TARGET DETECTION AND CLASSIFICATION IMPROVEMENTS USING CONTRAST ENHANCED 16-BIT INFRARED VIDEOS

Chiman Kwan and David Gribben

Applied Research, LLC, Rockville, Maryland, USA

ABSTRACT

In our earlier target detection and classification papers, we used 8-bit infrared videos in the Defense Systems Information Analysis Center(DSIAC) video dataset. In this paper, we focus on how we can improve the target detection and classification results using 16-bit videos. One problem with the 16-bit videos is that some image frames have very low contrast. Two methods were explored to improve upon previous detection and classification results. The first method used to improve contrast was effectively the same as the baseline 8-bit video data but using the 16-bit raw data rather than the 8-bit data taken from the avi files. The second method used was a second order histogram matching algorithm that preserves the 16-bit nature of the videos while providing normalization and contrast enhancement. Results showed the second order histogram matching algorithm improved the target detection using You Only Look Once (YOLO) and classificationusing Residual Network (ResNet) performance. The average precision (AP) metric in YOLO was improved by 8%. This is quite significant. The overall accuracy (OA) of ResNet has been improved by 12%. This is also very significant.

KEYWORDS

Deep learning, mid-wave infrared (MWIR) videos, target detection and classification, contrast enhancement, YOLO, ResNet

1. INTRODUCTION

Target detection algorithms for infrared videos can be divided into two groups. One group is to utilize supervised machine learning algorithms. For instance, some conventional target tracking methods [1]-[5] belong to this group. Target locations may need to be specified in the first frame of the videos. The second group of target detection and classification schemes uses deep learning algorithms such as You Only Look Once (YOLO) for optical and infrared videos [6]-[27]. Training videos are required in these algorithms. However, there is no need to specify the target locations in the first frame when using YOLO. Among those deep learning algorithms, it is worth mentioning that some of them [6]-[16] are using compressive measurements directly for target detection and classification. This means that no reconstruction of compressive measurements is needed and hence fast target detection and classification can be achieved. The algorithms in [17]-[26] require target locations to be known.

The image quality of infrared videos in ground based imagers is of low quality due to the presence of air turbulence, sensor noise, etc. In addition, the image contrast may also be poor in these infrared videos. In practical applications, the image quality of infrared videos may seriously affect the target detection and classification performance. In our earlier papers[12][13], we used 8-bit videos even though the raw DSIAC videos are in 16-bit format; someone else already converted 16-bit videos to 8-bit in the DSIAC database. Some contrast enhancement was applied

to those 8-bit videos in our earlier papers. In this research, we would like to investigate the use of 16-bit videos for target detection and classification and see how much improvement we can achieve over our earlier results of using 8-bit videos. When the raw 16-bit data were used at the beginning of this research, the images were much darker and hence requiredsignificant contrast enhancement. We then decided to focus on evaluating the impact of contrast enhancement techniques on target detection and classification performance of deep learning algorithms using 16-bit videos. In particular, we carried out extensive evaluations of two contrast enhancement techniques, which are all simple and efficient to implement. The first method uses an 8-bit image with decent contrast as reference and all the 16-bit videos are histogram matched to the reference image. Compared with the previous results using 8-bit videos, the new detection and classification results with 16-bit videos improved over the earlier results. The second method uses a second order contrast enhancement algorithm which achieves contrast enhancement and at the same time retains the 16-bit video format. Experiments showed that the results using the second method improved over the first method as well as earlier results in [12][13].

Our contributions are as follows. First,we evaluated two simple and efficient contrast enhancement algorithms that can improve the video quality in real 16-bit infrared videos. Second, using many DSIAC videos, we demonstrated that using 16-bit videos with contrast enhancement can significantly improve the target detection and classification performance.

Our paper is organized as follows. Section 2 describes the contrast enhancement methods, target detection and classification algorithms, performance metrics, and infrared videos. Section 3 summarizes the experimental results. Finally, some remarks are included in Section 4.

2. METHODS, PERFORMANCE METRICS AND DATA

2.1. Contrast Enhancement Methods

The raw videos are with a bit depth of 16. Figure 1shows one frame with 16-bit. It can be seen that the image can be quite dark and hence contrast enhancement is needed in order for target detection and tracking algorithm to function properly.

Figure 1. 16-bit raw frame.

Although there are quite a few image enhancement algorithms in the literature, we experimented with two simple and efficient approaches to enhancing the contrast of these raw frames. In particular, we used the following approaches.

Approach 1: Histogram matching to an 8-bit reference frame

Approach 1 was implemented using the MATLAB function imhistmatch. It should be noted that the contrast enhancement was done off-line in the pre-processing step. One issue with this approach is that when a 16-bit image is histogram matched to an 8-bit reference image with good contrast, the bit depth of the resulting image is still 8-bit. In addition, it can be seen from Figure 2that there are overly bright areas in the background.

(a) Raw image (b) Contrast enhanced image

Figure 2. Before and after comparison of applying Approach 1.

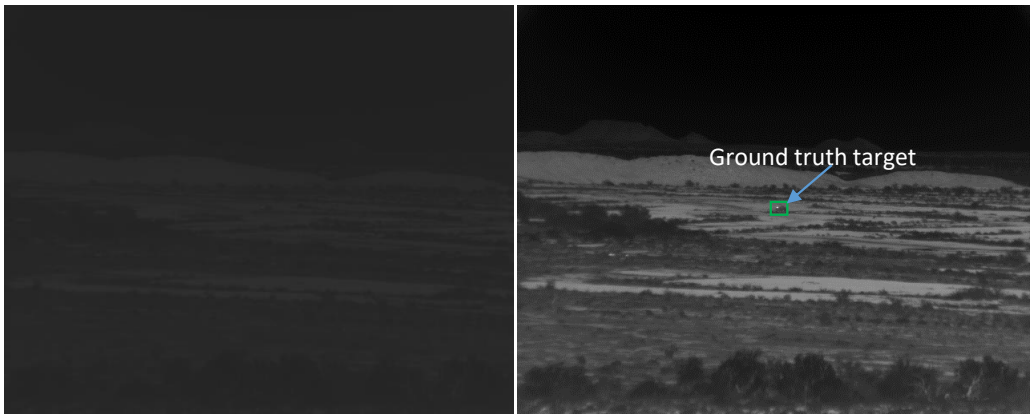
Approach 2: Second order histogram matching

Approach 1 used an 8-bit video frame with decent quality as reference. Consequently, the 16-bit low contrast videos are matched to 8-bit intervals. In practical applications, it will be important to retain the 16-bit data quality in the raw videos. Here, we applied a simple second-order contrast enhancement method that has been widely used in remote sensing [32]. This method preserves the 16-bit data quality in the raw video and is a simple normalization and histogram matching algorithm denoted by

$$J = \frac{\mathrm{ref_{std}}}{\mathrm{I_{std}}}(I - \mathrm{I_{mean}}) + \mathrm{ref_{mean}} \quad (1)$$

In the equation, J is the resulting image, $\mathrm{ref_{std}}$ is the numeric distance between standard deviations in the reference image, $\mathrm{I_{std}}$ is the numeric distance between standard deviations in the original image, $\mathrm{I_{mean}}$ is the mean value of the original image, and $\mathrm{ref_{mean}}$ is the mean value of the reference image. The reference image used was an image from the middle of one DSIAC video that is compressed to 8 bits as it has the best histogram of any set of images.

Approach 2 uses a simple formula found in [32] to perform histogram matching.Figure 3illustrates the differences between images before and after applying Approach 2. Comparing Figure 2and Figure 3shows that the frame of Approach 2 appears to be have better contrast.



(a) Raw image (b) Contrast enhanced image

Figure 3. Before and after comparison of Approach 3.

2.2. YOLO for Target Detection

In the literature, there are some deep learning based object detectors such as YOLO and Faster R-CNN that do not require initial bounding boxes and can simultaneously detect objects. The YOLO detector [29] is fast and has similar performance to the Faster R-CNN [30]. The input image is resized to 448x448. There are 24 convolutional layers and two fully connected layers. The output is 7x7x30. We have used YOLOv2 because it is more accurate than YOLO version 1. The training of YOLO is quite simple. Images with ground truth target locations are needed. 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.

YOLO also comes with a built-in classification module. However, based on our earlier evaluations, the classification accuracy using YOLO's built-in module is not good as compared to ResNet [6]-[11].

2.3. ResNet for Target Classification

As mentioned In Section 1, YOLO's built-in classifier did not perform well, which is probably due to limited training data. Moreover, we think that, although YOLO is good for object detection, its built-in classifier is probably more suitable for inter-class (humans, traffic signs, vehicles, buses, etc.) discrimination and not good for inter-class (e.g. BTR70 vs. BMP2) discrimination.The ResNet-18 model[31] is an 18-layer convolutional neural network (CNN) that can avoid performance saturation when training deeper layers.

It is important to explain the relationship between YOLO and ResNet in our paper. YOLO[29] was used to determine where, in each frame, the vehicles were located. YOLO generated bounding boxes for those vehicles and that data were used to crop the vehicles from the image. The cropped vehicles would be fed into the ResNet-18 for classification and classification results were generated. To be more specific, ResNet-18 is used directly after bounding box information is obtained from YOLO.

Training of ResNet requires target patches. The targets are cropped from training videos. Mirror images are then created. We then perform data augmentation using scaling (larger and smaller), rotation (every 45 degrees), and illumination (brighter and dimmer) to create more training data.

For each cropped target, we are able to create a data set with 64 more images. For ResNet, the last layer of the deep learning model was re-trained. The ResNet model was trained until the validation score plateaued.

2.4. Performance Metrics for Assessing Target Detection and Classification Performance

The six different performance metrics to quantify the detection performance are: Center Location Error (CLE), Distance Precision at 10 pixels (DP@10), Estimates in Ground Truth (EinGT), Intersection over Union (IoU), Average Precision (AP), and number of frames with detection. These metrics are detailed below:

- Center Location Error (CLE): This is the error between the center of the bounding box and the ground-truth bounding box. Smaller means better. CLE is calculated by measuring the distance between the ground truth center location $(C_{x,gt}, C_{y,gt})$ and the detected center location $(C_{x,est}, C_{y,est})$. Mathematically, CLE is given by

$$\mathrm{CLE} = \sqrt{\left(C_{x,est} - C_{x,gt}\right)^2 + \left(C_{y,est} - C_{y,gt}\right)^2}. \quad (2)$$

- Distance Precision (DP): This is the percentage of frames where the centroids of detected bounding boxes are within 10 pixels of the centroid of ground-truth bounding boxes. Close to 1 or 100% indicates good results.

- Estimates in Ground Truth (EinGT): This is the percentage of the frames where the centroids of the detected bounding boxes are inside the ground-truth bounding boxes. It depends on the size of the bounding box and is simply a less strict version of the DP metric. Close to 1 or 100% indicates good results.

- Intersection over the Union (IoU): It is the ratio of the intersected area over the union of the estimated and ground truth bounding boxes.

$$\mathrm{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (3)$$

- Average Precision (AP): AP is the ratio between the intersection area and the area of the estimated bounding box and the value is between 0 and 1, with 1 or 100% being the perfect overlap. The AP being used can be computed as

$$\mathrm{AP} = \frac{\text{Area of Intersection}}{\text{Area of estimated bounding boxes}}. \quad (4)$$

  As shown in Equation (4), AP is calculated by taking the area of intersection of the ground truth bounding box and the estimated bounding box, then dividing that area by the union of those two areas.

- Number of frames with detection: This is the total number of frames that have detection.

We used confusion matrices for evaluating vehicle classification performance using ResNet. From the confusion matrix, we can also evaluate overall accuracy (OA), average accuracy (AA), and kappa coefficient.

2.5. DSIAC Data

We selected five vehicles in the DSIAC videos for detection and classification. There are optical and mid-wave infrared (MWIR) videos collected at distances ranging from 1000 m to 5000 m with 500 m increments. The five types of vehicles are shown inFigure 4. These videos are challenging for several reasons. First, the target sizes are small due to long distances. This is quite different from some benchmark datasets such as MOT Challenge [28] where the range is short and the targets are big. Second, the target orientations also change drastically. Third, the illuminations in different videos are also different. Fourth, the cameras also move in some videos.

In this research, we focus mostly on MWIR night-time videos because MWIR is more effective for surveillance during the nights.

(a) (b) (c)

(d) (e)

Figure 4. Five vehicles in DSIAC: (a) BTR70; (b) BRDM2; (c) BMP2; (d) T72; and (e) ZSU23-4.

3. EXPERIMENTAL RESULTS

In this section, we summarize two experiments. First, we histogram matched the 16-bit low contrast raw videos to a reference video in 8-bit format. The resulting videos, however, are still 8-bit. Although the enhanced videos are still 8-bit, we have seen some positive improvements in terms of target detection using YOLO and target classification using ResNet. Second, we investigated another contrast enhancement method that can preserve the 16-bit videos and yet can generate better contrast videos. The performance of target detection and classification has been further improved.

3.1. Histogram Matching Results (16-bit to 8-bit)

Previous processes in our earlier papers to generate images from the raw data did not preserve the 16-bit nature of the data.The raw 16-bit infrared image for every different distance and vehicle is quite dark, as seen inFigure 1. In order for it to be used by the different detection and classification methods, different processes need to be made to the images to enhance contrast while preserving the data as much as possible.

The first method attempted was simply histogram matching the 16-bit image to an 8-bit reference image, which was found in our earlier studies. The biggest difference in quality was that the image was no longer being saved in JPEG format and therefore was not losing information each time it was saved, in combination with the increased image quality from using a 16-bit image rather than an 8-bit image. The issue with this method is that because it was histogram matching to an 8-bit reference image, the new histogram matched image was then converted to the quality of the reference image, which is still 8-bit. Regardless, a model for both YOLO and ResNet was trained and performance metrics were generated to see what improvement, if any, there would be. Figure 5 shows one histogram matched frame.

Figure 5. Histogram matched MWIR image (8-bit) at 3500 meter distance with BRDM2 vehicle.

Table 1 contains all detection results generated from the YOLO model while Table 2 contains all classification results generated by ResNet. 1500 m videos were used for training and videos in other ranges were used for testing. Further analysis will be provided in the observational remarks section but in general each distance slightly improves upon the original histogram matched image (8-bit to 8-bit). However, the improvements are not much.

Table 1. YOLO results for the new model trained on the new histogram matched images. This is the 16-bit to 8-bit case.

1000 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	3.677	100.00%	100.00%	67.89%	74.48%	99.89%
BRDM2	3.829	100.00%	100.00%	72.02%	85.33%	92.22%
BMP2	3.762	100.00%	100.00%	70.52%	93.22%	99.22%
T72	3.638	100.00%	100.00%	72.90%	81.78%	85.17%
ZSU23-4	3.316	100.00%	100.00%	75.61%	81.40%	94.28%
Avg	3.645	100.00%	100.00%	71.79%	83.24%	94.15%

1500 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	1.401	100.00%	100.00%	83.00%	87.72%	100.00%
BRDM2	1.266	100.00%	100.00%	83.16%	89.76%	100.00%
BMP2	1.293	100.00%	100.00%	86.42%	93.06%	100.00%
T72	1.491	100.00%	100.00%	85.90%	90.95%	100.00%
ZSU23-4	1.387	100.00%	100.00%	81.45%	84.93%	100.00%
Avg	1.368	100.00%	100.00%	83.99%	89.29%	100.00%

2000 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	2.039	100.00%	100.00%	46.21%	46.28%	91.72%
BRDM2	2.328	100.00%	100.00%	52.79%	52.88%	99.39%
BMP2	2.005	100.00%	100.00%	59.22%	59.44%	68.61%
T72	1.467	100.00%	100.00%	51.99%	52.05%	94.44%
ZSU23-4	2.208	99.95%	99.95%	50.97%	51.04%	99.06%
Avg	2.009	99.99%	99.99%	52.24%	52.34%	90.64%

2500 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	2.804	99.89%	99.89%	16.85%	16.86%	36.61%
BRDM2	3.193	100.00%	99.12%	18.85%	18.85%	71.44%
BMP2	22.030	91.97%	91.97%	21.60%	21.60%	28.78%
T72	2.978	100.00%	100.00%	24.18%	24.18%	45.44%
ZSU23-4	3.046	100.00%	100.00%	19.23%	19.23%	37.89%
Avg	6.810	98.37%	98.20%	20.14%	20.14%	44.03%

3000 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	1.843	100.00%	100.00%	8.44%	8.44%	10.33%
BRDM2	4.367	100.00%	98.52%	11.16%	11.16%	13.94%
BMP2	5.242	100.00%	0.00%	11.80%	11.80%	0.11%
T72	5.033	100.00%	93.14%	14.12%	14.12%	18.00%
ZSU23-4	3.137	100.00%	100.00%	12.15%	12.15%	15.00%
Avg	3.924	100.00%	78.33%	11.54%	11.54%	11.48%

3500 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	1.860	100.00%	71.05%	2.50%	2.50%	1.83%
BRDM2	3.795	100.00%	45.24%	2.79%	2.79%	2.28%
BMP2	n/a	n/a	n/a	n/a	n/a	0.00%
T72	4.692	100.00%	25.43%	3.51%	3.51%	16.06%
ZSU23-4	3.578	100.00%	57.61%	2.98%	2.98%	10.78%
Avg	3.481	100.00%	49.83%	2.94%	2.94%	6.19%

Table 2. ResNet results for new histogram matched images. This is the 16-bit to 8-bit case.

1000 m	5	6	9	11	12	
BTR70	1810	9	0	0	5	
BRDM2	3	1996	8	1	110	
BMP2	2	0	1892	115	10	
T72	19	6	92	1676	18	
ZSU23-4	5	20	2	119	1767	
Class Stats	OA	94.38%	AA	94.42%	kappa	0.9298

1500 m	5	6	9	11	12	
BTR70	1800	0	0	0	0	
BRDM2	0	1800	0	0	0	
BMP2	0	0	1812	0	0	
T72	0	0	0	1800	0	
ZSU23-4	0	0	0	0	1800	
Class Stats	OA	100%	AA	100%	kappa	1.000

2000 m	5	6	9	11	12	
BTR70	2032	0	14	0	0	
BRDM2	122	1930	26	0	8	
BMP2	0	0	1223	0	12	
T72	42	1	409	1702	298	
ZSU23-4	9	1	29	1	1922	
Class Stats	OA	90.06%	AA	91.65%	kappa	0.8758

2500 m	5	6	9	11	12	
BTR70	564	53	61	190	67	
BRDM2	18	989	128	1	453	
BMP2	4	0	479	23	67	
T72	0	5	141	454	468	
ZSU23-4	0	19	144	35	627	
Class Stats	OA	62.38%	AA	64.93%	kappa	0.5298

3000 m	5	6	9	11	12	
BTR70	11	17	96	44	86	
BRDM2	0	10	9	1	250	
BMP2	0	0	0	0	2	
T72	0	88	29	3	303	
ZSU23-4	0	50	26	24	224	
Class Stats	OA	19%	AA	16%	kappa	-0.006

3500 m	5	6	9	11	12	
BTR70	15	0	20	3	0	
BRDM2	0	0	8	7	27	
BMP2	0	0	0	0	0	
T72	5	12	126	121	86	
ZSU23-4	0	0	80	22	174	
Class Stats	OA	43.91%	AA	27.42%	kappa	0.2989

3.2. Enhanced Results Using a Second Order Contrast Enhancement Method (16-bit to 16-bit)

The previous histogram matching in Section 3.1 used an 8-bit video as reference. As a result, the 16-bit low contrast videos are matched to 8-bit intervals. Here, we applied a simple second-order contrast enhancement method mentioned in Section 2. This method can preserve the 16-bit data and is a simple normalization and histogram matching algorithm. The reference image used was an image from the middle of the BTR70 vehicle video that is compressed to 8 bits as it has the

best histogram of any set of images. Figure 6 shows the resulting image generated from this method.

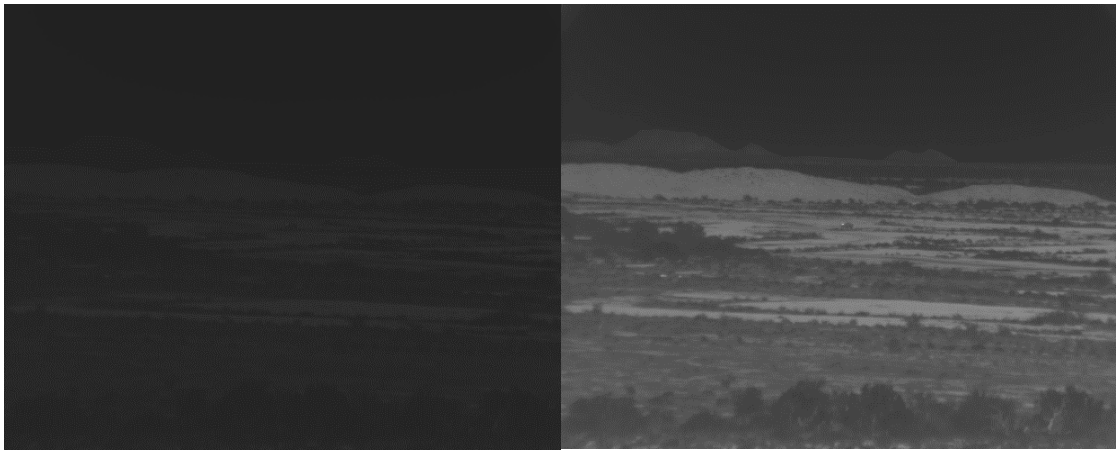


Figure 6. Enhanced MWIR image (16-bit, right) and original MWIR image (16-bit, left) of BRDM2 vehicle at 3500 meters.

The results for this method prove to be an improvement on the other method. Table 3contains the detection results for the YOLO model and Table 4 shows the classification results for ResNet. Additional remarks will be given in Section 3.3.

Table 3. Detection statistics for the YOLO model using second order contrast enhancement videos (16-bit to 16-bit).

1000 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	3.695	100.00%	100.00%	72.66%	79.27%	93.78%
BRDM2	2.956	100.00%	100.00%	77.04%	90.96%	99.94%
BMP2	4.613	100.00%	100.00%	72.24%	86.72%	90.89%
T72	3.929	100.00%	100.00%	76.31%	86.35%	99.89%
ZSU23-4	3.901	99.87%	99.87%	74.64%	86.93%	99.00%
Avg	3.819	99.97%	99.97%	74.58%	86.05%	96.70%

1500 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	1.363	100.00%	100.00%	87.38%	93.31%	100.00%
BRDM2	1.334	100.00%	100.00%	87.65%	93.45%	100.00%
BMP2	1.271	100.00%	100.00%	83.94%	98.66%	100.00%
T72	1.579	100.00%	100.00%	87.61%	94.49%	100.00%
ZSU23-4	2.169	100.00%	100.00%	81.90%	93.64%	98.89%
Avg	1.543	100.00%	100.00%	85.70%	94.71%	99.78%

2000 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	1.604	100.00%	100.00%	53.21%	53.27%	96.78%
BRDM2	2.719	100.00%	100.00%	54.07%	54.12%	99.72%
BMP2	1.639	100.00%	100.00%	64.43%	64.49%	87.22%
T72	1.634	100.00%	100.00%	60.79%	60.92%	95.50%
ZSU23-4	1.886	100.00%	100.00%	60.44%	60.51%	97.89%
Avg	1.896	100.00%	100.00%	58.59%	58.66%	95.42%

2500 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	2.762	99.90%	99.81%	18.99%	18.99%	51.28%
BRDM2	3.253	100.00%	99.32%	21.60%	21.60%	40.56%
BMP2	4.086	99.55%	99.55%	26.32%	26.34%	24.00%
T72	3.414	100.00%	100.00%	25.43%	25.43%	46.67%
ZSU23-4	2.915	100.00%	100.00%	17.97%	17.97%	38.00%
Avg	3.286	99.89%	99.73%	22.06%	22.07%	40.10%

3000 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	2.337	100.00%	99.64%	7.77%	7.77%	23.78%
BRDM2	4.588	100.00%	98.18%	11.72%	11.72%	13.72%
BMP2	4.123	100.00%	100.00%	17.63%	17.63%	0.06%
T72	4.478	100.00%	82.84%	14.67%	14.67%	29.72%
ZSU23-4	2.413	100.00%	100.00%	12.03%	12.03%	15.22%
Avg	3.588	100.00%	96.13%	12.76%	12.76%	16.50%

3500 m	CLE	DP	EinGT	IoU	AP	% det.
BTR70	2.750	100.00%	71.19%	2.14%	2.14%	3.22%
BRDM2	4.181	100.00%	36.67%	2.83%	2.83%	1.67%
BMP2	n/a	n/a	n/a	n/a	n/a	0.00%
T72	3.867	100.00%	62.11%	3.51%	3.51%	9.67%
ZSU23-4	2.564	100.00%	45.49%	2.75%	2.75%	13.39%
Avg	3.340	100.00%	53.86%	2.81%	2.81%	5.59%

Table 4. Classification results for the ResNet model using second order contrast enhancement videos (16-bit to 16-bit).

1000 m	5	6	9	11	12	
BTR70	2139	0	0	50	0	
BRDM2	0	1875	0	13	0	
BMP2	2	0	1713	446	1	
T72	1	0	0	1964	0	
ZSU23-4	28	0	3	326	1934	
Class Stats	OA	91.71%	AA	92.13%	kappa	0.9171

1500 m	5	6	9	11	12	
BTR70	1800	0	0	0	0	
BRDM2	0	1800	0	0	0	
BMP2	0	0	1800	0	0	
T72	0	0	0	1800	0	
ZSU23-4	0	0	0	1	1779	
Class Stats	OA	99.99%	AA	99.99%	kappa	0.9999

2000 m	5	6	9	11	12	
BTR70	2097	0	25	1	0	
BRDM2	0	1837	0	0	0	
BMP2	0	0	1570	0	0	
T72	0	0	40	2123	10	
ZSU23-4	0	3	93	1	1740	
Class Stats	OA	98.19%	AA	98.24%	kappa	0.9819

2500 m	5	6	9	11	12	
BTR70	244	71	9	663	42	
BRDM2	0	540	7	135	51	
BMP2	0	38	330	9	64	
T72	0	0	10	1015	0	
ZSU23-4	2	0	143	267	386	
Class Stats	OA	62.47%	AA	63.92%	kappa	0.6247

3000 m	5	6	9	11	12	
BTR70	7	132	46	329	43]	
BRDM2	2	187	15	29	41	
BMP2	0	1	0	0	0	
T72	20	55	42	321	133	
ZSU23-4	0	5	81	68	135	
Class Stats	OA	38.42%	AA	34.49%	kappa	0.3842

3500 m	5	6	9	11	12	
BTR70	0	29	9	21	0	
BRDM2	0	4	14	11	1	
BMP2	0	0	0	0	0	
T72	1	0	22	155	12	
ZSU23-4	0	0	160	69	26	
Class Stats	OA	34.64%	AA	21.02%	kappa	0.3464

3.3. Comparisons and Key Observations

There are several important observations to be seen from the two methods that had new YOLO and ResNet models generated for them. Compared to the original histogram matching results shown in Table 5, there is overall improvement. Taking the average results for each distance and combining it into one table where an average of those values is also taken shows a good representation of the overall performance of a given model. Table 5 shows the results in that format for the original baseline data. Table 6 shows the results for the new histogram matched data. Table 7 shows the results for the second order matching algorithm. Looking at these tables together helps show the improvement between the original method and the two generated for this work.

Comparing the YOLO results between each method there is incremental improvement when looking between each method. The new histogram matched method (16-bit to 8-bit) generates better results for CLE, DP, IoU, and AP when compared to the original method (8-bit to 8-bit) in our previous papers. The same is true for the second order contrast enhancement algorithm (16-bit to 16-bit), but the difference is that CLE is much improved than the original. The AP metric has been improved by 8% using the second order histogram matching method. This is very significant. The other improved statistics are also improved from the new histogram matching method but less extremely.

The next set of data to observe is the ResNet classification results. Again, for each method, the averages were taken for each distance as well as an average of all distances. Table 8 shows the results for the original baseline data. Table 9 shows the results for the new histogram matched data. Table 10 shows the results for the second order contrast enhancement algorithm. The overall accuracy (OA) has been improved by 12%. This is very significant.

Table 5. Average performance metrics for each vehicle distance for the earlier model in our previous papers (8-bit to 8-bit). 1500 m videos were used for training; other videos were used for testing.

Distance (m)	CLE	DP	EinGT	IoU	AP	% det.
1000	3.698	100.00%	100.00%	72.12%	76.63%	95.56%
1500	1.260	100.00%	100.00%	79.92%	80.45%	90.88%
2000	3.931	99.57%	99.57%	40.86%	41.00%	81.76%
2500	7.366	97.95%	97.80%	18.33%	18.33%	81.91%
3000	3.225	100.00%	94.82%	11.57%	11.57%	25.58%
3500	2.641	99.87%	76.69%	2.84%	2.84%	31.03%
Avg	3.687	99.57%	94.81%	37.61%	38.47%	67.79%

Table 6. Average performance metrics for each vehicle distance for the new histogram matched model (16-bit to 8-bit). 1500 m videos were used for training; other videos were used for testing.

Distance (m)	CLE	DP	EinGT	IoU	AP	% det.
1000	3.645	100.00%	100.00%	71.79%	83.24%	94.15%
1500	1.368	100.00%	100.00%	83.99%	89.29%	100.00%
2000	2.009	99.99%	99.99%	52.24%	52.34%	90.64%
2500	6.810	98.37%	98.20%	20.14%	20.14%	44.03%
3000	3.924	100.00%	78.33%	11.54%	11.54%	11.48%
3500	3.481	100.00%	49.83%	2.94%	2.94%	6.19%
Avg	3.540	99.73%	87.73%	40.44%	43.25%	57.75%

Table 7. Average performance metrics for each vehicle distance for the second order contrast enhancement algorithm model (16-bit to 16-bit). 1500 m videos were used for training; other videos were used for testing.

Distance (m)	CLE	DP	EinGT	IoU	AP	% det.
1000	3.819	99.97%	99.97%	74.58%	86.05%	96.70%
1500	1.543	100.00%	100.00%	85.70%	94.71%	99.78%
2000	1.896	100.00%	100.00%	58.59%	58.66%	95.42%
2500	3.286	99.89%	99.73%	22.06%	22.07%	40.10%
3000	3.588	100.00%	96.13%	12.76%	12.76%	16.50%
3500	3.340	100.00%	53.86%	2.81%	2.81%	5.59%
Avg	2.912	99.98%	91.62%	42.75%	46.18%	59.01%

Table 8. Average performance metrics for the original histogram matching ResNet model (8-bit to 8-bit). 1500 m videos were used for training; other videos were used for testing.

Distance (m)	OA	AA	kappa
1000	89.76%	89.74%	0.900
1500	99.98%	99.98%	1.000
2000	84.99%	86.50%	0.850
2500	50.89%	52.61%	0.510
3000	10.22%	27.53%	0.100
3500	16.66%	27.06%	0.170
Avg	58.75%	63.90%	0.588

Table 9. Average performance metrics for the new histogram matching ResNet model (16-bit to 8-bit). 1500 m videos were used for training; other videos were used for testing.

Distance (m)	OA	AA	kappa
1000	94.38%	94.42%	0.930
1500	100.00%	100.00%	1.000
2000	90.06%	91.65%	0.876
2500	62.38%	64.93%	0.530
3000	19.48%	15.58%	-0.006
3500	43.91%	27.42%	0.299
Avg	68.37%	65.67%	0.605

Table 10. Average performance metrics for the second order histogram matched ResNet model (16-bit to 16-bit). 1500 m videos were used for training; other videos were used for testing.

Distance (m)	OA	AA	kappa
1000	91.71%	92.13%	0.917
1500	99.99%	99.99%	1.000
2000	98.19%	98.24%	0.982
2500	62.47%	63.92%	0.625
3000	38.42%	34.49%	0.384
3500	34.64%	21.02%	0.346
Avg	70.90%	68.30%	0.709

4. CONCLUSIONS

In this paper, we focus on target detection and classification performance improvements using contrast enhanced infrared videos. Overall, the new histogram matching method from 16-bit to 8-bit improves on the old method (8-bit videos) in each category with the largest jump coming in the overall accuracy. The second order contrast enhancement model not only improves on the OA of both competing methods, in the original methods case a 12% improvement. It also beats the overall kappa by 12%. This improvement is systematic with each distance of the second order contrast enhancement method beating each value for the original. It can definitively be determined that the second order contrast enhancement method is worth using in the future over the past baseline for image correction with this dataset.

The detection and classification results in this paper used off-line image contrast enhancement methods. The YOLO detector and ResNet classifier are also not fast enough for real-time applications. One future direction is to integrate contrast enhancement algorithms with a fast object detector and classifier so that real-time experiments can be carried out. Another direction is to further investigate the use of super-resolution algorithms for target detection and classification performance enhancement.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by the US Army under contract W909MY-20-P-0024. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

[1] C. Kwan, B. Chou, andL. M. Kwan, "A Comparative Study of Conventional and Deep Learning Target Tracking Algorithms for Low Quality Videos," 15th International Symposium on Neural Networks, 2018.

[2] C. Kwan and B. Budavari, "Enhancing Small Target Detection Performance in Low Quality and Long Range Infrared Videos Using Optical Flow Techniques," Remote Sensing, 12(24), 4024, December 9, 2020.

[3] Y. Chen, G. Zhang, Y. Ma, J. U. Kang, and C. Kwan, "Small Infrared Target Detection based on Fast Adaptive Masking and Scaling with Iterative Segmentation," IEEE Geoscience and Remote Sensing Letters, January 2021.

[4] C. Kwan and B. Budavari, "A High Performance Approach to Detecting Small Targets in Long Range Low Quality Infrared Videos," arXiv:2012.02579, 2020.

[5] H. S. Demir and A. E. Cetin,"Co-difference based object tracking algorithm for infrared videos," IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 434-438

[6] C. Kwan, B. Chou, J. Yang, and T. Tran, "Compressive object tracking and classification using deep learning for infrared videos," Proc. SPIE 10995, Pattern Recognition and Tracking (Conference SI120). 2019.

[7] C. Kwan, B. Chou, J. Yang, and T. Tran, "Target tracking and classification directly in compressive measurement for low quality videos," SPIE 10995, Pattern Recognition and Tracking XXX, 1099505, 13 May 2019.

[8] C. Kwan, B.Chou, A. Echavarren, B. Budavari, J. Li, and T. Tran, "Compressive vehicle tracking using deep learning," IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York City. 2018.

[9] C. Kwan, D. Gribben, and T. Tran, "Multiple Human Objects Tracking and Classification Directly in Compressive Measurement Domain for Long Range Infrared Videos," IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York City. 2019.
[10] C. Kwan, D. Gribben, and T. Tran, "Tracking and Classification of Multiple Human Objects Directly in Compressive Measurement Domain for Low Quality Optical Videos," IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York City. 2019.
[11] C. Kwan, B. Chou, J. Yang, and T. Tran, "Deep Learning based Target Tracking and Classification Directly in Compressive Measurement for Low Quality Videos," Signal & Image Processing: An International Journal (SIPIJ), November 16, 2019.
[12] C. Kwan, B. Chou, J. Yang, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, "Target tracking and classification directly using compressive sensing camera for SWIR videos," Journal of Signal, Image, and Video Processing, June 7, 2019.
[13] C. Kwan, B. Chou, J. Yang, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, "Target tracking and classification using compressive measurements of MWIR and LWIR coded aperture cameras,"Journal Signal and Information Processing. 10, 73–95, 2019.
[14] C. Kwan, D. Gribben, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, "Detection and Confirmation of Multiple Human Targets Using Pixel-Wise Code Aperture Measurements," J. Imaging. 6(6), 40, 2020.
[15] C. Kwan, B. Chou, J. Yang, and T. Tran, "Deep Learning based Target Tracking and Classification for Infrared Videos Using Compressive Measurements," Journal Signal and Information Processing, November 2019.
[16] C. Kwan, B. Chou, J. Yang, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, "Deep Learning based Target Tracking and Classification for Low Quality Videos Using Coded Aperture Camera," Sensors, 19(17), 3702, August 26, 2019.
[17] S. Lohit, K. Kulkarni, and P. K. Turaga, "Direct inference on compressive measurements using convolutional neural networks," Int. Conference on Image Processing. 2016. 1913-1917.
[18] A. Adler, M. Elad, and M. Zibulevsky, "Compressed Learning: A Deep Neural Network Approach," arXiv:1610.09615v1 [cs.CV]. 2016.
[19] Y. Xu and K. F. Kelly, "Compressed domain image classification using a multi-rate neural network," arXiv:1901.09983 [cs.CV]. 2019.
[20] Z. W. Wang, V. Vineet, F. Pittaluga, S. N. Sinha, O. Cossairt, and S. B. Kang, "Privacy-Preserving Action Recognition Using Coded Aperture Videos," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2019.
[21] H. Vargas, Y. Fonseca, and H. Arguello, "Object Detection on Compressive Measurements using Correlation Filters and Sparse Representation," 26th European Signal Processing Conference (EUSIPCO). 1960-1964, 2018.
[22] A. Değerli, S. Aslan, M. Yamac, B. Sankur, and M. Gabbouj, "Compressively Sensed Image Recognition," 7th European Workshop on Visual Information Processing (EUVIP), Tampere, 2018.
[23] P. Latorre-Carmona, V. J. Traver, J. S. Sánchez, and E. Tajahuerce, "Online reconstruction-free single-pixel image classification," Image and Vision Computing, 86, 2018.
[24] C. Li and W. Wang, "Detection and Tracking of Moving Targets for Thermal Infrared Video Sequences," Sensors, 18, 3944, 2018.
[25] Tan Y., Guo Y., Gao C., Tan Y., Guo Y., Gao C. Background subtraction based level sets for human segmentation in thermal infrared surveillance systems. Infrared Phys. Technol. 2013; 61: 230–240.
[26] A. Berg, J. Ahlberg, and M. Felsberg, "Channel Coded Distribution Field Tracking for Thermal Infrared Imagery," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; Las Vegas, NV, USA. pp. 1248–1256, 2016.
[27] C. Kwan, D. Gribben, B. Chou, B. Budavari, J. Larkin, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, "Real-Time and Deep Learning based Vehicle Detection and Classification using Pixel-Wise Code Exposure Measurements," Electronics, June 18, 2020.
[28] MOT Challenge, motchallenge.net/
[29] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arxiv, 2018.
[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems, 2015.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

[32] C. Kwan, L. Hagen, B. Chou, D. Perez, J. Li, Y. Shen, and K. Koperski, "Simple and effective cloud- and shadow-detection algorithms for Landsat and Worldview images," Signal, Image and Video Processing, 1-9, 2019.

Authors

Chiman Kwan received his Ph.D. degree in electrical engineering from the University of Texas at Arlington in 1993. He has written one book, four book chapters, 15 patents, 75 invention disclosures, 380 technical papers in journals and conferences, and 550 technical reports. Over the past 25 years, he has been the PI/Program Manager of over 120 diverse projects with total funding exceeding 36 million dollars. He is also the founder and Chief Technology Officer of Signal Processing, Inc. and Applied Research LLC. He received numerous awards from IEEE, NASA, and some other agencies and has given several keynote speeches in several international conferences.

David Gribben received his B.S. in Computer Science and Physics from McDaniel College, Maryland, USA, in 2015. He is a software engineer at ARLLC. He has been involved in diverse projects, including mission planning for UAVs, target detection and classification, and remote sensing.