International Journal of Multimedia Data Engineering and Management
Volume 12 • Issue 3

An Image Quality Adjustment Framework for Object Detection on Embedded Cameras

Lingchao Kong, University of Cincinnati, USA*
Ademola Ikusan, University of Cincinnati, USA
Rui Dai, University of Cincinnati, USA
Dara Ros, University of Cincinnati, USA

ABSTRACT

Automatic analysis tools are ubiquitously applied on wireless embedded cameras to extract high-level information from raw data. The quality of images may be degraded by factors such as noise and blur introduced during the sensing process, which could affect the performance of automatic analysis. Object detection is the first and the most fundamental step for the automatic analysis of visual information. This paper introduces a quality adjustment framework to provide satisfactory object detection performance on wireless embedded cameras. Key components of the framework include a blind regression model for predicting the performance of object detection and two distortion type classifiers for determining the presence of noise and blur in an image. Experimental results show that the proposed framework achieves accurate estimations of image distortion types, and it can be easily applied on embedded cameras with low computational complexity to improve the quality of captured images.

KEYWORDS

Embedded Cameras, Image Distortion, Image Quality Adjustment, No-Reference, Object Detection

INTRODUCTION

The ubiquitous deployment of wireless embedded camera sensors for various imaging applications has resulted in an exponential growth of image and video data. There is an emerging need to automatically extract meaningful information from the raw data generated by camera sensors. For this reason, automatic image and video analysis tools, which are enabled by machine learning and other mathematical approaches, have been applied in computing devices to integrate data into information. For imaging applications, automatic analysis results could be used to trigger alarms for abnormal events, provide situational awareness to human operators, and facilitate automatic control of physical systems, among many uses.

To assure the accuracy of automatic analysis methods, the embedded cameras in a system should provide images with satisfactory quality. Many image quality assessment (IQA) models have been designed to evaluate the perceptual quality of images judged by human users (Wang, Bovik, Sheikh, & Simoncelli, 2004). However, the quality of an image evaluated by an automatic analysis tool is not necessarily sensitive to the same factors that drive human perceptions. Characteristics of the human visual system (HVS) such as the visual attention and the contrast sensitivity mechanisms are considered in perceptual image quality assessment. Due to the foveation feature of the HVS, at an instance, only

DOI: 10.4018/IJMDEM.291557 *Corresponding Author

This article published as an Open Access Article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

a local area in the image can be perceived with high resolution at typical viewing distances (Gu, et al., 2016), and the HVS is sensitive to the relative rather than the absolute luminance change (Wang, Bovik, Sheikh, & Simoncelli, 2004). In contrast, automatic analysis methods executed by computing devices can have a global "view" of an image and "perceive" the absolute luminance change precisely. A few recent studies have explored the unique characteristics of image quality for automatic analysis algorithms. For example, a study of motion imagery quality for tracking in airborne reconnaissance systems shows that factors such as temporal jitter, level of noise, and edge sharpness have a strong effect on the accuracy of target detection, and unlike human users, the automatic detection algorithms are less sensitive to spatial resolution (Irvine & Wood, 2013). In our recent work (Kong, Dai, & Zhang, 2016), we have found that the performance of object detection algorithms can be affected by the quality of the background areas, which differs from the perception of human beings who can easily detect moving objects from blurred backgrounds. Therefore, the quality of images evaluated by automatic analysis requires further investigation, and new quality models are needed for automatic analysis algorithms.

For wireless imaging systems, an automatic analysis module could be deployed either on compressed videos in a central server or on uncompressed videos at local cameras. Some recent works (Kong & Dai, 2016; Kong & Dai, 2017; Wang, Li, Zhang, & Yang, 2018; Kong & Dai, 2018) have studied the impact of video compression on the accuracy of analysis algorithms. Apart from the distortion introduced by lossy compression, the quality of an image or a video could be degraded by factors such as blur or noise during the image sensing process. These factors should also be taken into consideration to evaluate image quality. It is challenging to obtain general solutions on image quality for automatic analysis since analysis methods are mostly dependent on specific applications. However, the common and the most fundamental step for automatic analysis is object detection, as the detected objects are the basis for higher-level analysis tasks such as object tracking and behavior understanding. Moreover, many existing embedded camera platforms, such as CITRIC (Chen, et al., 2013) and SWEETcam (Abas, Porto, & Obraczka, 2014), have incorporated light-weight object detection algorithms on board. For embedded camera platforms, it would be helpful if the quality of an image for object detection could be predicted and adjusted with light-weight solutions.

In this paper, we propose an image quality adjustment framework to provide satisfactory object detection performance for imaging applications based on embedded cameras. The framework includes a blind regression model based on a bagging ensemble of trees for predicting the performance of object detection on an image and two distortion type classifiers based on support vector machines for determining the presence of noise or blur in the image. The regression model and the classifiers utilize local features in an image, such as edge and oriented gradient, and global features, including image gradient, image contrast, and estimated object size. All of the features could be easily obtained from an image, providing a light-weight solution for embedded cameras. The regression model and the classifiers are trained using a large number of images with different scene characteristics and three types of distortions, including imaging noise, out-of-focus blur, and motion blur. Their performances are evaluated through extensive experiments on separate test data sets. Our preliminary results on the regression model have been presented in our recent publications (Kong, Ikusan, Dai, & Zhu, 2019; Kong, Ikusan, Dai, Zhu, & Ros, 2019). New contributions in this article include a systematic image quality adjustment framework, two distortion type classifiers, comprehensive experimental results, and discussions based on a larger data set.

The rest of this paper is organized as follows. First, related work in quality evaluation and quality adjustment for automatic analysis are summarized, followed by a description of the data set and the object detection measure used for this study. Then, the proposed image quality adjustment framework is introduced in details. Finally, performance evaluation results are discussed and concluding remarks are given.

RELATED WORK

There is a rich literature on enhancing the perceptual quality of images. Based on spectral total variation, an image fusion and enhancement framework extracted the main features of input images and achieved improvement in edge details and contrast (Zhao, Lu, & Wang, 2017). For low-light image enhancement, the robust Retinex model (Li, Liu, Yang, Sun, & Guo, 2018) additionally considered a noise map compared with the conventional Retinex model to improve the performance of enhancing low-light images accompanied by intensive noise. Wang et al. (2015) proposed a guided image contrast enhancement framework based on retrieved images in the cloud, in which context-sensitive contrast and context-free contrast were jointly improved via solving a multi-criteria optimization problem. A no-reference IQA model through analysis of contrast, sharpness, brightness, etc. was proposed by Gu, Tao, Qiao, & Lin (2017). Then, a robust image enhancement framework through histogram modification to rectify image brightness and contrast is established based on the NR-IQA model.

There are a few works on adjusting image quality for automatic analysis tasks. Based on the quantification of degradation in face detectors' performance, a new set of features were proposed for robust face detection that could augment face-indicative features with perceptual quality-aware spatial natural scene statistics features (Gunasekar, Ghosh, & Bovik, 2014). Based on aggregating IQA features that are more robust to image quality degradation, a similar enhancement strategy was employed by face recognition in infrared images (Pulecio, Benitez-Restrepo, & Bovik, 2017). A closed-loop computing framework of enhancing image steganography through optimizing picture quality was proposed by Lin, Chang, & Lie (2010). To improve poor license plate recognition due to the low quality of images, a new mathematical model based on the Riesz fractional operator for enhancing details of edge information in license plate images was proposed to improve the performances of text detection and recognition methods (Raghunandan, et al., 2017).

The aforementioned studies proposed quality enhancing or control approaches for either perceptual quality or specific applications like face recognition, image steganography, and plate recognition. However, two more challenges in image quality adjustment remain unsolved. First, there are different types of distortion during the imaging process, and the types of distortion, e.g., noise or blur, should be determined before applying image enhancement algorithms. Second, the estimation of distortion types should be achieved through low complexity considering the limitation of processing on embedded cameras. Existing studies on noise and blur estimation involve intensive computation (Pyatykh, Hesser, & Zheng, 2013; Yan & Shao, 2016). Our work aims to advance the state of the art by addressing these two challenges.

DATA SET AND OBJECT DETECTION MEASURE

In this section, we introduce the data set, the object detection algorithms and the measure used for evaluating object detection accuracy in our study.

Figure 1. Snapshots of video data set

Data Set

Ten high-resolution videos are chosen to cover diverse scenes, light conditions, and object scales. Five videos are selected from the Multiple Object Tracking (MOT) dataset (Milan, Leal-Taixé, Reid, Roth, & Schindler, 2016), and five videos are selected from the Duke Multi-Target Multi-Camera Tracking (DM) dataset (Ristani, Solera, Zou, Cucchiara, & Tomasi, 2016). Most of the videos have 1920×1080 resolution except for one video with 640×480 size. The average number of images in each video is around 740. The snapshots of these sequences are shown in Figure 1.

Blur and noise are major factors that degrade imaging quality for surveillance or mobile cameras. To explore the impact of blur and noise during the image sensing process on the performance of object detection, we have generated distorted video sequences based on the original 10 videos, including videos with out-of-focus blur, motion blur, and imaging noise. For each type of distortion, 5 distortion levels are simulated: low, medium, high, higher, and extreme levels.

We have also included reduced spatial resolution versions of the original videos to count in the effect of spatial resolution on object detection.

- The out-of-focus blur is generated by 2D circularly symmetric Gaussian blur kernels with standard deviations of [1.2, 2.5, 6.5, 15.2, 33.2] for 5 levels, respectively.
- The motion blur is simulated to approximate the linear motion of a camera at an angle of 45 degrees by [5, 12, 20, 40, 100] pixels.
- Imaging noise is simulated by white Gaussian noise, which has variances of [0.001, 0.006, 0.022, 0.088, 1].
- For reduced spatial resolution, 1:2 and 1:4 down-sampling rates are applied.

For each original sequence shown in Figure 1, we have generated a total number of 17 distorted sequences, including 2 videos from reduced spatial resolution and 5 videos from each of the three distortion types. This results in a total number of 180 video sequences, which corresponds to 133344 images in total for the entire data set.

OBJECT DETECTION ALGORITHMS AND MEASURE

There are two categories of object detection algorithms in the field of computer vision: background modeling based detection and objects modeling based detection. Algorithms based on background modeling require multiple frames to learn an initial background, while methods based on object modeling could work on a single image. Our study targets at enabling a wireless embedded camera to adjust its sensing strategy dynamically based on the predicted quality and its available resources. Therefore, the property of evaluating an individual image's quality is desired, and we focus on object modeling based detection. Furthermore, we consider the scenario that the embedded camera performs local object detection in a fast manner, so low-complexity object detection algorithms are preferred.

While deep learning solutions, such as those based on Convolutional Neural Networks (CNN), are gaining popularity in vision-based applications, their application on embedded devices involves significant complexity compared to traditional hand-crafted approaches (Suleiman, Chen, Emer, & Sze, 2017; Sze, Chen, Emer, Suleiman, & Zhang, 2017; Zhu, Samajdar, Mattina, & Whatmough, 2018). Based on the analysis of two chips using 65nm technology in (Suleiman, Chen, Emer, & Sze, 2017; Sze, Chen, Emer, Suleiman, & Zhang, 2017), the computational complexity of learned approaches (i.e. AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) and VGG (Simonyan & Zisserman, 2014)) is 36.9 and 871.9 times over the typical hand-crafted approach (HOG (Dalal & Triggs, 2005)), and the CNN-based chip consumes several orders of magnitude more energy than the HOG-based chip. Zhu, Samajdar, Mattina, & Whatmough, (2018) found that today's CNN-based approaches have at least one order of magnitude higher compute requirements than accommodated in a mobile device.

Considering the limited energy and computing resources in embedded systems, three typical lightweight algorithms based on object modeling are considered in this work:

- Histograms of Oriented Gradients (HOG) (Dalal & Triggs, 2005);
- Discriminatively Part Models (DPM) (Felzenszwalb, Girshick, McAllester, & Ramanan, 2010);
- Locally Decorrelated Channel Features (LDCF) (Nam, Dollár, & Han, 2014).

With the goal of predicting the performance of object detection once an image is captured, we evaluate the object detection accuracy of each image. We make use of the revised Frame Detection Accuracy (rFDA) measure proposed in our recent work (Kong, Ikusan, Dai, Zhu, & Ros, 2019). rFDA is the average of frame detection accuracy based on different thresholds (T) of detection confidence levels (C). For one image, where there are N_G ground-truth objects G and N_D detected objects D, N_m is the number of mapped object pairs, *rFDA* is given by

$$rFDA = \sum_{j=1}^{N_m} \left(\frac{\sum_{i=1}^{N_{T_j}} \frac{G_i \cap D_i}{G_i \cup D_i}}{\frac{N_G + N_D}{2}} \right) / N_m$$

where N_{Tj} is the number of true positives when the threshold of detection confidence T_j equals to C_j, $j \in \{1,...,N_m\}$, and C_j denotes the detection confidence level of the j-th mapped detected object.

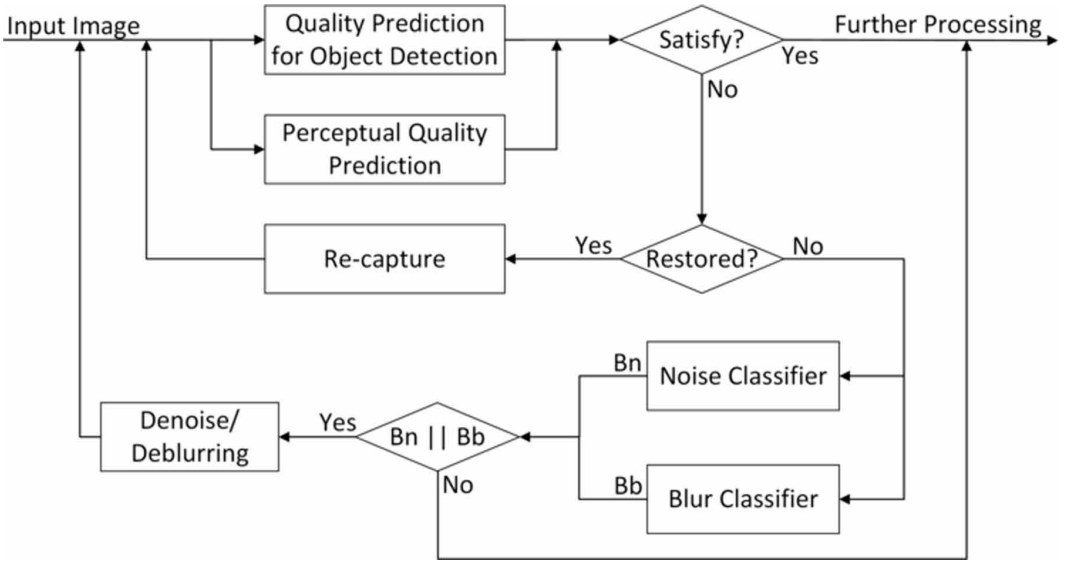
The performance of rFDA was investigated systematically in the entire dataset for HOG, DPM, and LDCF detectors in our recent work. rFDA can reflect the detection performance for individual images and achieve stabilized measurements. The detection performances measured in rFDA of the three different detectors are coherent. Thus, we focus on estimating the average results of the three detectors in this work.

IMAGE QUALITY ADJUSTMENT FOR OBJECT DETECTION

We propose a quality adjustment framework to provide satisfactory image quality for object detection during the image sensing process. The components of the framework are shown in Figure 2. Once an image is captured, both the quality for object detection and the perceptual quality of the image are estimated, as the image may be used for further automatic analysis or be delivered in front of human users. The prediction of perceptual quality could be achieved through existing no-reference perceptual quality models, such as BRISQUE (Mittal, Moorthy, & Bovik, 2012) and BLIINDS-II (Saad, Bovik, & Charrier, 2012); however, the evaluation of object detection quality requires further studies. We propose to build a new image quality model to estimate the performance of object detection. If the image has satisfactory perceptual and object detection quality, it will be further processed or analyzed. Otherwise, the framework will determine if there is any noise or blur in the image, which are common types of distortions in the imaging process. If so, image restoration approaches for removing noise or blur will be applied to enhance its quality. If the restored image still cannot provide satisfactory quality, an image re-capturing action will be executed. Although several existing denoise and deblurring algorithms could be applied here to restore the images (Pang, Zhou, Wu, & Li, 2019; Bai, Cheung, Liu, & Gao, 2019), there is a lack of mechanisms to distinguish noisy or blurred images with normal ones. To solve this problem, we propose to build a noise classifier and a blur classifier.

In the rest of this section, we explain the three core components in the proposed framework: quality prediction for object detection, blur classifier, and noise classifier. Using the aforementioned data set, we apply supervised learning algorithms to build these components. We introduce a total

Figure 2. The quality adjustment framework



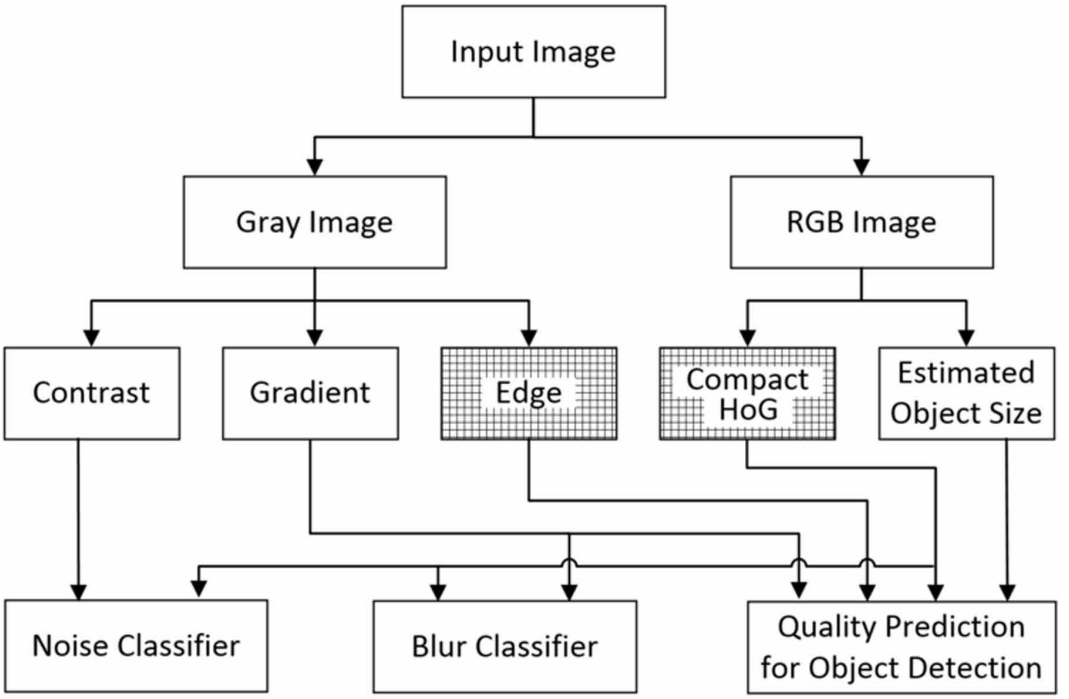
number of 18 local and global features, all of which could be obtained from an image with low computational complexity. Figure 3 illustrates the high-level relationship of the features. All of them are extracted from the converted gray image and the RGB image, and they can be shared among the three core components. The "Edge" and "Compact HOG" modules with grids in the background denote that these features are collected locally. Other features are summarized over an entire image.

BLIND MODEL FOR OBJECT DETECTION QUALITY

We build a blind/no-reference regression model to predict the performance of object detection on individual images, given by rFDA defined in the previous section. As illustrated in Figure 3, the regression model utilizes four categories of features: gradient (No. (1)-(4)), compact HOG (No. (5)-(8)), edge (No. (9)-(12)), and estimated object size (No. (13)), which are listed as follows:

(1) meanGmag: the average of gradient magnitude;
(2) stdGmag: the standard deviation of gradient magnitude;
(3) meanGdir: the average of gradient direction;
(4) stdGdir: the standard deviation of gradient direction;
(5) hog_mm: the average of every blocks' average frequency of HOG histogram's bins (w_m);
(6) hog_ms: the standard deviation of every blocks' w_m;
(7) hog_sm: the average of every blocks' average frequency variation level of HOG histogram's bins (w_s);
(8) hog_ss: the standard deviation of every blocks' w_s;
(9) edge_mm: the average of every blocks' average edge, which is obtained by applying convolution masks (Sobel operator) to the image in horizontal and vertical directions;
(10) edge_ms: the standard deviation of every blocks' average edge;
(11) edge_sm: the average of every blocks' standard deviation of edge;
(12) edge_ss: the standard deviation of every blocks standard deviation of edge;
(13) estimated object size: the estimated object scale relative to the size of the image, which also affect the performance of object detection.

Figure 3. The architecture of feature extraction



More details on the computation of features No. (1)-(12) are given in our recent work (Kong, Ikusan, Dai, Zhu, & Ros, 2019). For feature No. (13), object size is estimated after image segmentation based on Otsu's method. Four feature channels (the gradient's magnitude, red color channel, green color channel, and blue color channel of the image) are employed together to determine the optimal threshold for binarizing the gray image. This method is used because it can achieve good estimation through low computational overhead.

Utilizing these features, the bagging ensemble of trees is then employed to train a regression model to estimate object detection performance.

Blur And Noise Classifiers

Blur and noise in certain extents usually accompany with the image formation procedure (Wang & Tao, 2014), which can be described as:

$$g = x * h + n,$$

where g is the observed distorted image, x is the latent clean image, which does not exist actually, * denotes the convolution operator, h is the point spread function (PSF), or known as the blur kernel, and n represents an additive random noise. An image g is distorted when the latent image x is convolved with the PSF h, i.e., individual image pixels absorb information from multiple sources, and combined with an additive noise n. Different PSFs generate different types of blur, e.g., motion blur and out-of-focus blur. A typical noise on natural images is an additive random noise modeled as a zero-mean Gaussian distribution.

Features For Blur Classification

Blur, including motion blur and out-of-focus blur, can smooth the boundary information in an image. Thus, statistical features based on image gradient, a good indicator for the variance of image intensities, are extracted from the gray image. Except for the average and standard deviation based features introduced previously, we introduce two concepts known as Skewness and Kurtosis to better describe the properties of the distribution. Skewness means asymmetry. A distribution, or data set, is symmetrical (the corresponding skewness is 0) if the data points are evenly distributed around its mean. The skewness of a random variable x is the third standardized moment, defined as

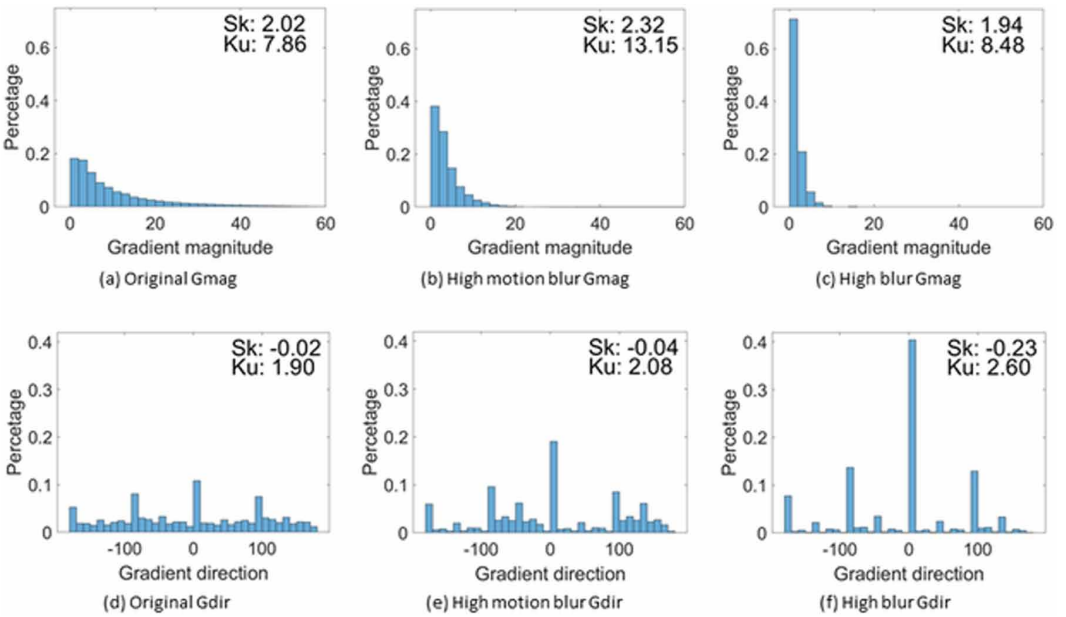
$$s = \frac{E\left(x-\mu\right)^3}{\sigma^3}$$

where μ is the mean value of x, σ is the standard deviation value of x, and $E(.)$ denotes the expectation operator.

Kurtosis is a measure of the sharpness or convexity relative to that of a normal distribution. The kurtosis equals 3 for a normal distribution. A high kurtosis value (larger than 3) indicates that the distribution has heavy tails or more outliers, and a low kurtosis value (less than 3) means the data set has light tails or a lack of outliers. The kurtosis of a random variable x is the fourth standardized moment, defined as

$$k = \frac{E\left(x-\mu\right)^4}{\sigma^4}$$

Figure 4. Distribution comparison of original and blur images



where μ is the mean value of x, σ is the standard deviation value of x, and $E(.)$ represents the expectation operator.

One sample of histograms for gradient of the original and the blur images is shown in Figure 4. The first row is the histograms of gradient magnitude (Gmag) for the original image, high blur version, and high motion blur version. The second row is the corresponding histograms of gradient direction (Gdir) calculated from images. The skewness (Sk) and kurtosis (Ku) for gradient's direction and magnitude are calculated and shown on the top-right corner of each figure. We notice that the shape/distribution for the histograms of blur version is different from the one of the original images, and such difference is also reflected in the skewness and the kurtosis values. For example, the Ku value of Gmag for high motion blur image (13.15) is almost double the one of Gmag for the original image (7.86); the Sk value of Gdir for high blur image (-0.23) is much larger than the one of Gdir for the original image (-0.02).

Thus, we calculate 4 more related global features based on the image gradient:

(14) skewGmag: the skewness of gradient magnitude;
(15) kurtGmag: the kurtosis of gradient magnitude;
(16) skewGdir: the skewness of gradient direction;
(17) kurtGdir: the kurtosis of gradient direction.

FEATURES FOR NOISE CLASSIFICATION

Noise has always been associated with the image acquisition procedure, and it is a setback for object detection and further analysis or processing. Due to the randomness of noise, it can cause arbitrary changes of intensities locally, which will bring more inconsistency of intensities compared with normal or natural images. Thus, we utilize one feature:

(18) image contrast.

It is introduced to depict the inconsistency of intensities. The image contrast is defined as

$$C = \sqrt{\sum_{x=1}^{M}\sum_{y=1}^{N}\frac{\left(I\left(x,y\right)-\mu\right)^{2}}{M\times N}}$$

where $I(x,y)$ denotes the intensity value of the gray image at location (x,y), μ is the mean value of the entire gray image, and $M \times N$ stands for image size.

BINARY CLASSIFIERS FOR BLUR AND NOISE

We propose to build binary classifiers to indicate whether there is blur or noise in an image. Considering the requirements of high accuracy and good robustness as well as the property of binary classification, we train two Support Vector Machines (SVM) as the blur classifier and the noise classifier. The mechanism of finding the maximal margin introduced by SVM can bring good tolerance for the classifier, which is a key point in our proposed framework. The training procedures of an SVM include transforming original data to a high-dimensional space using a kernel and constructing an optimal hyperplane to distinguish the transformed data into two classes.

Finally, the blur classifier is established based on 8 gradient related features (No. (1)-(4) and (14)-(17)) and 4 compact HOG features (No. (5)-(8)). The noise classifier is constructed based on the image contrast feature (No. (18)) and 4 compact HOG features (No. (5)-(8)).

Table 1. The setting of training and testing sets

Category	Video name	Image number	Percentage
Training set	MOT17-02, MOT17-10, MOT17-13, MOT15-02 DMcam01, DMcam02, DMcam04, DMcam08	100044	75.03%
Testing set	MOT17-04, DMcam06	33300	24.97%

PERFORMANCE EVALUATION

In this section, we first investigate the performance of the proposed framework on the aforementioned data set, which includes images of different types and levels of simulated distortions; we then extend our evaluation on a separate set of naturally distorted images.

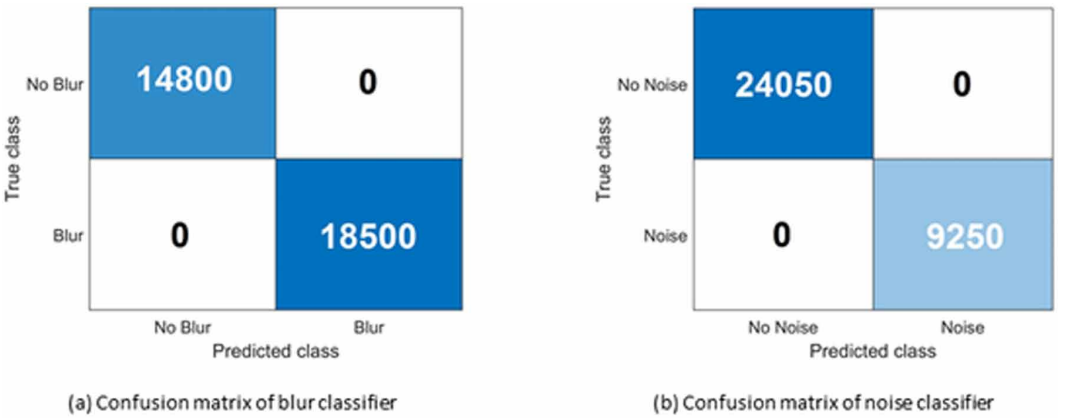
The entire data set in this study is divided exclusively into a training set and a testing set, which are described in Table 1. Through 5-fold cross-validation for the proposed quality model, 30 base trees and minimum leaf size of 8 are used to construct the ensemble of trees. For blur and noise classifiers, the box constraint parameter with 1 and the standardized predictor data are utilized to train the two linear kernel-based SVMs.

EVALUATION OF BLUR AND NOISE CLASSIFIERS

We use the same training and testing sets to train and test two linear kernel-based SVMs for the proposed blur and noise classifiers. The confusion matrices of the two classifiers are shown in Figure 5. The total amount of test images is 33300, which includes original images, 2 versions of down-sampled images, 5 distorted levels of out-of-focus blur, motion blur, and imaging noise. Since there are two types of blur and one type of noise, the number of blurry images is twofold the number of noisy images. In such a diversified data set, both the noise classifier and blur classifier can precisely distinguish noise and blur distortions among other interference factors. The accuracy and sensitivity for both classifiers reach 100%. The accurate classifications can help to determine whether denoise and/or deblurring algorithms could be applied to restore a distorted image.

All the features in the blur and noise classifiers could be obtained through light-weight computation, and it is worth noticing that the majority of the features for the classifiers are reused

Figure 5. Classification results for two classifiers of distortion types



from the features extracted for the proposed quality model. The entire computational complexity of the proposed quality model and two classifiers is compared with two no-reference IQA methods: BLIINDS-II and BRISQUE. Computational complexity is measured exclusively on a computer based on an Intel Xeon E5-2637 v3 (3.50GHz) processor running on a Windows 7 Enterprise operating system. It is evaluated by the average computational time over 800 images of the original 1080p resolution (1920×1080), half resolution (960×540), and quarter resolution (480×270) of DMcam06 video. As shown in Table 2, the total time consumption is still smaller than the BLIINDS-II and BRISQUE algorithms, which indicates that the proposed classifiers can be implemented on embedded cameras with low complexity.

Table 2. Average computational complexity measured in seconds

Algorithms	1920×1080	960×540	480×270
BLIINDS-II	158.987 (±1.491)	40.065 (±0.698)	10.140 (±0.202)
BRISQUE	0.785 (±0.070)	0.324 (±0.061)	0.235 (±0.053)
Proposed	0.473 (±0.028)	0.245 (±0.017)	0.153 (±0.014)

EVALUATION OF IMAGE QUALITY ADJUSTMENT

After determining the distortion type, the quality adjustment framework can call appropriate image restore algorithms, i.e., denoise and deblurring, to adjust image quality. We briefly introduce denoise and deblurring methods used, then we present the improvement of image quality adjustment for object detection.

We employ an efficient blind image deblurring method via dark channel prior (Pan, Sun, Pfister, & Yang, 2017), which is based on the observation that the dark channel, i.e. the minimum value in an image patch of blurred images, is less sparse than clean or sharp images. This constraint is added to the conventional optimization for estimating the latent image.

Image denoise can be conducted by linear filtering, and in this work, the Wiener filter is first deployed on three single color channels to suppress noise and to preserve edge, texture, and other high-frequency details based on noise level estimation (Pyatykh, Hesser, & Zheng, 2013). Then, the color image is composed of three channels' outputs.

From each of the 10 scenes shown in Figure 1, one image frame is randomly selected, and the corresponding distorted images with five levels of distortion are restored. Object detection performances on the distorted images and the restored images are compared. Due to space limit, results from only the low level and the medium level distortions are visualized in Figure 6. Results from high, higher, and extreme levels are consistent with the ones of low and medium levels. The results on distorted images are labeled with light color bars with a dashed line, and the ones on restored images are labeled with deep color bars with a solid line. From Figure 6(a), we notice that the restored motion blur images result in better object detection accuracy in most cases. Figure 6(b) describes the detection performance on out-of-focus blur images, and overall the restored images perform better than the distorted ones except in rare cases. The reason for different gains in Figure 6(a) and Figure 6(b) is that the algorithm used performs better with motion blur than out-of-focus blur (Pan, Sun, Pfister, & Yang, 2017). Figure 6(c) presents the detection results on noisy images and their restored counterparts, and the results indicate that restored images for the majority of scenes have higher detection accuracy than the distorted ones. The improvement on low-level distortions is not as much as the ones on medium level distortions. This is because distortion at a low level does not degrade the performance of object detectors significantly. For most cases, the restored images

produce a better detection accuracy performance. The average improvements for restoring the five-level distortions on detection accuracy are 72.26%, 18.93%, and 42.87% for motion blur, out-of-focus blur, and imaging noise, respectively. One reason for different ranges of gain is that the content characteristics of different scenes can also affect the object detection performance.

EVALUATION ON NATURALLY DISTORTED IMAGES

The performance of the proposed framework is also evaluated on naturally distorted images that are not related to our training data set. Two public datasets containing natural images with noise or blur are identified: "Live in the Wild Image Quality Challenge Database" (Ghadiyaram & Bovik, 2015) and "CERT Image Blur Dataset" (Mavridaki & Mezaris, 2014). From these two large public datasets, we have selected images with human objects and with different backgrounds and lighting conditions. As shown in Table 3, we have selected a total number of 68 images for our evaluation, including 10 noisy images, 18 blurry images, and 40 normal images without noise or blur. Snapshots of three representative images are shown in Figure 7.

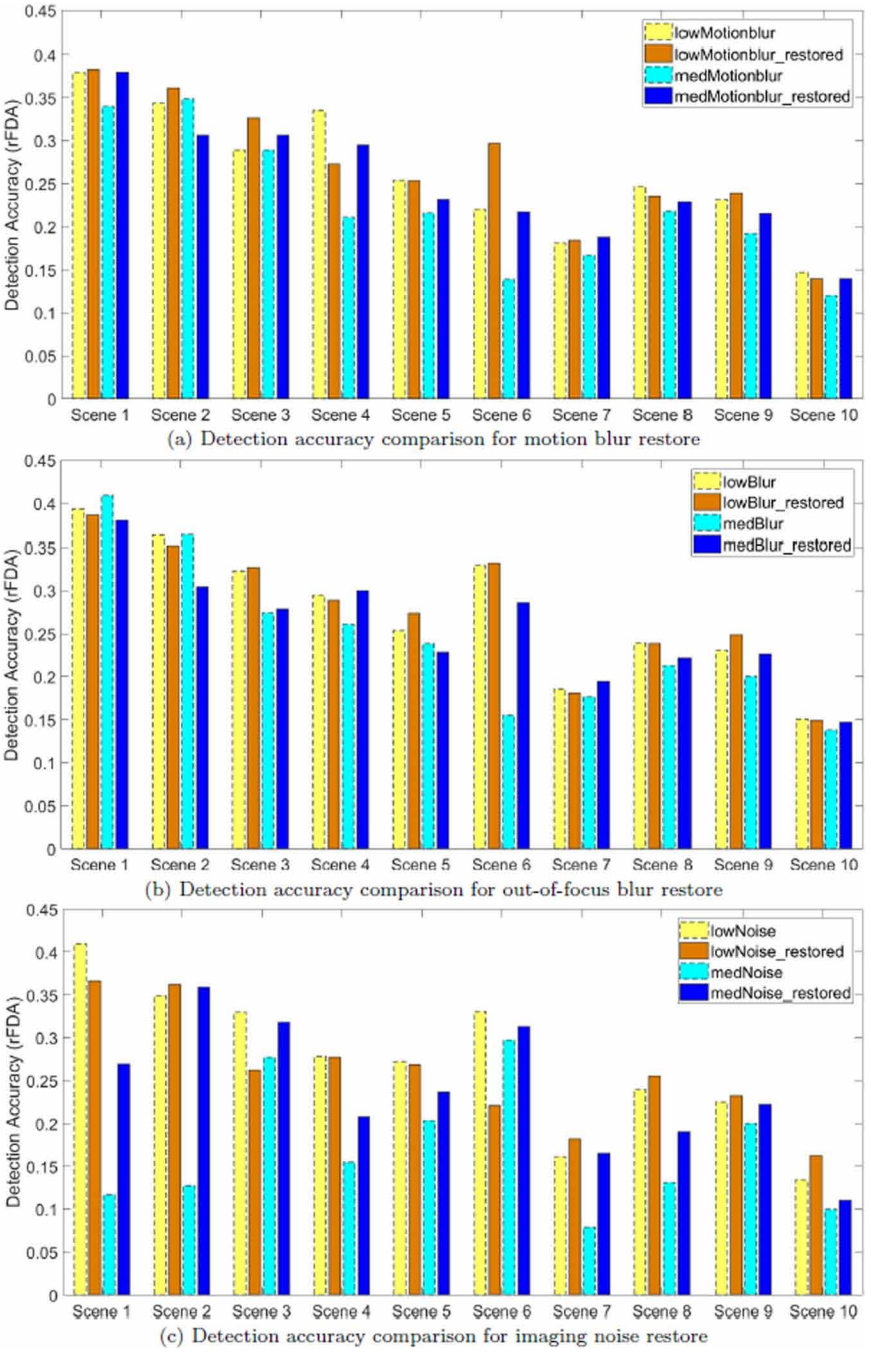
The performance of noise and blur classifiers are tested on the 68 naturally distorted images, and the results are shown in Table 4 and Table 5. Both blur and noise classifiers have an accuracy of 91.2%, while the sensitivity of the noise and blur classifier are 100% and 83.3%, respectively. These results indicate that the noise and blur classifiers could be applied to natural images to predict their distortion types.

We have also tested to what extent these naturally distorted images could be restored. The aforementioned denoise algorithm (Pyatykh, Hesser, & Zheng, 2013) and deblur algorithm (Pan, Sun, Pfister, & Yang, 2017) are applied on the natural images which are classified as noisy or blurry, respectively. After this process, we compare the quality of the restored images with that of the original ones. Unfortunately, the proposed rFDA metric cannot be obtained because the ground truths of these natural images are not available. Alternatively, we test the perceptual quality of the images (using no-reference models BLINDS II and BRISQUE), and we manually check the quality of restored images by visualizing the object detection results. We have observed that 12 out of the 16 noisy images have improved quality after denoising while only 5 out of the 18 blurry images have improved quality after deblurring. We evaluated the perceptual quality of these images based on BLINDS II Algorithm. The overall perceptual quality improved by 80% for the noise corrected images and only by 1.5% for the blur corrected images. The reason for lower performance in blur correction is because of the complexity of blur distortion: it is very difficult to predict the kernel of the blur without a reference image. Object detection results on some representative images are shown in Figures 8-11, where the detection results are described as the number of objects in TP, TN, FN, FP categories. Figure 8 and Figure 10 show examples that restored images result in better detection performance, which includes lower numbers of false positive objects. Figure 9 and Figure 11 show examples that denoise or deblur algorithms fail to improve detection performance, in which cases images should be re-captured if users require better quality.

CONCLUSION

In this paper, we have proposed an image quality adjustment framework to provide satisfactory object detection performance on embedded cameras. The core components of the framework are an image quality model that could predict the performance of object detection, a classifier for detecting out-of-focus and motion blur, and a classifier for detecting imaging noise. Given the predictions of quality and distortion types, restoration or re-capturing steps could be further applied to improve the quality of the image. Both the quality model and the distortion type classifiers are designed based on a data set that includes diverse scene characteristics and commonly used light-weight object detection algorithms. The proposed framework has been evaluated on both images with different

Figure 6. Detection accuracy comparisons for three-distortion restore



levels of simulated distortions and naturally distorted images. Evaluation results have shown that it achieves accurate estimations of both image quality and image distortion types with low computational complexity. The proposed framework could be easily applied to resource-constrained embedded cameras for controlling image quality. is work was supported by the National Institute of Standards and Technology under Grant 60NANB17D193.

Table 3. Breakdown of naturally noisy and blurry data set

Noisy Images (10 images)	**Noisy Images (10 images)**	**Non-Blurry Images (50 images)**
Non-Noisy Images (58 images)	Normal Images (40 images)	
		Blurry Images (18 images)

Figure 7. These are examples of normal (undistorted), noisy and blurry types of images contained in the used data set

Table 4. Confusion matrix of noisy classification

	Predicted Non-Noisy Images	**Predicted Noisy Images**
Actual Non-Noisy	52	6
Actual Noisy	0	10

Table 5. Confusion matrix of blurry classification

	Predicted Non-Noisy Images	**Predicted Noisy Images**
Actual Non-Blurry	47	3
Actual Blurry	3	15

Figure 8. Example of better detection performance after denoising (Before denoising: TP=2, TN=1, FN = 2, FP = 1; After denoising: TP=2, TN=2, FN=2, FP=0).

Figure 9. Example of no improvement in detection performance after denoising (Before denoising: TP=0, TN=0, FN = 1, FP = 1; After denoising: TP=0, TN=0, FN=1, FP=1).

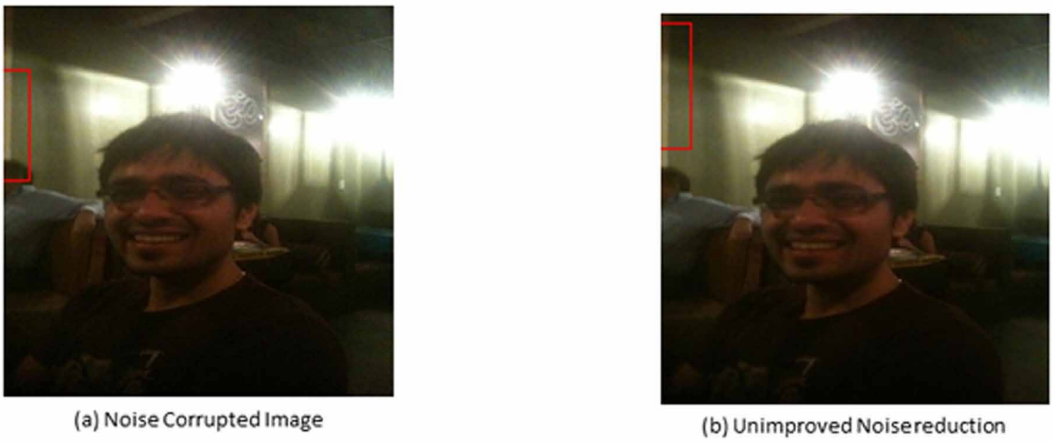


Figure 10. Example of better detection performance after deblurring (Before deblurring: TP=5, TN=0, FN = 0, FP = 6; After deblurring: TP=4, TN=2, FN=1, FP=4).

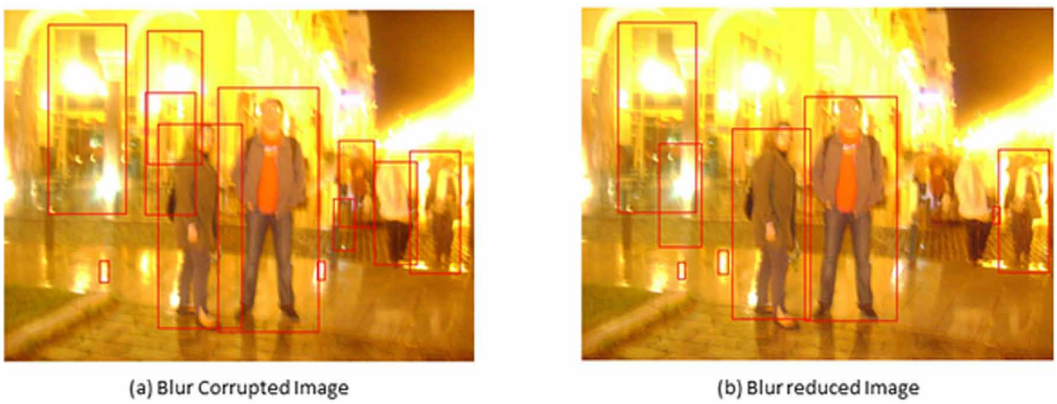


Figure 11. Example of no improvement in detection performance after deblurring (Before deblurring: TP=0, TN=0, FN = 11, FP = 13; After deblurring: TP=0, TN=0, FN=11, FP=22).

(a) Blur Corrupted Image

(b) Unimproved Blur reduction

REFERENCES

Abas, K., Porto, C., & Obraczka, K. (2014). Wireless smart camera networks for the surveillance of public spaces. *Computer*, *47*(5), 37–44. doi:10.1109/MC.2014.140

Bai, Y., Cheung, G., Liu, X., & Gao, W. (2019). Graph-based blind image deblurring from a single photograph. *IEEE Transactions on Image Processing*, *28*(3), 1404–1418. doi:10.1109/TIP.2018.2874290 PMID:30307861

Chen, P., Hong, K., Naikal, N., Sastry, S. S., Tygar, D., Yan, P., Yang, A. Y., Chang, L.-C., Lin, L., Wang, S., Lobatón, E., Oh, S., & Ahammad, P. (2013). A low-bandwidth camera sensor platform with applications in smart camera networks. *ACM Transactions on Sensor Networks*, *9*(2), 21. doi:10.1145/2422966.2422978

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE. doi:10.1109/CVPR.2005.177

Dollár, P., Appel, R., Belongie, S., & Perona, P. (2014). Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(8), 1532–1545. doi:10.1109/TPAMI.2014.2300479 PMID:26353336

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*(2), 303–338. doi:10.1007/s11263-009-0275-4

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1627–1645. doi:10.1109/TPAMI.2009.167 PMID:20634557

Ghadiyaram, D., & Bovik, A. (2015). *LIVE In the Wild Image Quality Challenge Daatabase*. Retrieved from http://live.ece.utexas.edu/research/ChallengeDB/index.html

Gu, K., Tao, D., Qiao, J.-F., & Lin, W. (2017). Learning a no-reference quality assessment model of enhanced images with big data. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(4), 1301–1313. doi:10.1109/TNNLS.2017.2649101 PMID:28287984

Gu, K., Wang, S., Yang, H., Lin, W., Zhai, G., Yang, X., & Zhang, W. (2016). Saliency-guided quality assessment of screen content images. *IEEE Transactions on Multimedia*, *18*(6), 1098–1110. doi:10.1109/TMM.2016.2547343

Gunasekar, S., Ghosh, J., & Bovik, A. C. (2014). Face detection on distorted images augmented by perceptual quality-aware features. *IEEE Transactions on Information Forensics and Security*, *9*(12), 2119–2131. doi:10.1109/TIFS.2014.2360579

Irvine, J. M., & Wood, R. J. (2013). Real-time video image quality estimation supports enhanced tracker performance. In Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications X (Vol. 8713, p. 87130Z). International Society for Optics and Photonics. doi:10.1117/12.2016174

Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Bowers, R., & Zhang, J. et al. (2009). Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(2), 319–336. doi:10.1109/TPAMI.2008.57 PMID:19110496

Kong, L., & Dai, R. (2016). Temporal-fluctuation-reduced video encoding for object detection in wireless surveillance systems. In *2016 IEEE International Symposium on Multimedia (ISM)* (pp. 126-132). IEEE. doi:10.1109/ISM.2016.0032

Kong, L., & Dai, R. (2017). Object-detection-based video compression for wireless surveillance systems. *IEEE MultiMedia*, *24*(2), 76–85. doi:10.1109/MMUL.2017.29

Kong, L., & Dai, R. (2018). Efficient Video Encoding for Automatic Video Analysis in Distributed Wireless Surveillance Systems. *ACM Transactions on Multimedia Computing Communications and Applications*, *14*(3), 1–24. doi:10.1145/3226036

Kong, L., Dai, R., & Zhang, Y. (2016). A new quality model for object detection using compressed videos. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 3797-3801). IEEE. doi:10.1109/ICIP.2016.7533070

Kong, L., Ikusan, A., Dai, R., & Zhu, J. (2019). Blind Image Quality Prediction for Object Detection. *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 216-221. doi:10.1109/MIPR.2019.00046

Kong, L., Ikusan, A., Dai, R., Zhu, J., & Ros, D. (2019). A No-reference Image Quality Model for Object Detection on Embedded Cameras. *International Journal of Multimedia Data Engineering and Management*, *10*(1), 22–39. doi:10.4018/IJMDEM.2019010102

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105). Academic Press.

Leal-Taixé, L., Milan, A., Reid, I., Roth, S., & Schindler, K. (2015). *Motchallenge 2015: Towards a benchmark for multi-target tracking.* arXiv preprint arXiv:1504.01942.

Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z. (2018). Structure-revealing low-light image enhancement via robust Retinex model. *IEEE Transactions on Image Processing*, *27*(6), 2828–2841. doi:10.1109/TIP.2018.2810539 PMID:29570085

Lin, G.-S., Chang, Y.-T., & Lie, W.-N. (2010). A framework of enhancing image steganography with picture quality optimization and anti-steganalysis based on simulated annealing algorithm. *IEEE Transactions on Multimedia*, *12*(5), 345–357. doi:10.1109/TMM.2010.2051243

Mavridaki, E., & Mezaris, V. (2014). No-reference blur assessment in natural images using fourier transform and spatial pyramids. In *2014 IEEE International Conference on Image Processing (ICIP)* (pp. 566-570). IEEE. doi:10.1109/ICIP.2014.7025113

Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). *MOT16: A benchmark for multi-object tracking.* arXiv preprint arXiv:1603.00831.

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, *21*(12), 4695–4708. doi:10.1109/TIP.2012.2214050 PMID:22910118

Nam, W., Dollár, P., & Han, J. H. (2014). Local decorrelation for improved pedestrian detection. *Advances in Neural Information Processing Systems*, *27*, 424–432.

Pan, J., Sun, D., Pfister, H., & Yang, M.-H. (2017). Deblurring images via dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(10), 2315–2328. doi:10.1109/TPAMI.2017.2753804 PMID:28952935

Pang, Z.-F., Zhou, Y.-M., Wu, T., & Li, D.-J. (2019). Image denoising via a new anisotropic total-variation-based model. *Signal Processing Image Communication*, *74*, 140–152. doi:10.1016/j.image.2019.02.003

Pulecio, C. G., Benitez-Restrepo, H. D., & Bovik, A. C. (2017). Image quality assessment to enhance infrared face recognition. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 805-809). IEEE. doi:10.1109/ICIP.2017.8296392

Pyatykh, S., Hesser, J., & Zheng, L. (2013). Image noise level estimation by principal component analysis. *IEEE Transactions on Image Processing*, *22*(2), 687–699. doi:10.1109/TIP.2012.2221728 PMID:23033431

Raghunandan, K., Jalab, H. A., Ibrahim, R. W., Kumar, G. H., Pal, U., & Lu, T. (2017). Riesz fractional based model for enhancing license plate detection and recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(9), 2276–2288. doi:10.1109/TCSVT.2017.2713806

Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016). Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. In *European conference on computer vision* (pp. 17-35). Springer. doi:10.1007/978-3-319-48881-3_2

Saad, M. A., Bovik, A. C., & Charrier, C. (2012). Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Transactions on Image Processing*, *21*(8), 3339–3352. doi:10.1109/TIP.2012.2191563 PMID:22453635

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556.

Suleiman, A., Chen, Y.-H., Emer, J., & Sze, V. (2017). Towards closing the energy gap between HOG and CNN features for embedded vision. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-4). IEEE. doi:10.1109/ISCAS.2017.8050341

Sze, V., Chen, Y.-H., Emer, J., Suleiman, A., & Zhang, Z. (2017). Hardware for machine learning: Challenges and opportunities. In 2017 IEEE Custom Integrated Circuits Conference (CICC) (pp. 1-8). IEEE.

Wang, G., Li, B., Zhang, Y., & Yang, J. (2018). Background modeling and referencing for moving cameras-captured surveillance video coding in HEVC. *IEEE Transactions on Multimedia*, *20*(11), 2921–2934. doi:10.1109/TMM.2018.2829163

Wang, R., & Tao, D. (2014). *Recent progress in image deblurring.* arXiv preprint arXiv:1409.6838.

Wang, S., Gu, K., Ma, S., Lin, W., Liu, X., & Gao, W. (2015). Guided image contrast enhancement based on retrieved images in cloud. *IEEE Transactions on Multimedia*, *18*(2), 219–232. doi:10.1109/TMM.2015.2510326

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. doi:10.1109/TIP.2003.819861 PMID:15376593

Yan, R., & Shao, L. (2016). Blind image blur estimation via deep learning. *IEEE Transactions on Image Processing*, *25*(4), 1910–1921. PMID:26930680

Zhao, W., Lu, H., & Wang, D. (2017). Multisensor image fusion and enhancement in spectral total variation domain. *IEEE Transactions on Multimedia*, *20*(4), 866–879. doi:10.1109/TMM.2017.2760100

Zhu, Y., Samajdar, A., Mattina, M., & Whatmough, P. (2018). Euphrates: Algorithm-SoC Co-Design for Low-Power Mobile Continuous Vision. In *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)* (pp. 547-560). IEEE.

Lingchao Kong is a computer vision algorithm researcher at Zebra Technologies. He received a Ph.D. degree in the Department of Electrical Engineering and Computer Science at University of Cincinnati in 2019. His recent interests include computer vision, deep learning, and multimedia communications.

Ademola Ikusan is a PhD candidate at the University of Cincinnati, his research area is in computer vision, machine learning, networking, and security.

Rui Dai is an Associate Professor in the Department of Electrical Engineering and Computer Science at University of Cincinnati, Ohio, USA. She received her BS in Electronics and Information Engineering and MS in Communications and Information Systems from Huazhong University of Science and Technology, Wuhan, China, in 2004 and 2007, respectively. She received her PhD degree in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta, GA, USA in 2011, under the supervision of Professor Ian F. Akyildiz in the Broadband Wireless Networking Laboratory. She was a postdoctoral fellow at the Center for Assistive Technology and Environmental Access of Georgia Tech in 2012. She was an Assistant Professor at the Department of Computer Science at North Dakota State University from 2012 to 2014. Her recent research interests include multimedia communications and networking, wireless sensor networks, and cyber-physical systems.

Dara Ros received a B.E. degree from the University of Cincinnati, Ohio, USA, in 2016. He is currently working toward his M.Sc. degree at the University of Cincinnati. His research interests include image processing and multimedia processing.