DEEP LEARNING BASED TARGET TRACKING AND CLASSIFICATION DIRECTLY IN COMPRESSIVE MEASUREMENT FOR LOW QUALITY VIDEOS

Chiman Kwan\textsuperscript{1}, Bryan Chou\textsuperscript{1}, Jonathan Yang\textsuperscript{2}, and Trac Tran\textsuperscript{3}

\textsuperscript{1}Applied Research LLC, Rockville, Maryland, USA
\textsuperscript{2}Google, Inc., Mountain View, California, USA
\textsuperscript{3}Johns Hopkins University, Baltimore, Maryland, USA

ABSTRACT

Past research has found that compressive measurements save data storage and bandwidth usage. However, it is also observed that compressive measurements are difficult to be used directly for target tracking and classification without pixel reconstruction. This is because the Gaussian random matrix destroys the target location information in the original video frames. This paper summarizes our research effort on target tracking and classification directly in the compressive measurement domain. We focus on one type of compressive measurement using pixel subsampling. That is, the compressive measurements are obtained by randomly subsample the original pixels in video frames. Even in such special setting, conventional trackers still do not work well. We propose a deep learning approach that integrates YOLO (You Only Look Once) and ResNet (residual network) for target tracking and classification in low quality videos. YOLO is for multiple target detection and ResNet is for target classification. Extensive experiments using optical and mid-wave infrared (MWIR) videos in the SENSIAC database demonstrated the efficacy of the proposed approach.

KEYWORDS

Compressive measurements, target tracking, target classification, deep learning, YOLO, ResNet, optical videos, infrared videos, SENSIAC database

1. INTRODUCTION

Optical and infrared videos have been widely used for traffic monitoring, surveillance, and security monitoring applications [1]-[5]. Compared to radar based trackers [6][7], object features can be clearly seen in optical or infrared videos.

Compressive measurements [8]-[12] are normally collected by multiplying the original vectorized image with a Gaussian random matrix. Each measurement is a scalar and the measurement is repeated $M$ times where $M$ is much fewer than $N$ (the number of pixels). Using compressive measurements for target tracking and classification, the image frames need to be reconstructed and then conventional trackers are then applied. There are two drawbacks in this conventional approach. First, the reconstruction process using $L_0$ [13] or $L_1$ [14]-[16] based methods is time consuming. Second, there may be information loss in the reconstruction process.

In the SENSIAC data [9], there are optical and mid-wave infrared (MWIR) videos containing multiple vehicles collected at ranges between 1000 m and 5000 m. The data are challenging because there are strong illumination variations and other environmental factors. Moreover, the target sizes are small and the video quality (resolution) is low. One active research area in target tracking is the use of compressive measurement directly without reconstruction because...
reconstruction requires a lot of time and thereby limits real-time applications. Some initial results have been presented in [10]-[12].

In the literature, there are some trackers such as [17] that used the term compressive tracking. However, those trackers are not using compressive measurements directly. Conventional approaches usually need to reconstruct the video frames and no one, except our team, has developed any high performance algorithms to deal with target tracking and classification directly in the compressive measurement domain, which has two key advantages. First, using compressive measurement directly enables faster processing and real-time tracking is then feasible. Second, our study showed that there will be no loss of information due to reconstruction if one uses compressive measurement directly [18]-[21].

Recently, we developed a residual network (ResNet) based tracking and classification framework using compressive measurements [12]. Pixel subsampling was used to obtain the compressive measurements. ResNet was used in both target detection and classification. The tracking is done by detection. Although the performance in [12] is much better than conventional trackers, there is still room for further improvement. The key area is to improve the tracking part, which has a significant impact on the classification performance. That is, if the target area is not correctly located, the classification performance will degrade.

In this paper, we propose an alternative approach, which aims to improve the both tracking performance and classification using compressive measurements. The idea is to deploy a high performance detector known as YOLO [22] for target tracking. YOLO is fast, accurate, and has comparable performance as other trackers such as Faster R-CNN [23]. The improved tracking results will further improve the classification performance. The classification is still using ResNet [24] because ResNet has better classification than the built-in classifier in YOLO [10][11][25]. Experiments using optical and MWIR videos in the SENSIAC database clearly demonstrated the performance of the proposed approach.

We would like to briefly review some state-of-the-art algorithms that performs action inference or object classification directly using compressive measurements. We will highlight the key differences between our approach and others.

The paper in [28] presents a reconstruction-free approach to action inference. Smashed filters are built using training samples that are affine transformed to a canonical viewpoint. It works very well even for 100 to 1 compression. However, the approach is for action inference, not for target detection, tracking, and classification in compressed measurement domain. Moreover, the smashed filter assumes that the camera is stationary and the angle is fixed. Extending the approach to target tracking and classification with moving cameras may be non-trivial.

In [29], a convolutional neural network (CNN) approach was presented to perform image classification directly in compressed measurement domain. The input image is assumed to be cropped and centered, and there is only one target in each image.

Papers [30,31] are similar in spirit to [29]. Both papers discussed direct object classification using compressed measurement. However, both papers assumed that the targets/objects are already centered. The approach in [32] is not reconstruction free. The integral image is one type of reconstructed image. In contrast, our paper does not require any image reconstruction.

Reference [33] used a random mask to conceal the actual contents of the original video. In addition, the key idea in [33] is about action recognition (similar to [28]), not object tracking and classification. Extending the idea in [33] to object tracking and classification may not be an easy task.
Reference [34] presents an object detection approach using correlation filters and sparse representation. There was no object classification. No reconstruction of compressive measurements is needed. The results are quite good. Different from [34], our paper focuses on object detection, tracking, and classification.

In [35], the authors present an approach to extracting features out of the compressed measurements and then uses the features to create a proxy image. This approach may not be considered as a reconstruction free approach. Similar to Refs. [29]-[31], the approach is suitable for stationary camera cases and also the objects are already centered in the images.

The paper in [36] presents an online reconstruction free approach to object classification using compressed measurements. Similar to [29]-[31] and [35], the approach assumes the object is already at the center of the image. The methods in [29]-[31], [35] and [36] also did not address the above mentioned issue.

Instead of using Gaussian random measurements to obtain the compressive measurements, we emphasize that we have proposed two alternative compressive measurements. One is called subsampling and the other is called coded aperture. The coded aperture case has been summarized and reported in our recent papers [37-38].

An earlier version of this paper was presented in an SPIE conference [25]. One major change is that significant amount of new experiments were performed. In our earlier paper, we had only the tracking and classification results for the optical videos. Here, we have included tracking and classification results for MWIR daytime and night-time videos. The earlier version had 11 pages and the current version has more than 20 pages.

This paper is organized as follows. Section 2 describes the compressive sensing via subsampling, YOLO tracker, ResNet, SENSIAC videos, and performance metrics. Section 3 presents the tracking and classification results directly in the compressive measurement domain using SENSIAC optical videos. Section 4 includes the tracking and classification results for the MWIR daytime videos. Section 5 then presents the tracking and classification results for the MWIR night-time videos. Finally, some concluding remarks and future research directions are included in Section 6.

2. BACKGROUND

2.1. Compressive Sensing Via Subsampling

Using Gaussian random measurement makes the target tracking very difficult. This is because the targets can be anywhere in a frame and the target location information is lost in the compressive measurements. Recently, we proposed an approach [12] using a random subsampling operator to perform compressive sensing. Figure 1 displays two examples of a random subsampling sensing matrices.

![Figure 1. (a) Visualization of the sensing matrix for a random subsampling operator with a compression factor of 2. The subsampling operator is applied to a vectorized image. This is equivalent to applying a random mask shown in (b) to an image.](image-url)
2.2. YOLO

YOLO tracker [22] has similar performance as Faster R-CNN [23]. YOLO also comes with a classification module. However, based on our evaluations, the classification accuracy using YOLO was not good [10][11][25]. This is perhaps due to a lack of training data.

The training of YOLO is quite simple. Images with ground truth target locations are needed. The input image is resized to 448x448. Figure 2 shows the architecture of YOLO version 1. There are 24 convolutional layers and 2 fully connected layers. The output is 7x7x30. We have used YOLOv2 because it is more accurate than YOLO version 1. The bounding box for each vehicle was manually determined using tools in MATLAB. For YOLO, the last layer of the deep learning model was re-trained. We did not change any of the activation functions. YOLO took approximately 2000 epochs to train.

![Figure 2. 24 convolutional layers followed by 2 fully connected layers for YOLO version 1 [22].](image)

2.3. ResNet

The ResNet-18 model is an 18-layer convolutional neural network (CNN) that avoids performance saturation and/or degradation when training deeper layers. Figure 3 shows the architecture of an 18-layer ResNet.

Training of ResNet requires target patches. The targets are cropped from training videos. Mirror images are then created. Data augmentation using scaling (larger and smaller), rotation (every 45 degrees), and illumination (brighter and dimmer) is usually deployed to create more training data. For each cropped target, we are able to create a data set with 64 more images.

![Figure 3. Architecture of ResNet-18. Figure from [24].](image)
2.4. SENSIAC Optical and MWIR Videos

Our research objective is to perform tracking and classification of seven vehicles using the SENSIAC videos in compressive measurement domain. Both optical and mid-wave infrared (MWIR) videos were collected at distances from 1000 m to 5000 m with 500 m increments. Optical videos have only daytime videos. In contrast, there are both daytime and night-time MWIR videos. There are seven types of vehicles, which are shown in Figure 4. These videos are challenging for several reasons. First, the target sizes are small due to long distances. Second, the target orientations also change drastically. Third, the illuminations in different videos are also different. Here, the compressive measurements are collected via directly sub-sampling. That is, 50% or 75% of the pixels are thrown away during the data collection process.

2.5. Performance Metrics for Tracking and Classification

In our earlier paper [25], we have included some tracking results where conventional trackers such as GMM [26] and STAPLE [27] were used. The tracking performance was poor in the presence of missing data. We experimented with a YOLO tracker, which has been determined to perform better tracking than our earlier ResNet based tracker described in [12]. We used the following metrics [9]-[12] for evaluating the tracker performance:

- Center Location Error (CLE): It is the error between the center of the bounding box and the ground-truth bounding box.
- Distance Precision (DP): It is the percentage of frames where the centroids of detected bounding boxes are within 20 pixels of the centroid of ground-truth bounding boxes.
- EinGT: It is the percentage of the frames where the centroids of the detected bounding boxes are inside the ground-truth bounding boxes.
- Number of frames with detection: This is the total number of frames that have detection.

For classification, we used confusion matrix and classification accuracy as performance metrics.

3. Tracking and Classification Results Using SENSIAC Optical Videos

Here, we focus on the optical videos.

3.1. Conventional Trackers Performance

Figure 5 shows the tracking performance of a conventional tracker known as GMM [26] for three compressive measurement cases: 0%, 50%, and 75%. It can be seen that the tracking is only effective for the 0% missing case. For the other two cases, the green bounding boxes are mostly off the target. Figure 6 shows the tracking results using another conventional tracker known as...
STAPLE [27]. STAPLE appears to yield reasonable tracking results up to 50% missing. However, for 75% missing case, the tracker loses all the targets. This is because STAPLE was not designed to handle videos with missing data.

![Tracking using GMM for 3 missing cases of an optical video containing a truck from 1000 m range.](image1)

![Tracking using STAPLE for 3 missing cases of an optical video containing a truck from 1000 m range.](image2)

3.2. Train using 1500 m and 3000 m videos and Test using 1000 m, 2000 m, 2500 m, 3500 m videos

Here, we applied YOLO and used videos from ranges of 1500 m and 3000 m for training and videos from ranges of 1000 m, 2000 m, 2500 m, and 3500 m for testing. Table 1 - Table 3 summarize the performance metrics for 0%, 50%, and 75% missing cases, respectively. Our first observation is that the number of frames with detection decreases when we have more missing pixels. This is reasonable. For the same missing rate case, the tracking performance drops with increasing range, which is also reasonable. For 75% missing case, the tracking performance is only effective up to 1500 m. Figure 7 -
Figure 9 shows the tracking results in some selected frames for 0%, 50%, and 75% missing cases, respectively. It can be seen that there are more missed detections in those cases of high missing rates.

**Table 1. Tracking metrics for 0% missing case.**

<table>
<thead>
<tr>
<th>Model</th>
<th>1000m</th>
<th>2000m</th>
<th>2500m</th>
<th>3000m</th>
<th>3500m</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>BTRM2</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>SUV</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>T72</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Truck</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

**Table 2. Tracking metrics for 50% missing case.**

<table>
<thead>
<tr>
<th>Model</th>
<th>1000m</th>
<th>2000m</th>
<th>2500m</th>
<th>3000m</th>
<th>3500m</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>BTRM2</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>SUV</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>T72</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Truck</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Figure 7. Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 0% missing case. Optical – Subsampling Mode. Tracking at 0%; vehicle in the frames is SUV.

**Table 2. Tracking metrics for 50% missing case.**

<table>
<thead>
<tr>
<th>Model</th>
<th>1000m</th>
<th>2000m</th>
<th>2500m</th>
<th>3000m</th>
<th>3500m</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>BTRM2</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>SUV</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>T72</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Truck</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

**Figure 7.** Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 0% missing case. Optical – Subsampling Mode. Tracking at 0%; vehicle in the frames is SUV.
Figure 8. Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 50% missing case. Vehicle in the frames is SUV.

Table 3. Tracking metrics for 75% missing case.

<table>
<thead>
<tr>
<th>1000m</th>
<th>2000m</th>
<th>3500m</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>0.99</td>
<td>0.85</td>
</tr>
<tr>
<td>BRDM2</td>
<td>0.95</td>
<td>0.80</td>
</tr>
<tr>
<td>BTR70</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>SUV</td>
<td>0.85</td>
<td>0.70</td>
</tr>
<tr>
<td>T72</td>
<td>0.80</td>
<td>0.65</td>
</tr>
<tr>
<td>Truck</td>
<td>0.75</td>
<td>0.60</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0.70</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Figure 9. Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 75% missing case. Vehicle in the frames is SUV.
3.3. Classification Results

For vehicle classification, we deployed ResNet. For the ResNet classifier, we performed customized training where the training data are augmented with rotation, scaling, and illumination variations. We used videos from ranges 1500 m and 3000 m for training and videos from other ranges for testing.

Classification is only applied to frames with detection of targets from the tracker. Table 4-Table 6 summarize the ResNet classifiers for 0%, 50%, and 75% missing cases, respectively. The highlighted yellow cells indicate the biggest number of each row. The column labels in the confusion matrices are the ground truth and the row labels are the classifier output labels. For 0% missing, the classification performance is good up to 3000 m. For 50% missing case, the performance is still reasonable up to 2000 m. However, for 75% missing case, the classification is only good up to 1500 m.

Table 4. Classification results for 0% missing case. Left columns are the confusion matrix; the last column is the classification results.

<table>
<thead>
<tr>
<th></th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>SUV</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>654</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>35%</td>
</tr>
<tr>
<td>BRDM2</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98%</td>
</tr>
<tr>
<td>BTR70</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>SUV</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>T72</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Truck</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5. Classification results for 50% missing case. Left columns are the confusion matrix; the last column is the classification results.

<table>
<thead>
<tr>
<th></th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>SUV</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>BRDM2</td>
<td>2</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>BTR70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>SUV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>T72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Truck</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 6. Classification results for 75% missing case. Left columns are the confusion matrix; the last column is the classification results.

<table>
<thead>
<tr>
<th></th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>SUV</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>BRDM2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>BTR70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>SUV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>T72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Truck</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 6. Classification results for 75% missing case. Left columns are the confusion matrix; the last column is the classification results.

3.4. Discussions: Tracking and Classification at 75% missing rate: Optical Videos

Table 7 summarizes the tracking (YOLO) and classification (ResNet) results for 75% missing data case. By looking at Table 7, we have the following observations:

- If we have trained models for some specific ranges, then the tracking accuracy is good. For example, we have trained models at 1500 m and 3000 m and the tracking results are good at 1500 m, but tracking results deteriorate at 3000 m. This means that, in theory, tracking can perform well if we have enough training data at various ranges.
- The classification is only effective up to 1500 m. This is because the target size is so small at farther ranges. The target information is further reduced due to missing pixels.

Table 7. Summary of Tracking and Classification Results at 75% missing rate.

4. Tracking and Classification Results Using SENSIAC MWIR Day Time Videos

There are two types of MWIR videos: daytime and night-time. Here, we focus on daytime MWIR videos.

4.1. Train using 1500 m and 3000 m videos and Test using 1000 m, 2000 m, 2500 m, 3500 m videos

For the YOLO models, we used 1500 m and 3000 m videos to train two separate models. The 1500m model will be used for 1000 m and 2000 m ranges and the 3000 m model will be for ranges of 2500 m and 3500 m. Table 8-Table 10 show the metrics. We can see that the numbers
of frames with detection are high for low missing rates. For frames with detection, the CLE values generally increase whereas the DP and EinGT values are relatively stable.

Figure 10-Figure 12 show the tracking results visually. The tracking results are good up to 2000 m at 0% missing. For longer ranges such as 2500 m and beyond, the detection is not good. For example, at 50% missing rate, the tracking is only good up to 1000 m. For 75% missing, the tracking is poor even at 1000 m range. This means MWIR is not suitable for daytime surveillance.

Table 8. Tracking metrics for 0% missing case. MWIR Day time videos.

<table>
<thead>
<tr>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000m</td>
<td></td>
<td></td>
<td></td>
<td>1500m</td>
<td></td>
<td></td>
<td></td>
<td>2000m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP2</td>
<td>1.00</td>
<td>12.64</td>
<td>0.00</td>
<td>69.61</td>
<td>BMP2</td>
<td>1.00</td>
<td>16.65</td>
<td>0.00</td>
<td>79.09</td>
<td>BMP2</td>
<td>1.00</td>
</tr>
<tr>
<td>BRDM2</td>
<td>1.00</td>
<td>25.09</td>
<td>0.73</td>
<td>77.77</td>
<td>BRDM2</td>
<td>1.00</td>
<td>18.94</td>
<td>0.66</td>
<td>99.96</td>
<td>BRDM2</td>
<td>1.00</td>
</tr>
<tr>
<td>BTR70</td>
<td>1.00</td>
<td>27.03</td>
<td>0.65</td>
<td>64.48</td>
<td>BTR70</td>
<td>1.00</td>
<td>17.11</td>
<td>0.88</td>
<td>99.72</td>
<td>BTR70</td>
<td>1.00</td>
</tr>
<tr>
<td>SV</td>
<td>1.00</td>
<td>25.09</td>
<td>0.65</td>
<td>64.48</td>
<td>SV</td>
<td>1.00</td>
<td>12.20</td>
<td>0.66</td>
<td>99.96</td>
<td>SV</td>
<td>1.00</td>
</tr>
<tr>
<td>T72</td>
<td>0.99</td>
<td>58.45</td>
<td>0.00</td>
<td>81.72</td>
<td>T72</td>
<td>1.00</td>
<td>26.71</td>
<td>0.00</td>
<td>95.61</td>
<td>T72</td>
<td>1.00</td>
</tr>
<tr>
<td>Truck</td>
<td>1.00</td>
<td>23.15</td>
<td>0.29</td>
<td>81.56</td>
<td>Truck</td>
<td>1.00</td>
<td>15.20</td>
<td>0.98</td>
<td>100.00</td>
<td>Truck</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 9. Tracking metrics for 50% missing case. MWIR Day time videos.

<table>
<thead>
<tr>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000m</td>
<td></td>
<td></td>
<td></td>
<td>1500m</td>
<td></td>
<td></td>
<td></td>
<td>2000m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP2</td>
<td>1.00</td>
<td>10.35</td>
<td>0.00</td>
<td>61.07</td>
<td>BMP2</td>
<td>1.00</td>
<td>7.36</td>
<td>1.00</td>
<td>100.00</td>
<td>BMP2</td>
<td>1.00</td>
</tr>
<tr>
<td>BRDM2</td>
<td>0.91</td>
<td>45.83</td>
<td>0.44</td>
<td>76.18</td>
<td>BRDM2</td>
<td>1.00</td>
<td>9.88</td>
<td>1.00</td>
<td>100.00</td>
<td>BRDM2</td>
<td>1.00</td>
</tr>
<tr>
<td>BTR70</td>
<td>0.74</td>
<td>52.61</td>
<td>0.76</td>
<td>57.00</td>
<td>BTR70</td>
<td>1.00</td>
<td>8.42</td>
<td>1.00</td>
<td>100.00</td>
<td>BTR70</td>
<td>1.00</td>
</tr>
<tr>
<td>SV</td>
<td>0.69</td>
<td>45.61</td>
<td>0.76</td>
<td>57.00</td>
<td>SV</td>
<td>1.00</td>
<td>7.36</td>
<td>1.00</td>
<td>100.00</td>
<td>SV</td>
<td>1.00</td>
</tr>
<tr>
<td>T72</td>
<td>0.85</td>
<td>20.31</td>
<td>0.72</td>
<td>81.94</td>
<td>T72</td>
<td>1.00</td>
<td>15.20</td>
<td>0.98</td>
<td>100.00</td>
<td>T72</td>
<td>1.00</td>
</tr>
<tr>
<td>Truck</td>
<td>0.65</td>
<td>34.39</td>
<td>0.99</td>
<td>56.33</td>
<td>Truck</td>
<td>1.00</td>
<td>10.77</td>
<td>1.00</td>
<td>73.11</td>
<td>Truck</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 10. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 0% missing case. Vehicle: SUV.

Table 9. Tracking metrics for 50% missing case. MWIR Day time videos.

<table>
<thead>
<tr>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
<th></th>
<th>EinGT</th>
<th>CLE</th>
<th>% of detections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000m</td>
<td></td>
<td></td>
<td></td>
<td>1500m</td>
<td></td>
<td></td>
<td></td>
<td>2000m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP2</td>
<td>0.99</td>
<td>37.65</td>
<td>0.02</td>
<td>69.60</td>
<td>BMP2</td>
<td>1.00</td>
<td>22.55</td>
<td>0.13</td>
<td>97.83</td>
<td>BMP2</td>
<td>1.00</td>
</tr>
<tr>
<td>BRDM2</td>
<td>1.00</td>
<td>26.90</td>
<td>0.12</td>
<td>83.00</td>
<td>BRDM2</td>
<td>1.00</td>
<td>18.94</td>
<td>0.66</td>
<td>99.96</td>
<td>BRDM2</td>
<td>1.00</td>
</tr>
<tr>
<td>BTR70</td>
<td>1.00</td>
<td>24.50</td>
<td>0.20</td>
<td>69.60</td>
<td>BTR70</td>
<td>1.00</td>
<td>16.65</td>
<td>0.77</td>
<td>99.96</td>
<td>BTR70</td>
<td>1.00</td>
</tr>
<tr>
<td>SV</td>
<td>1.00</td>
<td>19.51</td>
<td>0.62</td>
<td>76.00</td>
<td>SV</td>
<td>1.00</td>
<td>14.56</td>
<td>0.99</td>
<td>91.17</td>
<td>SV</td>
<td>1.00</td>
</tr>
<tr>
<td>T72</td>
<td>1.00</td>
<td>24.66</td>
<td>0.18</td>
<td>80.22</td>
<td>T72</td>
<td>1.00</td>
<td>18.18</td>
<td>0.86</td>
<td>84.89</td>
<td>T72</td>
<td>1.00</td>
</tr>
<tr>
<td>Truck</td>
<td>1.00</td>
<td>21.79</td>
<td>0.00</td>
<td>93.28</td>
<td>Truck</td>
<td>1.00</td>
<td>17.17</td>
<td>0.32</td>
<td>99.33</td>
<td>Truck</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 10. Tracking metrics for 75% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges. MWIR Day time videos.

<table>
<thead>
<tr>
<th></th>
<th>1000m</th>
<th>1500m</th>
<th>2000m</th>
<th>2500m</th>
<th>3000m</th>
<th>3500m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP2</td>
<td>0.89</td>
<td>31.78</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BRDM-2</td>
<td>0.97</td>
<td>31.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BTR70</td>
<td>0.98</td>
<td>29.84</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SUV</td>
<td>0.98</td>
<td>29.63</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>T72</td>
<td>0.99</td>
<td>28.98</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0.98</td>
<td>28.67</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Figure 11. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 50% missing case. Vehicle: SUV.
4.2. Classification Results

Classification is only applied to frames with detection of targets from the tracker. Table 11 and Table 13 summarize the ResNet classifiers for 0%, 50%, and 75% missing cases, respectively. The highlighted yellow cells indicate the biggest number of each row. The column labels in the confusion matrices are the ground truth and the row labels are the classifier output labels. As shown in the tables, the classification results using ResNet is in general not so good even at 1000 m with 75% missing. This means MWIR is not suitable for daytime surveillance.

Table 11. Classification results for 0% missing case. Left is the confusion matrix; the last column is the classification accuracy.

Table 12. Classification results for 50% missing case. Left is the confusion matrix; right is the classification results.
Table 13. Classification results for 75% missing case. Left is the confusion matrix; right is the classification results.

4.3. Discussions: Tracking and Classification at 75% rate at MWIR Day-Time

Since our research objective is to perform tracking and classification using compressive measurements, we focus on 75% missing pixel scenario where there are savings in both storage and transmission. From Table 14, we can see that both tracking and classification performance using MWIR videos is not good. Hence, MWIR is not suitable for target tracking and classification for day time videos.

Table 14. MWIR Day Time Summary at 75% missing rate.

5. TRACKING AND CLASSIFICATION RESULTS USING SENSIAC MWIR NIGHT-TIME VIDEOS

5.1. Train using 1500 m and 3000 m videos and Test using 1000 m, 2000 m, 2500 m, 3500 m videos

Now, we will focus on tracking and classification for MWIR night-time videos. Table 15-Table 17 show the metrics when we used videos from 1500 m and 3000 m for training and other videos from other ranges for testing. We can see that the numbers of frames with detection are high for low missing rates. For a given missing rate, the percentage of frames with detection decreases with longer ranges.

Figure 13-Figure 15 show the tracking results visually. It can be seen that we have good tracking up to 3500 m at 0% missing rate. At 50%, the tracking performance drops, as we can barely see some bounding boxes around vehicles up to range 2000 m. At 75% missing rate, the tracking is only good up to 1000 m.

Table 15. Tracking metrics for 0% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges.
Figure 13. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 0% missing case.

Vehicle is SUV.

Table 16. Tracking metrics for 50% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges.
Figure 14. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 50% missing case.

Vehicle is SUV.

Table 17. Tracking metrics for 75% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges.
5.2. Classification Results

The classification is still done with ResNet. As shown in Table 18-Table 20, we have the following observations:

- At 0% missing case, classification performance was reasonable for videos up to 2000 m. Some vehicles at 2000 m could be correctly classified even they could be correctly tracked at high percentage. The classification performance was not good for videos beyond 2500 m. For example, the accuracy at 3000 m was poor even using the same training videos for testing.

- At 50% missing case, we have the same observations as those at 0% missing rate. However, the percentages of correct classification were lower.

- Similarly, at 75% missing case, the percentages of correct classification dropped to even lower values even for ranges below 2000m.

Table 18. Classification results for 0% missing case. Left is the confusion matrix; the last column is the classification accuracy.

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>% of accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>137</td>
<td>208</td>
<td>0</td>
<td>172</td>
<td>587</td>
<td>4</td>
<td>3%</td>
</tr>
<tr>
<td>BRDM2</td>
<td>44</td>
<td>207</td>
<td>0</td>
<td>177</td>
<td>188</td>
<td>2</td>
<td>3%</td>
</tr>
<tr>
<td>BTR70</td>
<td>158</td>
<td>5</td>
<td>0</td>
<td>194</td>
<td>353</td>
<td>4</td>
<td>3%</td>
</tr>
<tr>
<td>T72</td>
<td>145</td>
<td>149</td>
<td>5</td>
<td>174</td>
<td>261</td>
<td>4</td>
<td>3%</td>
</tr>
<tr>
<td>Truck</td>
<td>41</td>
<td>47</td>
<td>1</td>
<td>34</td>
<td>183</td>
<td>4</td>
<td>3%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0</td>
<td>0</td>
<td>176</td>
<td>0</td>
<td>34</td>
<td>4</td>
<td>3%</td>
</tr>
</tbody>
</table>

At 0% missing case, classification performance was reasonable for videos up to 2000 m. Some vehicles at 2000 m could be correctly classified even they could be correctly tracked at high percentage. The classification performance was not good for videos beyond 2500 m. For example, the accuracy at 3000 m was poor even using the same training videos for testing.

At 50% missing case, we have the same observations as those at 0% missing rate.

Similarly, at 75% missing case, the percentages of correct classification dropped to even lower values even for ranges below 2000m.

Table 19. Classification results for 50% missing case. Left is the confusion matrix; the last column is the classification accuracy.

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>% of accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>23</td>
<td>27</td>
<td>0</td>
<td>124</td>
<td>403</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>BRDM2</td>
<td>133</td>
<td>252</td>
<td>0</td>
<td>136</td>
<td>278</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>BTR70</td>
<td>219</td>
<td>163</td>
<td>0</td>
<td>149</td>
<td>317</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>T72</td>
<td>156</td>
<td>147</td>
<td>0</td>
<td>133</td>
<td>327</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>Truck</td>
<td>57</td>
<td>61</td>
<td>0</td>
<td>136</td>
<td>385</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0</td>
<td>0</td>
<td>176</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>5%</td>
</tr>
</tbody>
</table>

Table 19. Classification results for 75% missing case. Left is the confusion matrix; the last column is the classification accuracy.

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>% of accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>57</td>
<td>274</td>
<td>0</td>
<td>149</td>
<td>323</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>BRDM2</td>
<td>274</td>
<td>144</td>
<td>0</td>
<td>149</td>
<td>323</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>BTR70</td>
<td>405</td>
<td>542</td>
<td>0</td>
<td>149</td>
<td>323</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>T72</td>
<td>102</td>
<td>372</td>
<td>5</td>
<td>349</td>
<td>428</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>Truck</td>
<td>35</td>
<td>51</td>
<td>0</td>
<td>39</td>
<td>49</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0</td>
<td>0</td>
<td>176</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>5%</td>
</tr>
</tbody>
</table>

Table 20. Classification results for 75% missing case. Left is the confusion matrix; the last column is the classification accuracy.

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>BMP2</th>
<th>BRDM2</th>
<th>BTR70</th>
<th>T72</th>
<th>Truck</th>
<th>ZSU23-4</th>
<th>% of accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>57</td>
<td>274</td>
<td>0</td>
<td>149</td>
<td>323</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>BRDM2</td>
<td>274</td>
<td>144</td>
<td>0</td>
<td>149</td>
<td>323</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>BTR70</td>
<td>405</td>
<td>542</td>
<td>0</td>
<td>149</td>
<td>323</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>T72</td>
<td>102</td>
<td>372</td>
<td>5</td>
<td>349</td>
<td>428</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>Truck</td>
<td>35</td>
<td>51</td>
<td>0</td>
<td>39</td>
<td>49</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td>ZSU23-4</td>
<td>0</td>
<td>0</td>
<td>176</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>5%</td>
</tr>
</tbody>
</table>
Signal & Image Processing: An International Journal (SIPIJ) Vol.10, No.6, December 2019

5.3. Discussions: Tracking and Classification at 75% rate at MWIR Night-Time

Tracking and Classification

Our main interest is the 75% missing data case. From Table 21, one can observe that tracking is good up to 3000 m and classification is good up to 1500 m. MWIR videos should be applied in night time.

Table 20. Classification results for 75% missing case. Left is the confusion matrix; the last column is the classification accuracy.

<table>
<thead>
<tr>
<th>Range</th>
<th>Average % of frames with detections</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>79.08</td>
<td>79%</td>
</tr>
<tr>
<td>1500</td>
<td>98.44</td>
<td>61%</td>
</tr>
<tr>
<td>2000</td>
<td>70.97</td>
<td>24%</td>
</tr>
<tr>
<td>2500</td>
<td>50.94</td>
<td>13%</td>
</tr>
<tr>
<td>3000</td>
<td>98.80</td>
<td>15%</td>
</tr>
<tr>
<td>3500</td>
<td>41.89</td>
<td>9%</td>
</tr>
</tbody>
</table>

Computational Efficiency

Here we briefly emphasize one key advantage of the proposed approach. For compressive measurements using the random subsampling approach, conventional trackers will require the
missing pixels to be reconstructed. For randomly missing pixels in the videos, simple interpolation algorithms can be used to reconstruct the videos. However, even the simple interpolation methods still take time. We ran some experiments to compare the execution times in terms of number of frames per second (fps). From Table 22, it can be seen that reconstruction slows down the execution time considerably. Tracking directly can achieve 7 fps for 50% and 75% missing cases, but the reconstruction based algorithm can only achieve 2 and 1 fps for 50% and 75% missing cases, respectively.

Table 22. Comparison of fps with and without missing data reconstruction for one video from 1000 m range.

<table>
<thead>
<tr>
<th>Missing rates</th>
<th>fps</th>
</tr>
</thead>
<tbody>
<tr>
<td>No pixel reconstruction</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
</tr>
<tr>
<td>With reconstruction</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

In this paper, a deep learning approach for multiple target tracking and classification directly in the compressive measurement domain is presented for low quality videos where the targets are small and the image resolution is poor. The compressive measurements are obtained via subsampling of the original image pixels. Experiments clearly demonstrated the performance of the proposed approach under different conditions even when the training data are limited.

YOLO is not suitable for classification of vehicles in low quality videos. However, it is good for tracking by detection. ResNet can be customized for classification and has achieved reasonable performance.

Optical camera is good for daytime up to 2000 m. Of course, if training data are available, tracking can be good up to 3000 m, but classification is not good. MWIR imager is not suitable for daytime tracking and classification. However, it is good for night-time up to 3000 m for tracking, but classification is not good even with training data at 3000m.

One future direction is to integrate the proposed approach with video cameras and perform real-time tracking and classification.

ACKNOWLEDGEMENTS

This research was supported by the US Air Force under contract FA8651-17-C-0017. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES


**AUTHORS**

**Chiman Kwan** received his BS with honours in Electronics from the Chinese University of Hong Kong in 1988, and MS and Ph.D. in Electrical Engineering from the University of Texas at Arlington in 1989 and 1993, respectively. Currently, he is the President of Applied Research LLC, leading research and development efforts in image and video processing, and remote sensing. He has 15 patents, 60 invention disclosures, 350 papers in journals and conferences, and over 500 proprietary technical reports. Over the past two decades, he has secured more than 120 diverse research projects totalling more than 36 million dollars from various government agencies and private companies. He also received numerous awards and recognitions from IEEE, NASA, and other agencies.

**Bryan Chou** graduated in Electrical Engineering from the George Mason University. In the past three years, he has worked on image demosaicing, image super-resolution, pansharpening, image fusion, and multi-target tracking.

**Jonathan Yang** graduated from the Computer Science Department of the University of Maryland. He was with Applied Research LLC and is now with Google Inc.

**Trac D. Tran** received his BS from MIT and Ph.D. from U. Wisconsin Madison. He is a Professor in the Department of Electrical and Computer Engineering, The Johns Hopkins University. His pioneering research on integer-coefficient transforms and pre-/post-filtering operators has been adopted as critical components of Microsoft Windows Media Video 9 and JPEG XR. He was the recipient of the NSF Career Award and is a Fellow of IEEE.