Information-Driven Video Communication for Public Safety Networks

Principal Investigator:

<u>Rui (April) Dai</u>

Assistant Professor

Department of Electrical Engineering and Computer Science

University of Cincinnati

July 9th, 2019

DISCLAIMER

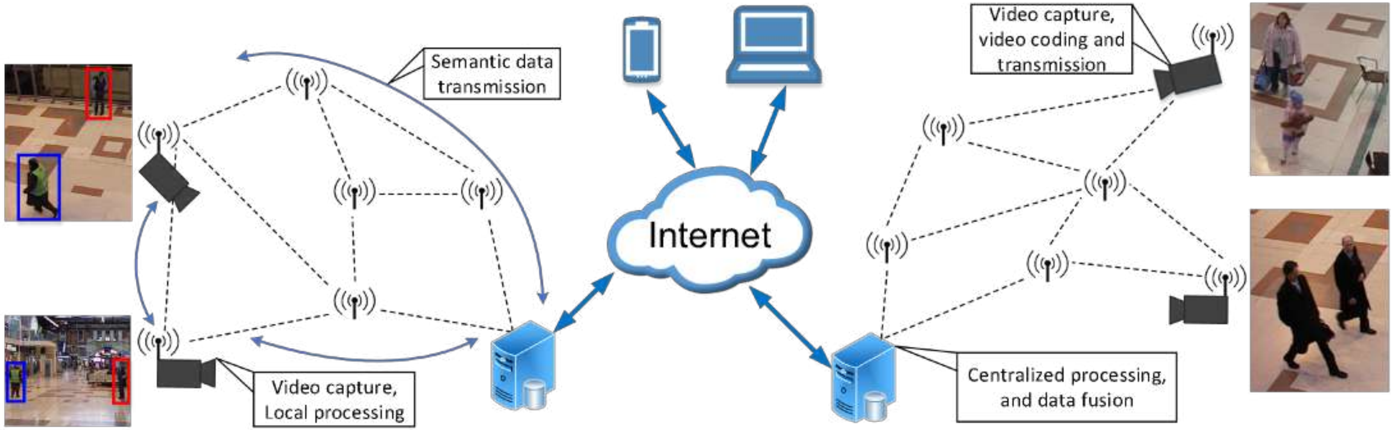
This presentation was produced by guest speaker(s) and presented at the National Institute of Standards and Technology's 2019 Public Safety Broadband Stakeholder Meeting. The contents of this presentation do not necessarily reflect the views or policies of the National Institute of Standards and Technology or the U.S. Government.

Posted with permission

Background

- Networked video surveillance systems have been used heavily in today's public safety infrastructure.
- More and more wireless cameras are deployed.
- Applying automatic video analysis in a distributed wireless camera network can
  - Alleviate the bandwidth pressure;
  - Provide real-time analysis results to enhance the situational awareness of first responders and increase the speed and precision of decision making for various events.

Sensing, Processing, And Communication In Wireless Camera Networks



- The process of video analysis is implemented either in the camera sensor nodes or at the central servers, depending on their computational capability, energy supply, and the purpose of the application.

Terminology

- Quality of Service (QoS)
  - The measurable end-to-end performance properties of a network service, which can be guaranteed in advance by a service level agreement between a user and a service provider.
- Quality of Experience (QoE)
  - The degree of delight or annoyance of human users when they are presented with raw data, such as video or audio streams, which is related to the perceptual quality of multimedia.

- Quality of Information (QoI)
  - Evaluate information that is valuable and actionable to the user.
  - QoI advances QoE by considering the quality of high level information, which could be extracted from raw data by automatic analysis tools.

QoE vs. QoI

- Perceptual image or video quality assessment solutions usually emulate known characteristics of the human visual system (HVS), such as:
  - *Contrast sensitivity*: HVS is sensitive to relative luminance changes rather than absolute luminance changes.
  - *Visual attention:* Only a local area can be perceived with high resolution at one time instance.

- The quality of an image or a video judged by an automatic analysis algorithm, is not necessarily sensitive to the same factors that drive human perceptions.
  - Automatic analysis methods run by machines can “perceive” the absolute luminance changes precisely and have a better global “view”.

Project Overview

- Quantify the QoI for wireless networked surveillance applications, considering common factors for information loss:
  - *Noise and motion blur during video capturing*
  - *Low spatial and temporal resolution videos from resource-constrained embedded cameras*
  - *Lossy compression/encoding*
  - *Packet losses and delays during transmission*
  - *Camera coverage*
  - *Collaboration of different cameras*

- Design QoI-based video encoding, processing, and communication solutions to maximize the information gain.

Research Activities In Year 1

- *<u>Use case</u>:* Videos from camera sensors are compressed and transmitted to a central server for further analysis
- *<u>Research contributions</u>:*

1. Identified factors that could contribute to the quality of automatic object detection in the video compression process

2. Designed a new video encoder to improve the quality of object detection on compressed videos

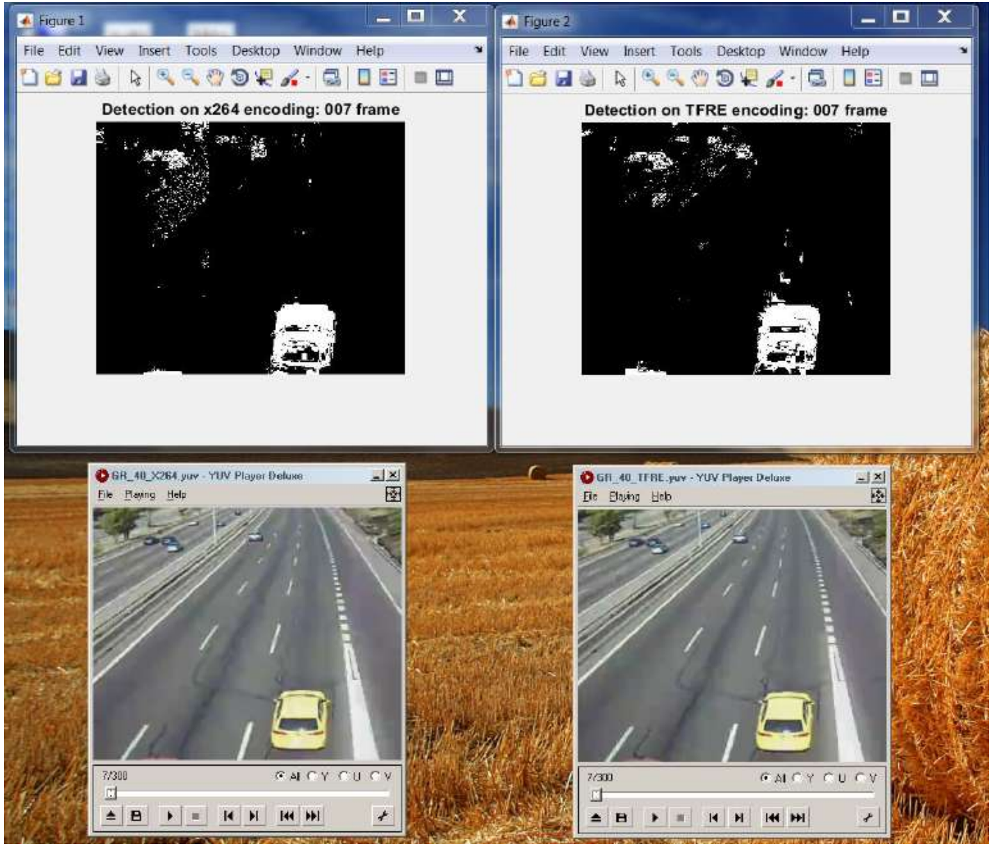
- L. Kong and R. Dai, "Efficient video encoding for automatic video analysis in distributed wireless surveillance systems", ACM Transactions on Multimedia Computing Communications and Applications, vol. 14, issue 3, Aug. 2018.

Object Detection Quality On Compressed Videos

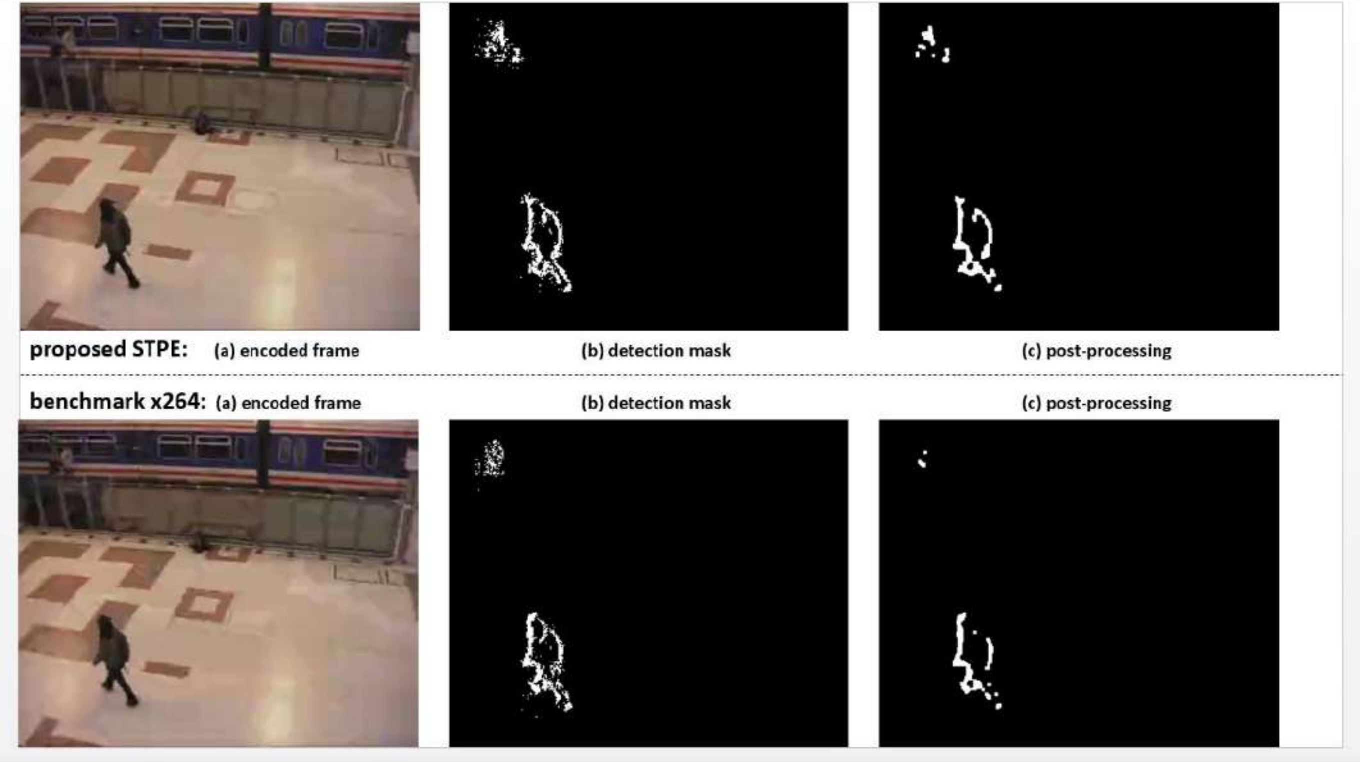
- Our statistical analysis results show that temporal domain fluctuation in stable background areas and spatial texture degradation in dynamic foreground areas could degrade the performance of object detection algorithms.

- We have designed a new video encoding scheme that aims to improve the performance of object detection
  - Temporal-Fluctuation-Reduced video Encoding (TFRE) that suppresses unnecessary temporal fluctuation in stable background areas
  - Spatial-Texture-Preserved video Encoding (STPE) that preserves spatial texture in dynamic foreground areas

Temporal-Fluctuation-Reduced video Encoding (TFRE): Demo



Spatial-Texture-Preserved video Encoding (STPE): Demo



Performance Evaluation

- Encoders for comparison:
  - **x264**: the H.264/AVC based open source encoder (**benchmark**)
  - **RFC**: the Reducing Flicker video Coding approach [Chun et al. 2006] (for improving video quality perceived by human users)
  - **cTwS**: the proposed scheme, combined TFRE and STPE

- Execute three object detection algorithms from different categories:
  - The statistical Gaussian Mixture Model (GMM) [Zivkovic et al. 2006]
  - The non-parametric GMG [Godbehere et al. 2012]
  - The basic Adaptive Background Learning (ABL) [Sobral et al. 2014]

Performance Evaluation: Object Detection

- Evaluate the overall performance of object detection in terms of

Configuration Distance (CD):

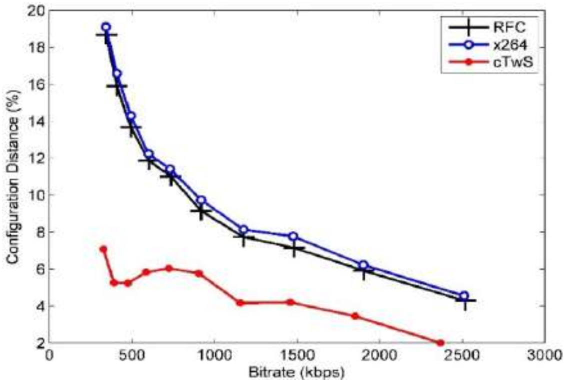
$$CD_f = \frac{|AR_o - GT_o|}{|\max(GT_o, 1)|}$$

where AR_o and GT_o are the numbers of AR (Algorithm Results) objects and GT (Ground Truth) objects in the given frame

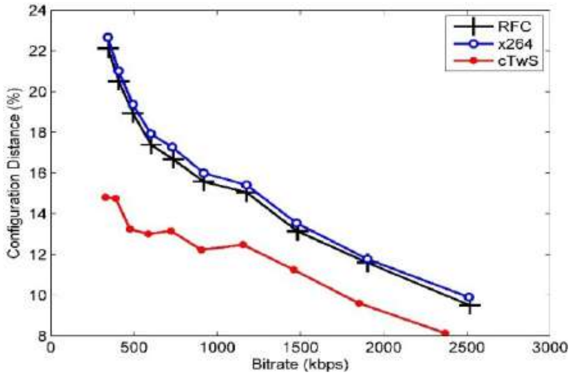
- Algorithm results (AR): object detection accuracy on compressed videos
- Ground truth (GT): object detection accuracy on uncompressed videos

Performance Evaluation: Object Detection

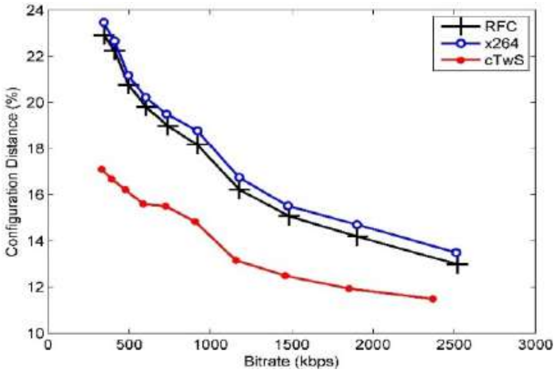
- The overall performance of object detection in terms of Configuration Distance (CD)
- CD vs. bitrate results on *8 videos with 1280*960 resolution*:
- Gain of cTwS over x264: ranging from 2.10% to 8.75%, with an average of 4.82%



CD vs. bitrate for ABL algorithm



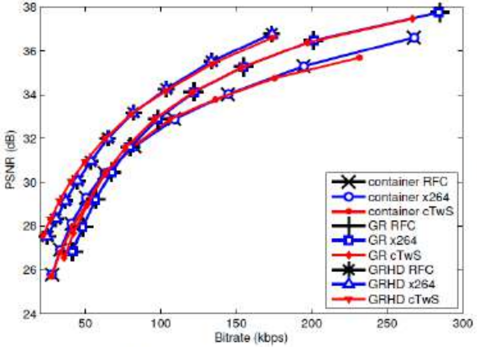
CD vs. bitrate for GMG algorithm



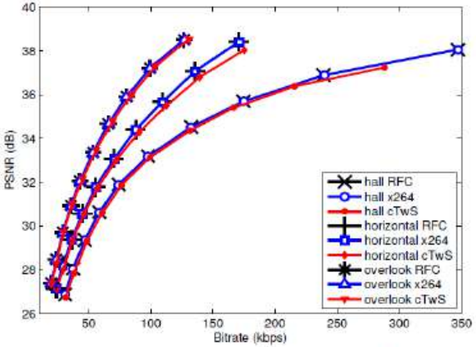
CD vs. bitrate for GMG algorithm

Performance Evaluation: Rate-Distortion

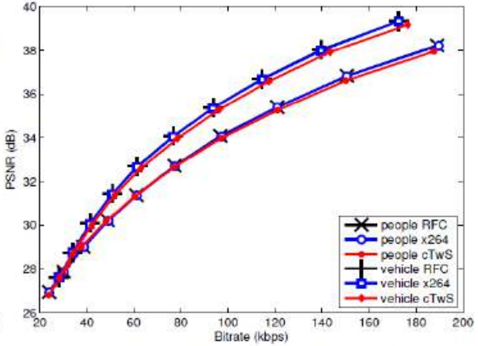
- The R-D performance of the proposed encoding scheme is comparable with that of the benchmarks (results on 8 CIF (352*288) videos)



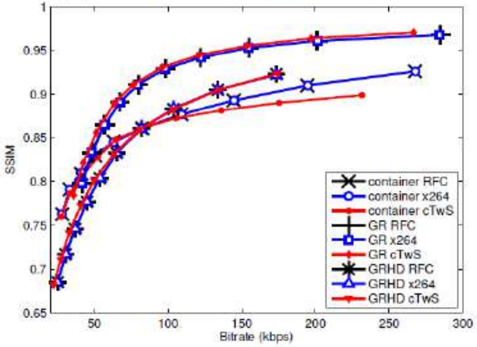
(a) PSNR vs. Bitrate of traffic videos



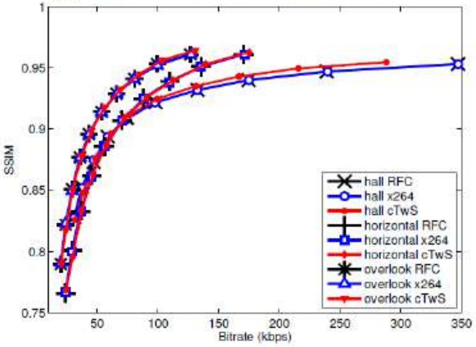
(b) PSNR vs. Bitrate of indoor videos



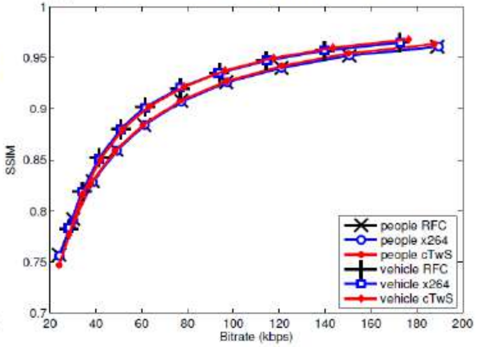
(c) PSNR vs. Bitrate of outdoor videos



(d) SSIM vs. Bitrate of traffic videos



(e) SSIM vs. Bitrate of indoor videos



(f) SSIM vs. Bitrate of outdoor videos

Research Activities In Year 2

- *Use case:* Local processing of raw videos captured by camera sensors
- *Research aims:*

  1. Predict the quality of a captured raw image for object detection algorithms
  2. Design a quality control framework to adjust image quality

- L. Kong, A. Ikusan, R. Dai, and J. Zhu, “Blind Image Quality Prediction for Object Detection”, in IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR), Mar. 2019.
- L. Kong, A. Ikusan, R. Dai, J. Zhu, and D. Ros, “A No-reference Image Quality Model for Object Detection on Embedded Cameras”, International Journal of Multimedia Data Engineering and Management (IJMDEM), vol. 10, issue 1, 2019.
- L. Kong, A. Ikusan, R. Dai, and D. Ros, “An Image Quality Adjustment Framework for Object Detection on Embedded Cameras”, submitted for journal publication, Jun. 2019.

Image Quality Adjustment Framework

- Many existing embedded camera platforms incorporate light-weight detection algorithms on board
  - E.g., some embedded platforms utilize background subtraction and frame differencing for fast object detection
- This work considers the following factors for information loss:
  - *Noise, out-of-focus blur, motion blur*
- Evaluate the performance of object detection algorithms with low or moderate complexity

Data set

- We have selected 10 high resolution original videos with different scene characteristics, illumination levels, and object scales
  - 5 videos from the Multiple Object Tracking (MOT) dataset

    A. Milan, L. Leal-Taix´e, I. Reid, S. Roth, and K. Schindler. Mot16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831, 2016.
  - 5 videos from the Duke Multi-Target Multi-Camera Tracking (DM) dataset

    E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking, 2016.

Data set

- Generate distorted videos based on the original data set
  - Out-of-focus (Gaussian) blur, five levels (low, medium, high, higher, extreme high)
  - Motion blur, five levels
  - Gaussian noise, five levels
  - Reduced resolution, 1:2 and 1:4 down-sampling rates
- Total number of videos: 180
- Total number of images: 133344

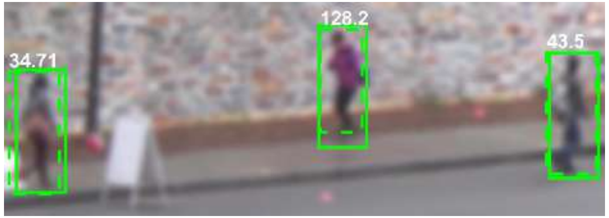
Candidate Object Detection Algorithms

- Two categories of object detection algorithms:
  - *Background modeling based methods*: require multiple frames to build a background
  - *Object modeling based methods*: could generate detection results on a single image

- We use the following three representative low-complexity algorithms based on object modeling:
  - Histograms of Oriented Gradients (HOG)
  - Discriminatively Part Models (DPM)
  - Locally Decorrelated Channel Features (LDCF)

Object Detection Performance Measures

- Frame Detection Accuracy (FDA):



$$FDA = \frac{\sum_{i=1}^{N_m} \frac{G_i \cap D_i}{G_i \cup D_i}}{(N_G + N_D)/2}$$

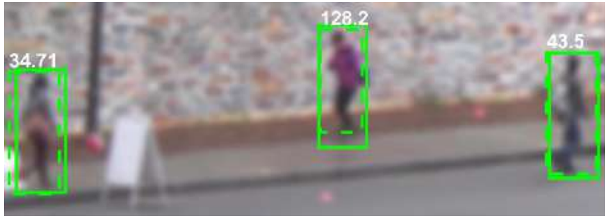
where there are N_G ground-truth objects G and N_D detected objects D, N_m is the number of mapped object pairs.

- For a given frame, the optimal matching pairs are assigned by computing the spatial overlap between ground truth and detected objects.
- Then, calculate the spatial overlap between the ground truth and system output objects as a ratio of the spatial intersection between the two objects and the spatial union of them.
- The sum of all of the overlaps is normalized over the average of the number of ground truth and detected objects.

Object Detection Performance Measures

- We introduce a revised FDA measure, rFDA, which is the average of FDA based on different confidence levels

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



N_m is the maximum number of mapped object pairs under different confidence levels.

$N_{T(j)}$ is the number of true positives when the threshold of detection confidence is T(j).

$N_{T(j)}$ could take values ranging from 1 to N_m.

Image Quality Adjustment Framework

Key components:

1. Quality prediction for object detection
2. Noise classifier
3. Blur classifier

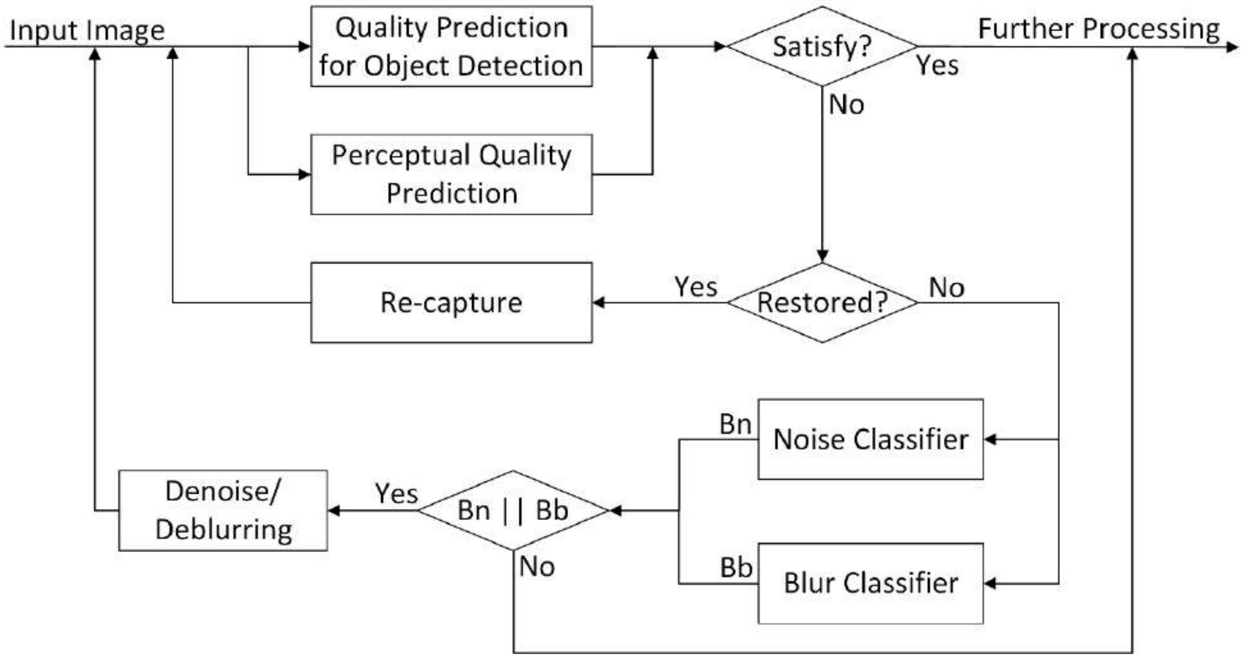
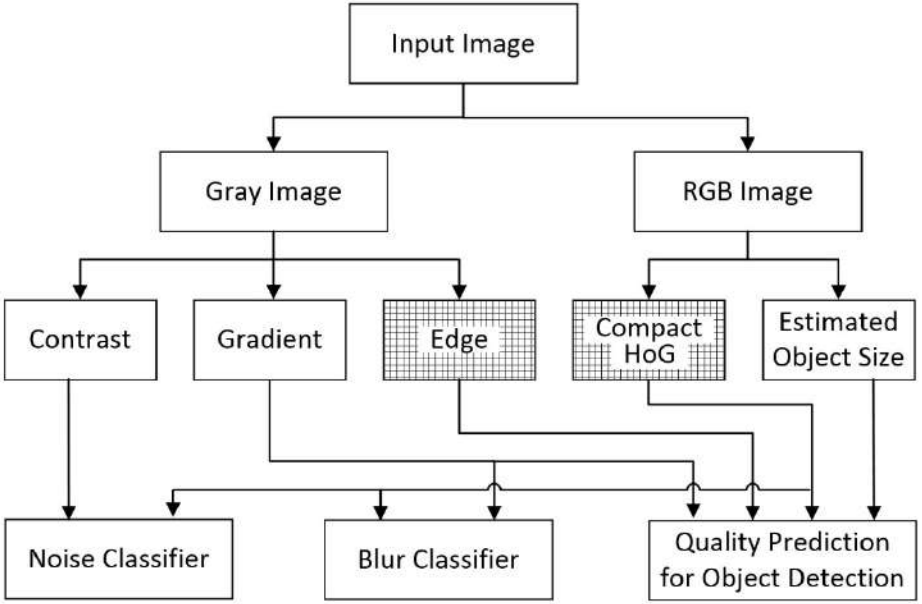


Image Quality Adjustment Framework

- Utilize machine learning to build the key components.
- Introduce 18 local and global features, all of which could be obtained from an image with low computational complexity.



Blind Model for Predicting Object Detection Quality

- We build a regression model for predicting rFDA using 4 categories of features
  - gradient, compact HoG (histogram of gradient), edge, and estimated object size

- Gradient magnitude and direction of an image: 4 features

  (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

Blind Model for Predicting Object Detection Quality

- Compact Histograms of Oriented Gradients (HOG):
  - The local window for one HOG descriptor is set as 16*16 pixels, and the average frequency w_m and the frequency's variation level w_s of the histogram's bins, are computed for each window.

- 4 features in this category:

(5) hog_mm: the average of every blocks' w_m

(6) hog_ms: the standard deviation of every blocks' w_m

(7) hog_sm: the average of every blocks' w_s

(8) hog_ss: the standard deviation of every blocks' w_s

(a) (b) (c) (d)

Blind Model for Predicting Object Detection Quality

- Boundary/edge features obtained using Sobel operator: 4 features

  (9) edge_mm: the average of every blocks' average

  (10) edge_ms: the standard deviation of every blocks' average

  (11) edge_sm: the average of every blocks' standard deviation

  (12) edge_ss: the standard deviation of every blocks' standard deviation

−1	0	1
−2	0	2
−1	0	1

(a) Horizontal mask

1	2	1
0	0	0
−1	−2	−1

(b) Vertical mask

Sobel masks

Blind Model for Predicting Object Detection Quality

- If the size of an object is too small or too large in the image, it is hard to detect the object from the background.

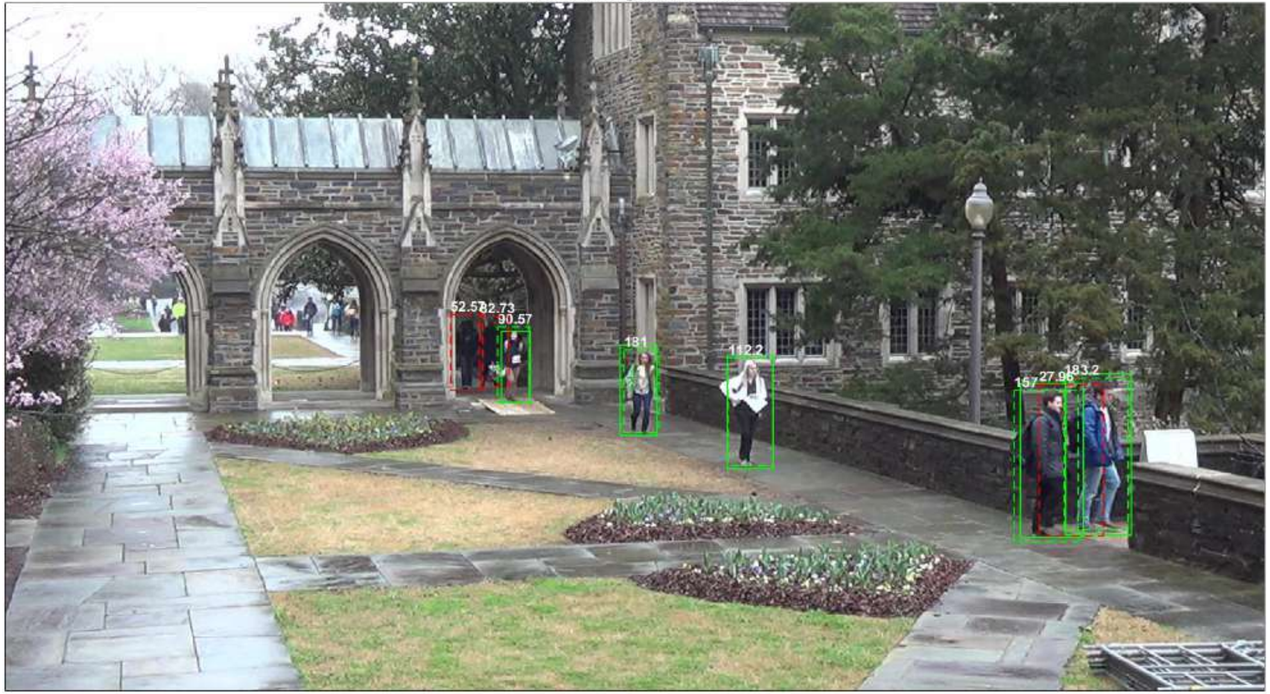
  (13) estimated object size (Obtained based on Otsu's method with low complexity)

- We use the bootstrap aggregating, or bagging, ensemble of trees to train a regression model to predict detection performance based on the aforementioned 13 features.

Predicting Object Detection Quality: Examples

- Object detection on an original high-resolution image

DMcam06-original-LDCF-fr24

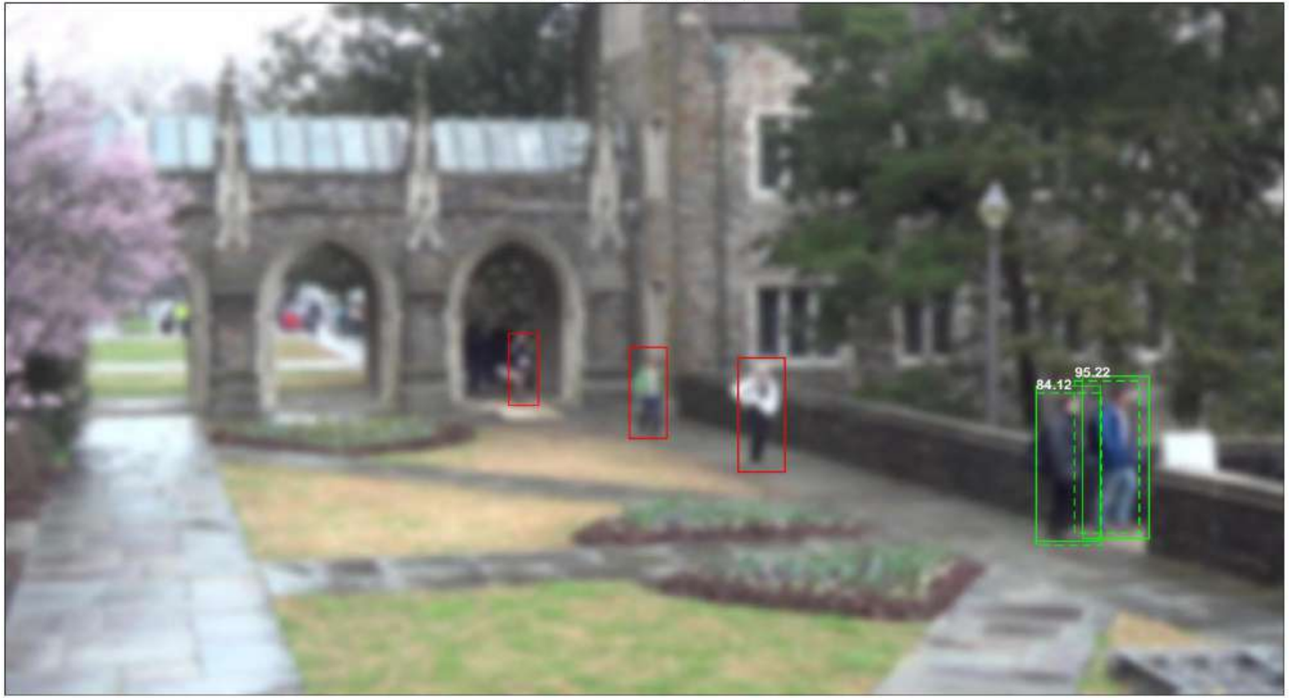


rFDA: 0.3633   predicted rFDA: 0.3198

Predicting Object Detection Quality: Examples

➢ Object detection on a blurred image (reduced accuracy indicated by smaller rFDA)

DMcam06-highBlur-LDCF-fr24

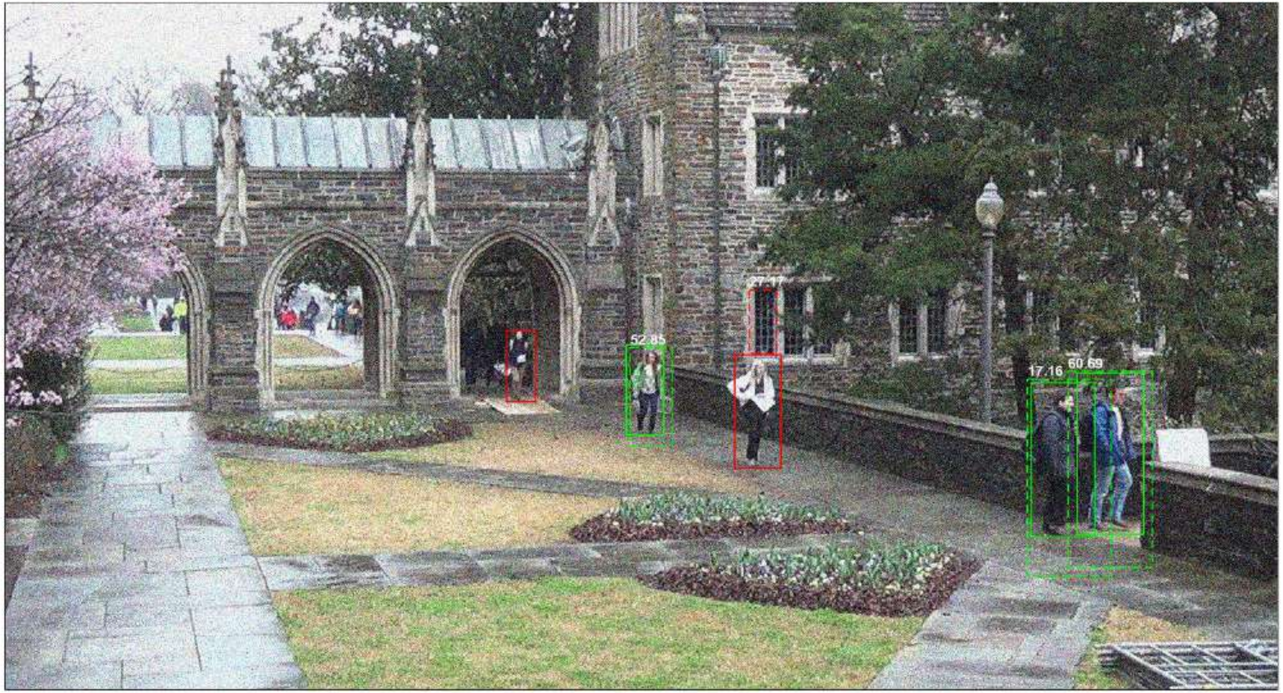


rFDA: 0.1721   predicted rFDA: 0.1474

Predicting Object Detection Quality: Examples

➢ Object detection on a noisy image (reduced accuracy indicated by smaller rFDA)

DMcam06-highNoise-LDCF-fr24



rFDA: 0.1981 predicted rFDA: 0.1090

Blur and Noise Classifiers

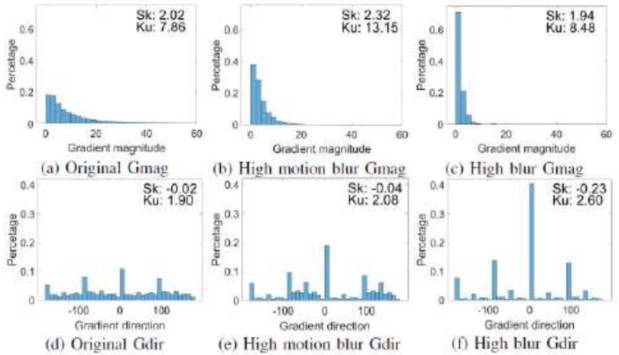
- More features for blur classification: Skewness and Kurtosis of gradient magnitude and direction

  (1) skewGmag: the skewness of gradient magnitude

  (2) kurtGmag: the kurtosis of gradient magnitude

  (3) skewGdir: the skewness of gradient direction

  (4) kurtGdir: the kurtosis of gradient direction



(a) Original Gmag (b) High motion blur Gmag (c) High blur Gmag
(d) Original Gdir (e) High motion blur Gdir (f) High blur Gdir

Skewness: describes lack of symmetry in a distribution

Kurtosis: indicates a distribution has heavy/light tails

Blur and Noise Classifiers

- One more feature for noise classification:

  (18) Image contrast

  The randomness of noise can cause arbitrary changes of local intensities, which bring more inconsistency of intensities compared with normal images.

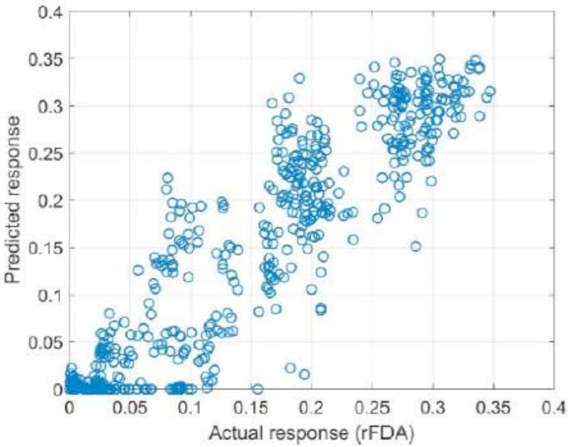
- Details of the classifiers:
  - The blur classifier is based on features (1)-(4), (5)-(8), and (14)-(17).
  - The noise classifier is based on features (5)-(8) and (18).
  - Both classifiers are based on support vector machines (SVM).

Performance Evaluation

- Settings for the regression model and the classifiers:

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%

- Evaluation of quality prediction accuracy:



Metrics	RMSE	R^2	$adj\, R^2$	MSE	MAE
Overall performance	0.0461	0.8416	0.8416	0.0021	0.0347

- Comparison with commonly-used quality metrics PSNR and SSIM:

Algorithms	LCC	KRCC	SROCC
SSIM	0.6187	0.5198	0.7068
PSNR	0.7792	0.6208	0.8165
Proposed	0.9205	0.7319	0.9049

Performance Evaluation

- Comparison with two no-reference perceptual quality models: BRISQUE and BLINDS-II
  - Accuracy:

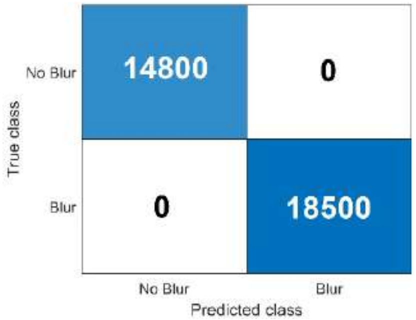
Algorithms	LCC	KRCC	SROCC
BRISQUE	0.4371	0.3737	0.5342
BLIINDS-II	0.5392	0.4337	0.6007
Proposed	0.9184	0.7204	0.9002

  - Computational complexity measured in seconds:

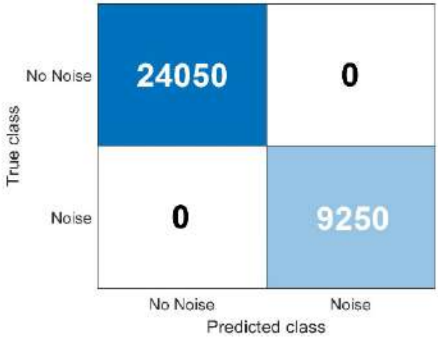
Algorithms	1920×1080	960×540	480×270
BLIINDS-II	158.987 (±1.491)	40.065 (±0.698)	10.140 (±0.202)
BRISQUE	0.785 (±0.070)	0.324 (±0.061)	0.235 (±0.053)
Proposed	0.425 (±0.044)	0.231 (±0.017)	0.143 (±0.014)

Performance Evaluation

- Results for blur classification:

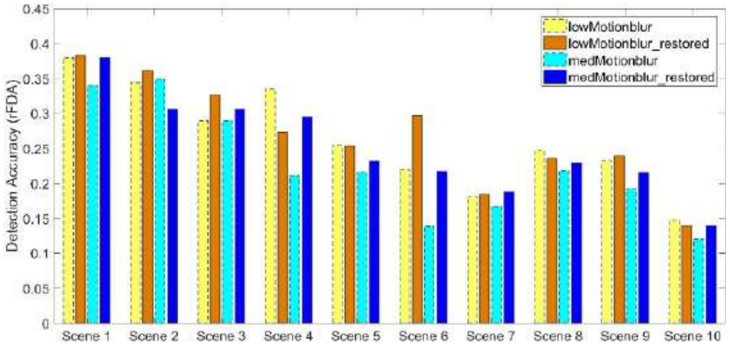


- Results for noise classification:

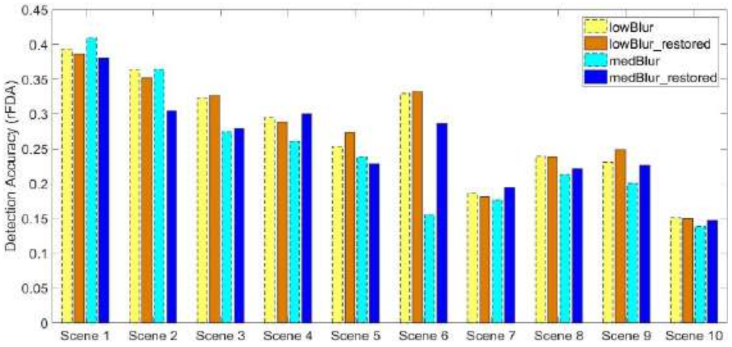


Performance Evaluation

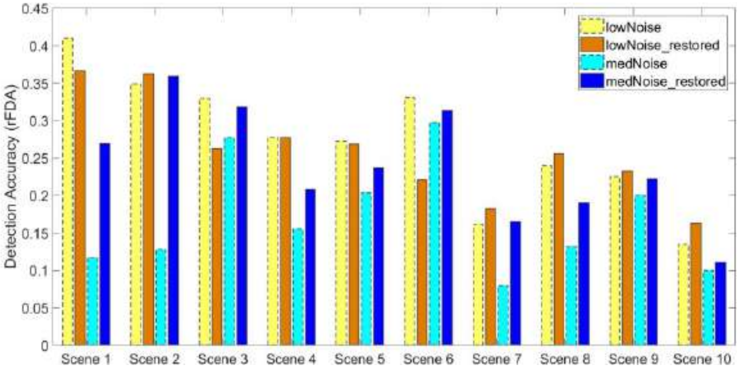
➢ Object detection accuracy before and after image quality adjustment:



Motion blur scenes



Out-of-focus blur scenes



Noisy scenes

Average percentages of detection accuracy improvement for motion blur, out-of-focus blur, and noisy images with all the five levels of distortion:
72.26%, 18.93%, and 42.87%

Conclusion and Future Work

- Summary of results in Year 1
  - Proposed new measures for object detection quality on compressed videos.
  - Designed a novel video encoding scheme that could improve object detection performance on compressed videos.

- Summary of results in Year 2
  - Investigated object detection quality for local processing on embedded cameras.
  - Designed an image quality adjustment framework that includes a quality prediction model and two distortion type classifiers.

Conclusion and Future Work

- Future work
  - Study automatic video analysis quality for applications based on multiple cameras.
  - Design QoI-based collaborative processing and communication solutions for wireless camera networks.

- Acknowledgement
  - This work has been supported by the National Institute of Standards and Technology under Grant 60NANB17D193.

#PSCR2019

Get your hands on the tech!

Demos Open

BACK TOMORROW

8:00 AM