design a low complexity application for UAVs. Therefore we use a BB approach for low delay coding of image sequences. The QP values and transform coefficients are extracted by parsing the bitstream. Then the Mean Square Error (MSE) and estimated PSNR are calculated using these values. In addition, the framerate of the IR sequences is 9 Hz which means that the frames do not have so much overlap with each other. The motion between

Fig. 1 First image of the 5 IR energy leakage videos (flight 1 to flight 5) from UAV



(a) 5 drone sequences



(b) 3 ground sequences

Fig. 2, Scatter plot of the NR BB PSNR estimation based on analysing the I-frames of IR sequences.

consecutive frames is larger than at a standard video frame rate (25Hz). Using P- or B- frames would increase the computational complexity. Therefore, the method has been designed for I–frame analysis. In the present paper, IR sequences were encoded with I-frames only using H.264/AVC. Below we describe the PSNR estimation for each I frame.

PSNR Estimation

In this subsection, we briefly describe the NR BB PSNR estimation method for I-frames in H.264/AVC IR videos. Although the PSNR estimation is not our only goal, it is used as an intermediate result for features generation for VQA. There are many works on NR PSNR estimation for H.264 compressed videos [1,2,6]. We applied the procedure described in [2], which is described in more details in the original publication. As detailed in [1,2], suppose we know the distribution of the original transform coefficient, in this case the local mean squared error ε_k^2 of a block at the kth coefficient band can be calculated by using the quantized value X_j

TABLE I

Number of features calculated for Machine Learning

	Features	***# of features***
QP	***High cluster U_h*** *Low cluster U_l* ***Weighted average U_ω*** ***Weight ω (averaging by U_h and U_l)*** ***Standard deviation(SD) σ*** ***Mean μ*** ***Maximum differences*** ***SD of good cluster*** *SD of bad cluster*	*9*
Estimated PSNR	*Same as above*	*9*
IQA	***SD μ and Mean μ***	*2*
SI/TI	***Maximum element of SI*** ***Minimum element of SI*** *Mean of SI σ* *SD of SI μ*	*7*
	Total	*27*

$$\varepsilon_k^2 \approx \sum_j P_k(X_j)\frac{\left(\int_{a_j}^{b_j} f_x(x)\,(X_j-x)\right)^2 dx}{\int_{a_j}^{b_j} f_x(x)\,dx} \quad (1)$$

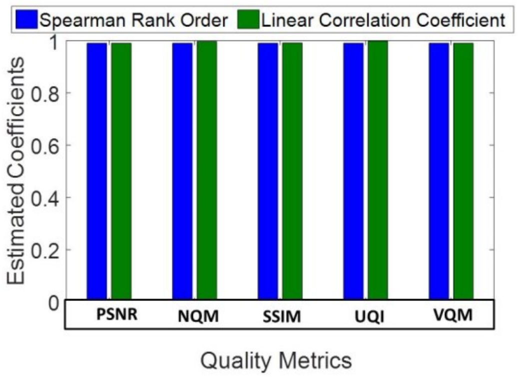
where, $P_k(X_j)$ is the ratio of the transform coefficients in band k which is inside the jth quantization interval belonging to X_j, $f_x(x)$ is the original coefficient data distribution in the quantization interval limits a_j and b_j .

We use single Cauchy distribution for each k to model the original distributions of the transform coefficients. The shape parameter γ_k of each Cauchy distribution at band k can be calculated by the percentage P_0 of coefficients which is smaller than $\alpha . qs$ in the reconstructed coefficients [2]. The local mean square error can be estimated by a given Cauchy distribution. Finally, with this MSE value, the PSNR can be estimated by the following equation as:

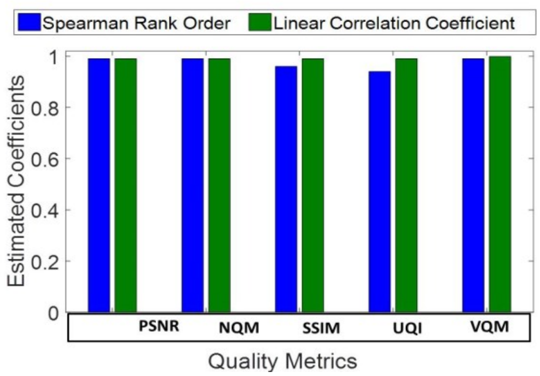
$$MSE_{est} = \frac{1}{16}\sum_{k=1}^{16}\sum_j \hat{\varepsilon}_{k,j}^2 \quad (2)$$

$$PSNR_{est} = 10log_{10}\frac{255^2}{MSE_{est}} \quad (3)$$

The estimated quality was compared with the true PSNR which were calculated from the first 200 frames of the IR sequences.



(b) 8 IR (5 drones and 3 ground)



(b) 5 drone IR energy leakage sequences

Fig. 3 Accuracy and estimated coefficients for prediction of full reference metrics using the proposed algorithm

Figure.2 shows that the estimated PSNR gives good results with a high SROCC (Spearman Rank Order Correlation Coefficients) for 5 IR energy leakage provided by Drone systems, Denmark and 3 ground IR sequences (among those, one is taken from the database made available by INO in Canada). More detailed description of the videos is available in Table II.

Features and Machine Learning

In this section, we describe the feature selection procedures using H.264/AVC codec analysis. In this paper, we compute features based on the true QP and estimated PSNR. Moreover, we compute Spatial Perceptual Information (SI) and Temporal Perceptual Information(TI) [13] and Image Quality Assessment (IQA) features. IQA features are estimated by modifying the BRISQUE features [14] which are based on natural scene statistics. In our work we use the implementation described in [2]. These features are used as an input to machine learning.

For prediction of quality measures, there are many kinds of machine learning algorithms such as Support vector Machine used for regression, known as Support Vector Regression (SVR), Elastic net and neural networks [7,8]. We use SVR in this paper to predict the objective quality grades.

Machine Learning Algorithm

Machine learning is used in our work for mapping the features to the quality score. We apply the SVR known as ε-SVR. The implementation is fully described in [2, 7]. Our goal is to find a function of feature vector x:

$$f(x) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)K(x_i, x) + \beta \quad (4)$$

where $(\alpha_i^* - \alpha_i)$ are called solution values, $K(x_i, x)$ is the kernel and β is an offset. The feature vectors above are x_i where $(\alpha_i^* - \alpha_i)$ are non-zero or so called support vectors. In our work we adapt the implementation described in [2]. The radial basis function is used in our implementation as a kernel function:

$$K(x_i, x) = e^{-\omega||x_i - x||^2} \quad (5)$$

In the training phase for this model, we need to search the optimal value of three parameters such as cost, ε in the SVR formulation and the ω in the radial basis function, in a 3-dimensional grid search.

Features Selection

For H.264/AVC codec analysis we measure 18 features based on true QP_{true} and estimated PSNR for I frame analysis. We take into consideration the Image Quality Assessment (IQA) features. We adapt this features as described in [2]. Additionally the SI/TI measures from distorted IR sequences are considered for VQA.

Feature selection is performed before applying the SVR to enhance the prediction ability. We adopt the feature selection procedure described in [9]. The prime idea of feature selection is to compare the weight of features in the model to that of random "noise features", and not to use the features with a weight lower than the average of the weights of the noise features. In addition, the features are also deselected where weights are negative in some folds and positive in other folds of the cross-validation.

In our evaluation, the experiments were carried out for three scenarios which are: codec only, codec with SI/TI and codec with IQA for 8 IR (5 drones and 3 grounds) and 5 IR sequences. In the case of codec only, feature selection procedures have provided 2 to 6 features (out of 18 features) for metric prediction. In the case of codec with SI, in most cases 2 SI features were selected whereas TI features were totally ignored for metric prediction. In the case of codec with IQA, 1 to 2 features were selected as well. The selection process underwent by the features is similar to that in [9] and the selected features are those marked in bold in Table 1. Overall, using features selection procedure, results significantly improve with fewer features.

Results

In our performance analysis, we use 5 IR energy leakage videos which are shown in Fig. 1. The Sequences were captured by an UAV. Additionally, 3 ground IR sequences are used for our experiment. In our experiments, the GOP structure used is only I-frames. The encoding parameters are shown in Table II, each of the videos have been encoded at 200, 1000, 3000 and 5000 kbps for 5 energy leakage sequences and 50, 200,400 and 1000 kbps for 3 ground IR sequences (the lower bitrate is explained by a lower resolution of the ground videos). Therefore a total of 20 sequences for 5 energy leakage and 12 for 3 ground IR sequences have been used for evaluating the quality prediction model. For 5 IR energy leakage sequences, 80% of the sequences have been used for training and the remaining 20% for testing (the folds being split by different content). In our experiment, we carried out experiment for 8 sequences (5 drones and 3 grounds). For 8 IR sequences, 7 contents out of 8 are used for training which mean 28 sequences for training and remaining 4 for testing. The proposed model has

TABLE III

CROSS-VALIDATION RESULTS OF SROCC, LCC AND RMSE USING FR METRICS

8 IR (5 IR and 3 ground)	PSNR			NQM			SSIM		
	SROCC	LCC	RMSE	SROCC	LCC	RMSE	SROCC	LCC	RMSE
Codec Only	**0.99**	**0.99**	**1.32**	**0.99**	0.99	7.21	**0.99**	0.96	0.03
IQA & Codec features	**0.99**	**0.99**	1.40	**0.99**	0.9934	4.33	**0.99**	**0.9913**	**0.03**
Codec & SI	**0.99**	**0.99**	1.56	**0.99**	**0.9965**	**3.71**	**0.99**	**0.9913**	0.04
	UQI			VQM					
Codec Only	0.95	0.97	**0.07**	0.95	0.98	0.04			
IQA & Codec features	**0.99**	**0.9986**	0.1512	**0.99**	**0.99**	**0.01**			
Codec & SI	0.93	0.90	0.1030	**0.99**	**0.99**	0.03			

(a) 8 IR sequences (5 drones and 3 ground)

5 IR	*PSNR*			*NQM*			*SSIM*		
	SROCC	*LCC*	*RMSE*	*SROCC*	*LCC*	*RMSE*	*SROCC*	*LCC*	*RMSE*
Codec Only	*0.97*	*0.99*	***0.65***	*0.97*	***0.99***	*1.02*	*0.95*	***0.99***	***0.01***
IQA & Codec features	***0.99***	***0.99***	*0.74*	***0.99***	***0.99***	*0.70*	***0.96***	***0.99***	***0.01***
Codec & SI	*0.98*	*0.99*	***0.65***	*0.97*	***0.99***	***0.60***	***0.96***	***0.99***	***0.01***
	UQI			VQM					
Codec Only	*0.93*	***0.99***	***0.02***	0.94	0.9939	**0.03**			
IQA & Codec features	***0.94***	***0.99***	***0.02***	**0.99**	**0.9987**	0.08			
Codec & SI	*0.93*	***0.99***	***0.02***	0.96	0.99	0.06			

(b) 5 IR energy leakage sequences

been tested for four different image FR metrics: PSNR, SSIM [10], UQI [11] and NQM [12] and one video quality metric VQM [15]. The temporal pooling applied for the image metrics is averaging. The results are reported by Spearman Rank Order Correlation Coefficients (SROCC), Linear Correlation Coefficients (LCC) and Root Mean Square Error (RMSE) which describe the prediction accuracy. The regression values can be seen in Figure 3. The proposed model is able to predict the quality metrics for H.264 compressed IR sequences. The results are shown in Table III.

TABLE I

INFRARED SEQUENCES AND X264 ENCODER PARAMETERS

Name	Setting
Resolution (5 IR)	640X512
Resolution (3 IR)	320X256 (Road & Crossing Road) and 328X254 (INO database)
Frames	300
FPS	9 and 10(INO database)
Bitrates [kbps] (5 IR)	200, 1000, 3000, 5000
Bitrates [kbps] (3 IR)	50,200,400,1000
Encoding	I – frame

For 8 IR sequences, SROCC for codec are 0.99, 0.99, 0.99, 0.95 and 0.95 for PSNR, NQM, SSIM, UQI and VQM respectively. Adding SI [13] and IQA features slightly increase the prediction accuracy of SROCC. In the case of Codec features with IQA, the reported SROCC are 0.99, 0.99, 0.99 and 0.99 for PSNR, NQM, SSIM, UQI and VQM respectively. Both LCC and RMSE results support the SROCC results.

In the case of 5 drone IR sequences, the reported SROCC for codec features are 0.97, 0.97, 0.95, 0.93 and 0.94 for PSNR, NQM, SSIM, UQI and VQM respectively. In the case of Codec features with IQA, the reported SROCC are 0.99, 0.99, 0.96, 0.94 and 0.99 for PSNR, NQM, SSIM, UQI and VQM respectively. It can be concluded that, most of the cases the codec features combined with IQA give better results compared to codec only and codec with SI.

Moreover, as shown in Figure 4, estimated quality values are compared with image and video metrics.

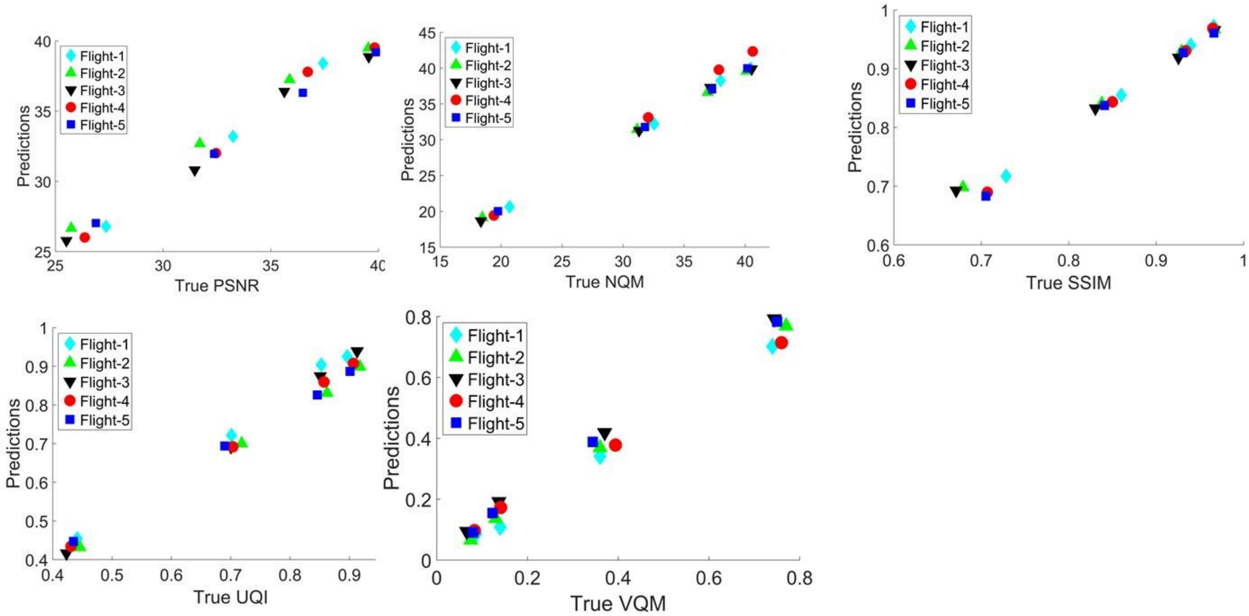


Fig. 4 Scatter plot of the NR quality estimation based on analysing the I-frames of 5 IR sequences.

Conclusion

In this paper, a framework for a BB NR prediction of quality metrics of IR sequences for low complexity UAV application was presented. In order to achieve this goal, we have presented a PSNR estimation technique based on QP and transform coefficients parsed from a bitstream. Since our target application is low complexity UAV applications, the codec analysis focuses on coding with I-frames only to perform low delay coding. The extracted QP, estimated PSNR, codec features combining with SI/TI and IQA features are used as an input for SVR to predict the quality metrics. Moreover, we have applied feature selection procedure to increase the prediction accuracy with fewer features.

In our evaluation, 8 IR sequences (5 drones and 3 grounds) and 5 drone sequences are used which were coded with H.264/AVC. For both of these sequences, we achieve high SROCC results. The experimental results prove the validity of our proposed work.

References

[1] X. Huang, J. Søgaard and S. Forchhammer, "No-reference video quality assessment by HEVC codec analysis," 2015 Visual Com munications and Image Processing (VCIP), Singapore, 2015, pp.1-4.

[2] J. Søgaard, S. Forchhammer and J. Korhonen, "No-Reference Video Quality Assessment Using Codec Analysis," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 10, pp. 1637-1650, Oct. 2015.

[3] T. R. Goodall, A. C. Bovik and N. G. Paulter, "Tasking on Natural Statistics of Infrared Images," in IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 65-79, Jan. 2016.

[4] Huang, Xin; Søgaard, Jacob; Forchhammer, Søren, "No-reference pixel based video quality assessment for HEVC decoded video", In: Journal of Visual Communication and Image Representation, Vol. 43, 2017, p. 173-184.

[5] M. Shahid, A. Rossholm and B. Lövström, "A no-reference machine learning based video quality predictor," 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt am Wörthersee, 2013, pp. 176-181.

[6] A. Eden, "No-Reference Estimation of the Coding PSNR for H.264-Coded Sequences," in IEEE Transactions on Consumer Electronics, vol. 53, no. 2, pp. 667-674, May 2007.

[7] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, 2011, Art. ID 27. [Online]. Available: http:// www. csie. ntu.edu.tw/~cjlin/libsvm/

[8] K. Sjöstrand, L. Harder Clemmensen, R. Larsen, B. Ersbøll "SpaSM: A Matlab Toolbox for Sparse Statistical Modeling",Fifth Intl Workshop Quality Multimedia Experience, Klagenfurt, Austria, July, 2013

[9] Claire Mantel, Jacob Søgaard, Søren Bech, Jari Korhonen, Jesper Melgaard Pedersen, and Søren Forchhammer, "Modeling the quality of videos displayed with local dimming backlight at different peak white and ambient light levels," IEEE Transactions on Image Processing, vol. 25, no. 8, pp. 3751–3761,2016.

[10] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.

[11] Zhou Wang and A. C. Bovik, "A universal image quality index," in IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002.

[12] D. Niranjan et al. "Image Quality Assessment Based on a Degradation Model," IEEE Transaction on Image Processing, VOL. 9, NO. 4, April 2000.

[13] Subjective Video Quality Assessment Methods for Multimedia Applications, document Rec. ITU-T P.910, 2008.

[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[15] M.H. Pinson, S. Wolf A new standardized method for objectively measuring video quality IEEE Trans. Broadcast., 50 (3) (2004), pp. 312-

Acknowledgment

This project was partially funded by the EUDP project "Energy efficient drone inspection system for spotting leakages" grant number 64015-0072.

Author Biography

Kabir Hossain received the B.E. degree in computer engineering and M.E. degree in electronics and computer engineering from Chonnam National University, South Korea in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree with the Coding and Visual Communication Group, Department of Photonics engineering of the Technical University of Denmark. His research interests include image and video coding, image and video quality assessment and machine learning.

Claire Mantel received the M.S. and Ph.D. degrees in signal processing from Grenoble Polytechnic Institute, France, in 2007 and 2011, respectively. She is currently working as a Researcher at the Department of Photonics Engineering of the Technical University of Denmark. Her research interests include image and video coding and visual quality assessment.

Søren Forchhammer received the M.S. degree in engineering and the Ph.D. degree from the Technical University of Denmark, Lyngby, in 1984 and 1988, respectively. Currently, he is a Professor with 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, two-dimensional information theory, and visual communications.