Hue-saturation-value feature analysis for robust ground moving target tracking in color aerial video

Virgil E. Zetterlind III., Stephen M. Matechik

The MITRE Corporation, 348 Miracle Strip Pkwy Suite 1A, Ft Walton Beach, FL 32548

ABSTRACT

Ground moving target tracking in aerial video presents a difficult algorithmic challenge due to sensor platform motion, non-uniform scene illumination, and other extended operating conditions. Theoretically, trackers which operate on color video should have improved performance vs. monochromatic trackers by leveraging the additional intensity channels. In this work, ground moving targets in color video are characterized in the Hue-Saturation-Value (HSV) color space. Using segmented real aerial video, HSV statistics are measured for multiple vehicle and background types and evaluated for separability and invariance to illumination change, obscuration, and aspect change. HSV statistics are then calculated for moving targets from the same video segmented with existing color tracking algorithms to determine HSV feature robustness to noisy segmentation.

1. INTRODUCTION

UAV's have revolutionized modern warfare by providing warfighers the unprecedented ability to see the battlespace. Initially, it was the real-time tactical value of improved battlefield situational awareness that lured the services and government agencies into increased UAV deployments. UAVs range in size from small backpack portable systems, to systems such as Global Hawk, which has a wingspan of 116 feet, has a range of approximately 12,000 nautical miles, and can fly at altitudes up to 65,000 feet.[1]. As UAV deployments continue to proliferate, more and more agencies are recognizing the forensics value of motion imagery collected by these platforms and are seeking technology solutions for the exploitation of archived video.

Unfortunately, current video archive capabilities lag the insatiable need that intelligence analysts have as they try to assemble and analyze evidence, their mission, in part, fighting the Global War on Terror. Leveraging archived UAV video has proven more challenging due to the current limitations in context-driven archiving and retrieval systems for aerial video. Current archiving systems are generally limited to searches in time and geographic location. The granularity of these searches depends on the system in use, but can be as broad as an entire UAV mission. Ideally, a system should allow for frame-level results to limit the amount subsequent human analysis. Further limiting could be obtained by also detecting and classifying moving targets during the archive process. Development of aerial video trackers is an active research area. The more mature techniques, such as the Sarnoff tracker[2], are based on kinematic tracking and change detection. More recent techniques, such as those being developed under the DARPA Video Verification of ID (VIVID) program[3] combine kinematic methods with adaptive target modeling in terms of shape and color to improve performance and persistence. As these hybrid techniques are refined, the motion, shape, and color attributes measured in the tracking process can be incorporated as additional metadata in video archive and retrieval systems.

For this paper, we characterized the color statistics of moving vehicles in UAV imagery collected and released by the DARPA VIVID program to gain insight into color characterization methods for content and model-based archiving and retrieval. We used the Hue, Saturation, Value (HSV) color model for our statistics based on the desire to maintain validity for spectral characterization under changes in scene illumination, scintillation, and other difficult imaging conditions. The HSV parameters encode the spectral color (Hue), purity (Saturation), and intensity (Value) and have a mapping to and from RGB[4]. A cone is normally used to represent the HSV space. Hue is represented as an angle about the vertical axis of the cone with Red set to $0°$ rotating counter clockwise through Yellow, Green, Cyan, Blue, and Magenta and back to Red. Saturation is the ratio of the purity of a selected Hue to its maximum purity at S=1. This is plotted radialy outward from the vertical axis along the hue angle. Value is measured along the vertical axis with 0 set to the tip of the cone.

Statistics were collected on moving targets using hand-segmented tracking 'masks'. We also collected statistics on the overall scene to evaluate separability. To simulate the use of real trackers, we performed morphological dilation on the

truth segmentations and compared this to the background. Finally, we evaluated separability between moving targets within a scene using the Histogram Ratio Shift (HRS) filter provided in the CMU 'CTracker'[5] toolbox.

2. EXPERIMENT

2.1 Aerial Video Data

Our testing was based on the Eglin Public Datasets provided by the DARPA VIVID program and Carnegie Melon University[6] The dataset contains three scenes of moving vehicles taken from an aircraft with a color video camera. Each scene contains multiple moving targets and instances of like and dissimilar targets. The background environment also varies from relatively unobstructed runways to narrow roads with adjacent tree lines. Target motion was scripted to include cases of proximity, crossing, and passing amongst the target vehicles. Table 1 provides a high-level description of each scene and shows chips of each target.   Scenes are about 1800 frames in length.
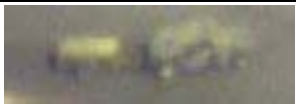
Scene Info	Frames	Targets
Elgin01 - No obscuration with vehicles driving on a paved runway. - Truthed target was Silver Car 2.	1820	Truck 1 Truck 2 Red Silver 1 **Silver 2** Blue Car
Eglin02 - No obscuration with vehicles driving on a paved runway. Vehicle groups cross in close proximity - Truthed target was Truck 1	1300	Blue Car Red Silver 1 Silver 2 **Truck 1** Truck 2
Eglin03 - No obscuration with vehicles driving on a paved runway. Groups cross in close proximity. - Truthed target was Jeep 1	2570	**Jeep 1** Jeep 2 Truck 1 Truck 2 Truck 3
Eglin04 - Light obscuration along tree line - Truthed target was Silver 1	1832	Blue Car **Silver 1** Silver 2 Truck 1 Truck 2
Eglin05 - Heavy obscuration along tree line - Truthed target was Truck	1763	Silver Car Blue Car **Truck**

Table 1: Scene and Target Summaries

Distributed with the datasets are CMU derived truth masks for 1 target per scene (shown in bold in Table 1). These masks appear to be hand generated and are very accurate. Figure 1 provides a typical example of one of these masks. A mask is provided once every 10 frames.

Figure 1: Truth mask for scene eglin01 frame 100

2.2 Experiments

For each scene, we conducted 3 experiments to extract target color statistics. The 1st experiment was based on the CMU truth masks and sought to quantify the temporal stability of our proposed color models under ideal segmentation. The 2nd experiment evaluated the stability of the proposed color models under imperfect segmentation by performing a morphological dilation on the truth masks and recalculating HSV statistics every 10 frames. Finally, we used 3 of the trackers implemented in the CMU tracking tool box version 2.2 to evaluate the HSV statistics of both target and confuser vehicles in each scene.

2.2.1 Truth Mask Processing

For each masked frame, we loaded the original image, converted it to HSV color space and calculated the overall image statistics as a baseline. Next, we used the binary image masks to extract the target chips, convert these values to HSV, and store them for further analysis. To simulate imperfect segmentation, we performed a morphological dilation of each truth mask using a disc structuring element. HSV statistics were collected for each masked image using a disc radius of 5, 15, and 45 pixels. Figure 2 shows examples of the resulting target chips for each dilation level.

5 pixel dilation

15 pixel dilation

45 pixel dilation
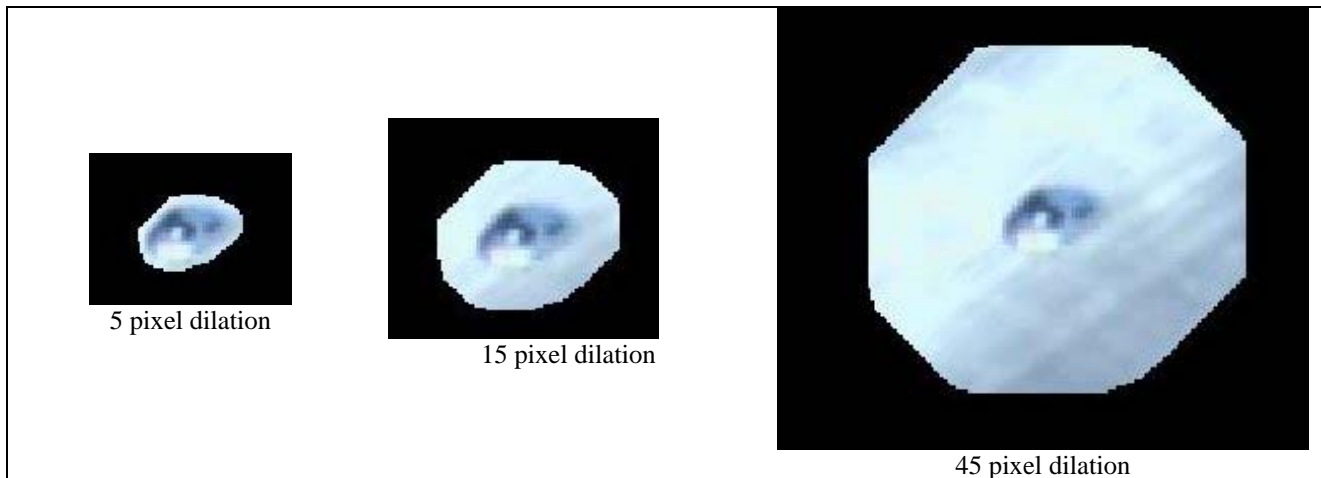
Figure 2: Target chips given 3 different dilation levels

2.2.2 Histogram Ratio Shift Tracker Processing

Tracking data was collected for 3-5 targets per scene using the Histogram Ratio Shift (HRS)[6] tracker implemented in the CMU Vivid Tracking Toolbox CTracker version 2.2. The tracker was stopped and restarted as necessary to maintain track through the majority of a scene. The HRS tracker generates a target mask file for each frame. These mask files were used to extract HSV statistics for each tracked target for our analysis.

3. RESULTS

3.1 Target vs. Background

Figure 3 shows a plot of the mean H, S, and V values for the truth mask and whole frame for scene eglin01. Error bars represent the 1 sigma values at each frame. For this scene, the tracked target had very similar Hue to the background runway but was well discriminated in Saturation and somewhat discriminated in Value. This behavior generally held in eglin01, eglin02, eglin04, and eglin05 which had civilian vehicles. In these scenes, large standard deviations in target Saturation and Value were expected due to specular paint and daylight viewing conditions. The natural backgrounds in these scenes were much more uniform in Saturation. The mean Value for both target and background trended based on overall scene illumination when vehicle aspect was relatively constant. Under strong lighting, rapid fluctuations in the target Saturation and Value correlated well with target aspect changes and relative orientation to the sun.

For the dilated truth masks, the relative difference between target and background HSV statistics across all three channels was greatly reduced. Figure 4 shows the 45 pixel dilation case for eglin01. Comparing this to the truth case in Figure 3, it is clear that the variance of the Saturation and Value has been reduced by the relatively uniform background. Further, the mean of the target Saturation has been pulled towards the background value. The Value statistics are actually more separated for this case due to the differences in the brightness of the runway along the target path vs. the overall runway brightness.

Eglin03 contained military vehicles driving on an abandoned runway. HSV statistics where somewhat different for this case as shown in Figure 5. Here there was somewhat better discrimination in Hue, but much less in Saturation and Value. Had this scene been run in more natural terrain, the discrimination would be poorer yet as the difference in Hue and background brightness would likely decrease further.

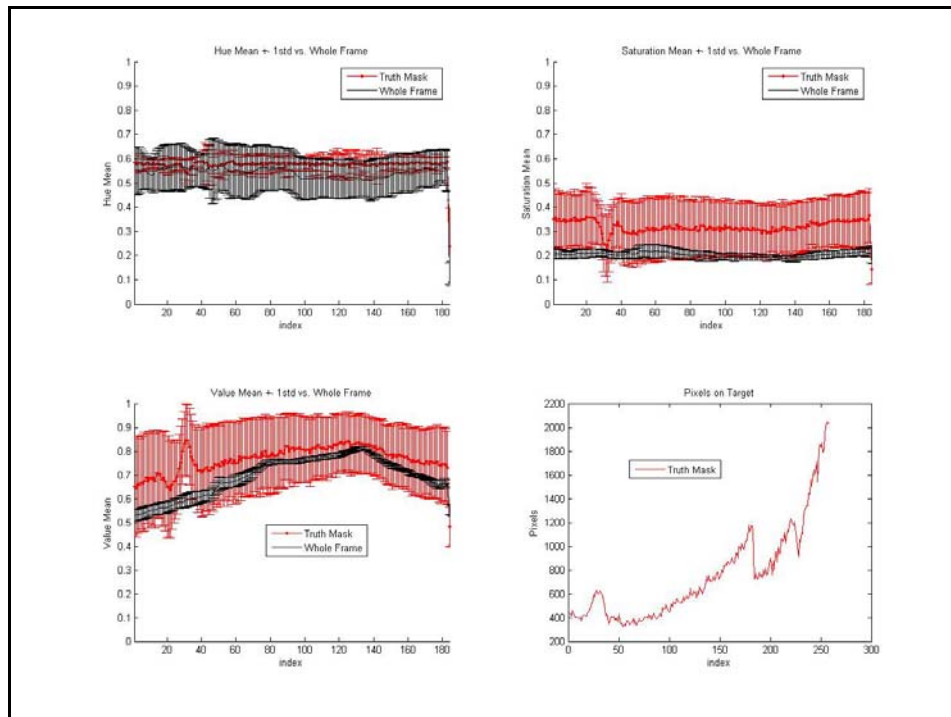

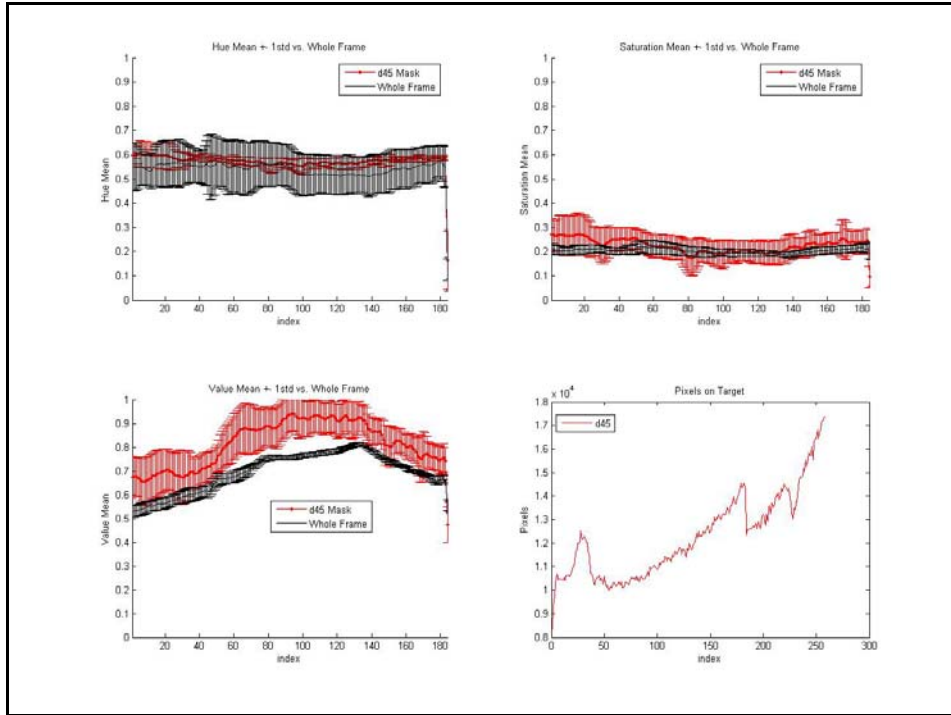**Figure 3: H, S, V, and pixels on target plot for eglin01 truth**

Figure 4: H,S,V, and pixels on target plot for eglin01 with 45 pixel dilation
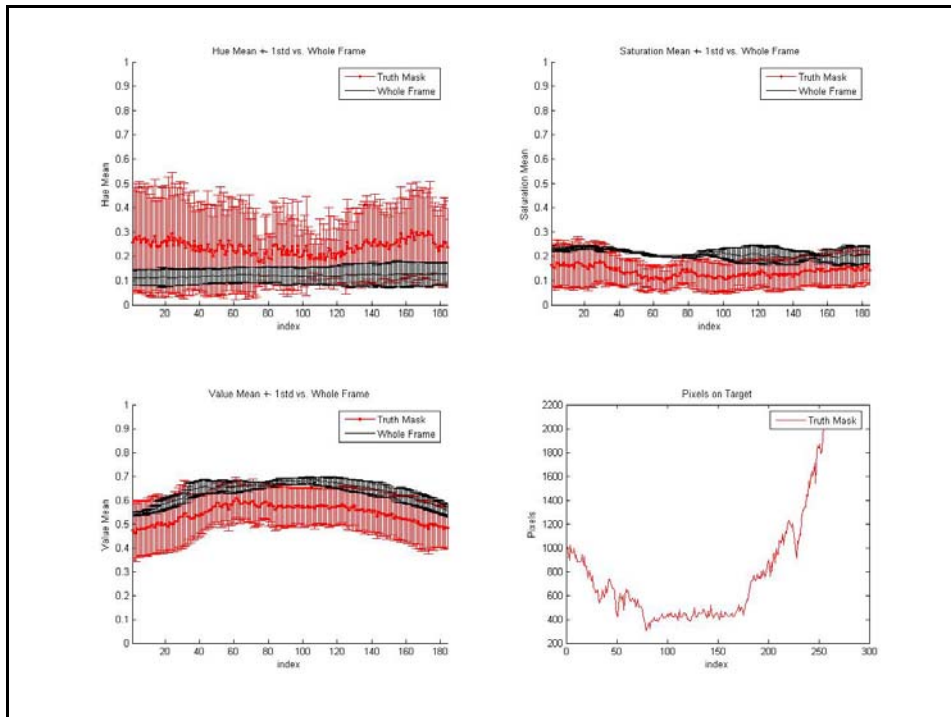

Figure 5: H,S,V, and pixels on target plot for eglin03 which contained military vehicles

3.2 Target vs. Confuser

Figures Figure 6 – Figure 10 provide scatter plots for the HSV values for each tracked target in the 5 scenes. The plots were constructed using 10 track masks for each target evenly sampled over 100 frames. To improve visibility, only those points within 1 standard deviation of the H, S, and V mean are plotted. For each figure, the plot on the left shows the Hue distribution as seen looking 'down' the HSV cone. The zero degree axis is horizontal to the right of the origin and represents reds. Distance from the origin represents Saturation. The right plot shows a side view of the HSV cone with Value running along the vertical and Saturation outward in the horizontal from the origin. The orientation for each side view plot was selected to maximize visibility between targets.

As illustrated in Figure 6, the vehicles in scene eglin01 are quite similar, except for the red convertible. The identical silver cars (the light blue and black dots) essentially overlap in the HSV space while the 2 trucks are distinguished by differences in Saturation and Value extent. The dark blue car is surprisingly similar to the trucks, but is more saturated.
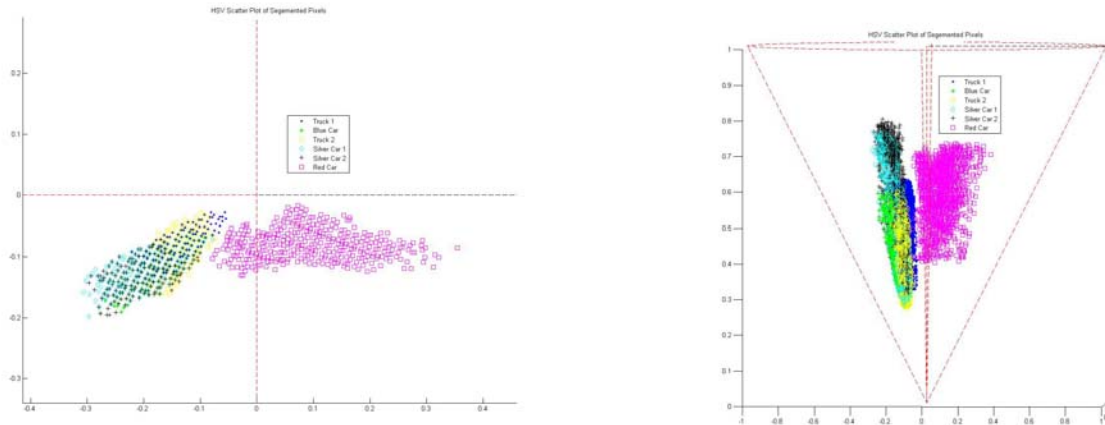
Figure 6: HSV scatter plots for scene eglin01

We tracked truck 1, silver car 1, and the red car for scene eglin02 as shown in Figure 7. Here the vehicles showed a wider difference in statistics in part because lighting conditions did not appear to be as harsh or perhaps the sensor exposure control was better. Once again, the red car is easily distinguished relative to the silver truck and car. Differences in S and V are also distinct for the silver car and truck even though their distributions overlap.

Figure 7: HSV scatter plots for scene eglin02

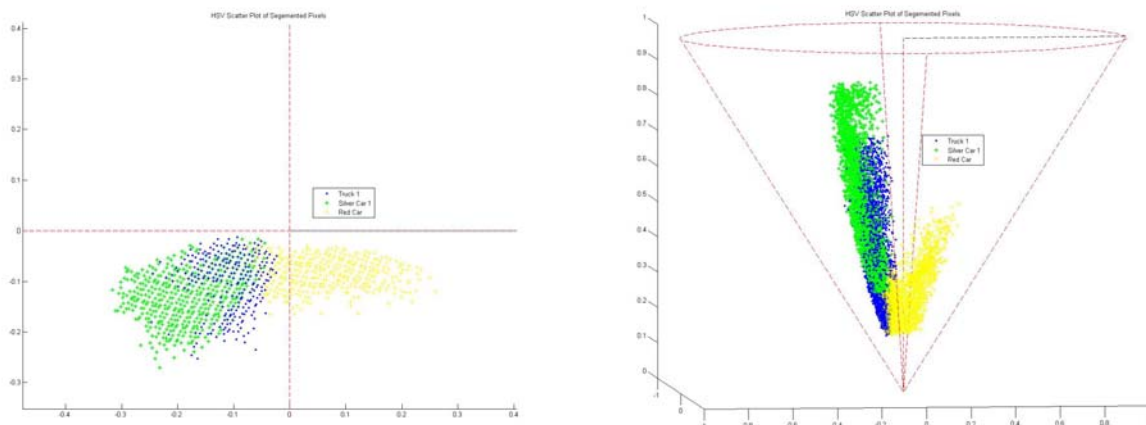Scene eglin03 contained a collection of military trucks and jeeps. We tracked trucks 1 through 3 and jeep 1. Truck 2 is clearly distinct in Hue with the light green paint. The other vehicles exhibit significant overlap in their distributions. The HSV distribution for truck 1 (as seen in Figure 8) shows the contribution of shadow pixels to the HSV statistics as the lower-left quadrant pixels in the top-down plot are contributions from the shadowed underside of the truck.
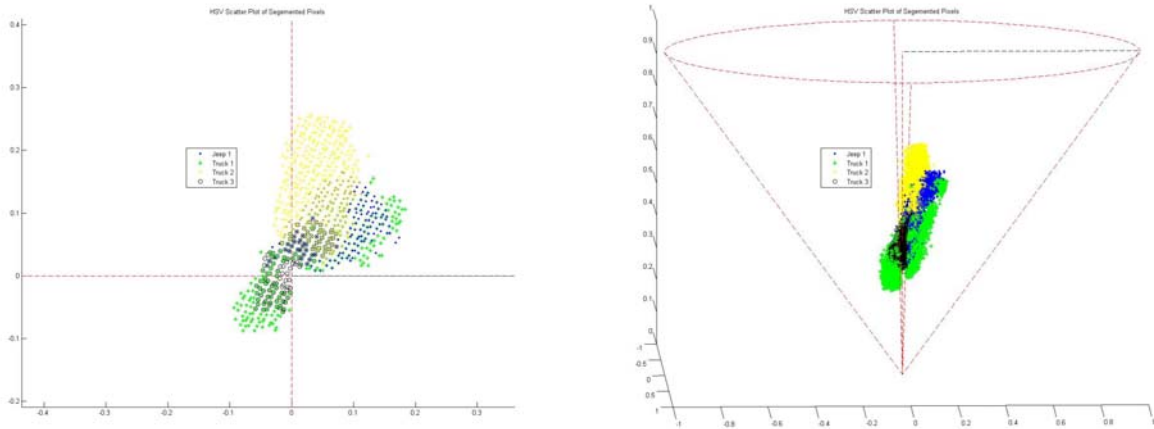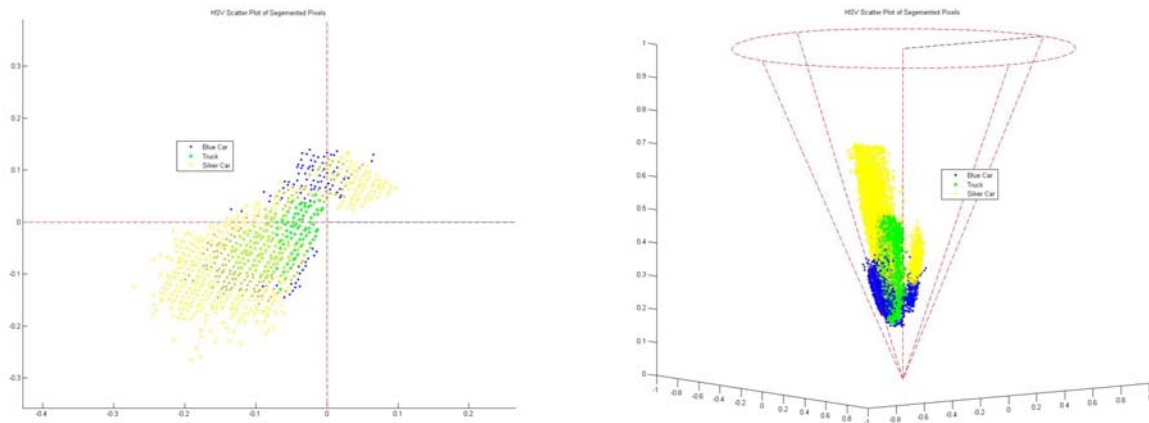

Figure 8: HSV scatter plots for scene eglin03

Of the 5 scenes, eglin04 had some of the toughest tracking conditions due to small targets and poor image exposure control. As seen in Figure 9, HSV distributions were very high on the Value scale and all vehicles had similar color. This scene was also the 1st to have obscuration in the form of adjacent treelines. The obscuration effect is seen in the upper-right quadrant pixels in the top-down plot for car silver 1 and the blue car. These pixels are include many green pixels from the treeline which the tracker included in the masks. The blue car is the most distinguishable from the HSV statistics as its brightness level was lower than the metallic colors in the silver cars and truck.
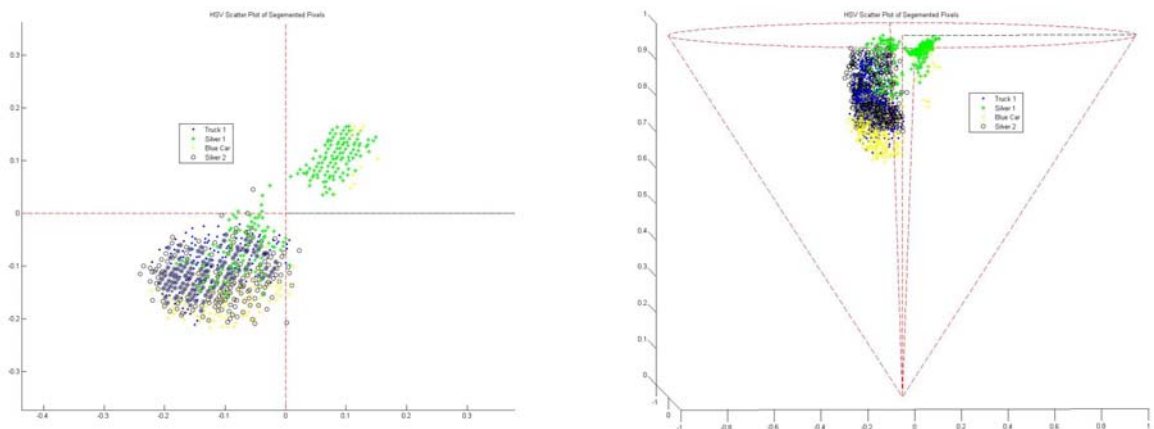
Figure 9: HSV scatter plots for scene eglin04

Finally, in scene eglin05 we saw more even lighting conditions with good pixels on target. In this case, while vehicles had a high degree of overlap in Hue, they were more readily separable based on Value and Saturation as seen in Figure 10. This scene showed the potential of HSV target modeling when given a reasonable number of target pixels (at least 2000 on average) and good exposure control of the sensor.

Figure 10: HSV scatter plots for scene eglin05

4. CONCLUSION AND FUTURE WORK

This research provides an initial characterization of HSV color statistics for typical ground targets imaged in color UAV video. As expected, civilian vehicles are most easily distinguished from natural backgrounds due to large variations in Saturation and Value relative to natural materials. They are sensitive though to the degree of correct segmentation and segmentation errors can quickly reduce the separability of the background and target statistics. For well segmented targets, rapid changes in Value or Saturation were good predictors of aspect change.

Target vs. confuser separation showed the advantage of obtaining more pixels on target whenever possible. Since this test was conducted using a real tracker, larger targets generally meant better initial segmentation from background and a lower overall contribution of error pixels into the HSV distribution. In terms of a content-based archive and retrieval system, it makes sense to measure target color characteristics during periods of maximum zoom during a track and weight this as part of the track query mechanism. As we implement color features into our archive, we will also begin to consider query methods for target color using queries closer to natural language[7].

We did not discuss any image preprocessing or inclusion of motion information to improve performance. These are areas of active research. One simple extension that might improve the HSV statistics from the real tracker would be to perform a small dilation ~5px on the output mask to help fill in missed pixels on the target. We found that the CMU tracker often created sparse masks on the small targets which emphasized the shadow and highlight features. A small dilation around these would capture more target pixels and likely improve the accuracy of the HSV statistics.

While simplistic, the HSV statistics covered here can provide important additional information within the context of a broader sensor exploitation system. Combined with motion and other information, they can improve overall confidence in correct ID and tracking of targets in cluttered environments. They also provide additional search parameters for a video archive system oriented to aerial video collections from UAVs.

REFERENCES

1. www.af.mil/factsheets/factsheet_print.asp?fsID=175&page=1
2. Kumar, Raeksh, et al, "Aerial Video Surveillance and Exploitation," Proc. of the IEEE, **89**(10), October 2001, pp. 1518-1539.
3. Arambel, Pablo, et al, "Performance Assessment of a Video-based Air-to-ground Multiple Target Tracker with Dynamic Sensor Control," Proc. of SPIE **5809**, 2005, pp.123-134.
4. Hearn, Donald, and Baker, M., *Computer Graphics*, Prentice Hall, New Jersey, 1994, pp. 575-576.
5. www.vividevaluation.ri.cmu.edu

6. Collins, Robert T., Zhou, Xuhui, and Teh, Seng Keat, "An Open Source Tracking Testbed and Evaluation Web Site," IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance, January, 2005.
7. Mojsilovic, Aleksandra, "A Computational Model for Color Naming and Describing Color Composition of Images," IEEE Trans. on Image Proc. **14**(5), May 2005, pp. 690-699.