VALIDATION OF VEHICLE CANDIDATE AREAS IN AERIAL IMAGES USING COLOR CO-OCCURRENCE HISTOGRAMS

Winfried Leister [a,b,*], Sebastian Tuermer [a,c], Peter Reinartz [a], Karl Heinz Hoffmann [b], Uwe Stilla [c]

[a] Remote Sensing Technology Institute, German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
{sebastian.tuermer, peter.reinartz}@dlr.de
[b] Institut fuer Physik, Technische Universitaet Chemnitz
09107 Chemnitz, Germany
winfried.leister@s2005.tu-chemnitz.de, hoffmann@physik.tu-chemnitz.de
[c] Photogrammetry and Remote Sensing, Technische Universitaet Muenchen (TUM)
80290 Munich, Germany
stilla@tum.de

Commission III/VII

KEY WORDS: vehicles, aerial imagery, traffic monitoring, color co-occurrence histograms, 3K+ camera system

ABSTRACT:

Traffic monitoring plays an important role in transportation management. In addition, airborne acquisition enables a flexible and real-time mapping for special traffic situations e.g. mass events and disasters. Also the automatic extraction of vehicles from aerial imagery is a common application. However, many approaches focus on the target object only. As an extension to previously developed car detection techniques, a validation scheme is presented. The focus is on exploiting the background of the vehicle candidates as well as their color properties in the HSV color space. Therefore, texture of the vehicle background is described by color co-occurrence histograms. From all resulting histograms a likelihood function is calculated giving a quantity value to indicate whether the vehicle candidate is correctly classified. Only a few robust parameters have to be determined. Finally, the strategy is tested with a dataset of dense urban areas from the inner city of Munich, Germany. First results show that certain regions which are often responsible for false positive detections, such as vegetation or road markings, can be excluded successfully.

1 INTRODUCTION

Within the last years smart routing of vehicles has become a highly acclaimed topic (Schofield, 2009). The objective is not only to increase efficiency and sustainability but also to achieve a maximum of personal comfort. The current development was made possible due to further developments in traffic acquisition techniques. A part of routing is the determination of a vehicle's position as well as the knowledge of whether a road is clear or not. This question can have even more impact when we talk about routing of rescue crews or emergency vehicles.

A common method to obtain the position of a vehicle is utilizing a global navigation satellite system (GNSS). Often it is combined with a system that collects data of all traffic participants registered in a certain program. Exemplary programs are operated by Google and its Android OS or the dutch navigation device manufacturer TomTom. However, this technique works only for cars which are in use. Once the driver has left the car and the engine is turned of no further signal is transmitted and the position of the car is uncertain. A statement whether a road is trafficable cannot be made without new cars frequently passing this section.

Of course, information of stationary video cameras can be used to address that problem (Zhou et al., 2007). However, stationary video cameras provide local information, usually from bigger roads or traffic system relevant spots only. In contrast, remote sensing enables gathering data within a wide area (Hinz et al., 2006). One choice could be satellite images (Leitloff et al., 2010) but they are not flexible enough due to fixed revisiting times. A better choice for real-time applications is the use of aerial images.

*Corresponding author.

Several methods have been developed where moving cars are recognized by airborne optical sensors (Cao et al., 2012, Cheng et al., 2012, Kirchhof and Stilla, 2006). However, in the case of the above described problem they are not helpful. Instead, approaches which focus on standing or parking vehicles are superior in that case. Research works within that field can be grouped according to their utilized features.

Many techniques of image recognition are based on gradients. These can be subdivided into 3d car models based on significant edges (Hinz, 2004), the surrounding contour (Kozempel and Reulke, 2009) or histograms of oriented gradients in combination with other high-level features (Kembhavi et al., 2011). A further class includes approaches utilizing algorithms which are region-based and try to see a car as an object or a blob (Holt et al., 2009, Lenhart et al., 2008).

Most methods have one thing in common. The main focus is on the car itself, and the background of the vehicles is only seldom exploited. Inspired by publications that show how important such information can be (Heitz and Koller, 2008, Divvala et al., 2009), the target of our research is to better incorporate the background in detection strategies for standing or parking vehicles.

In this paper a method is presented to validate vehicle detections based on their background and color information. The aim is to classify the foreground in a patch as one of the categories (car, street, vegetation). Patches are obtained either by a sliding window technique or by some other pre-classifications. Such patches of vehicle candidates and their surroundings are transformed to the HSV color space. The hypothetical background is separated from the foreground by applying a coarse car mask to the patches. Its size is based on an average car size. The orientation of the

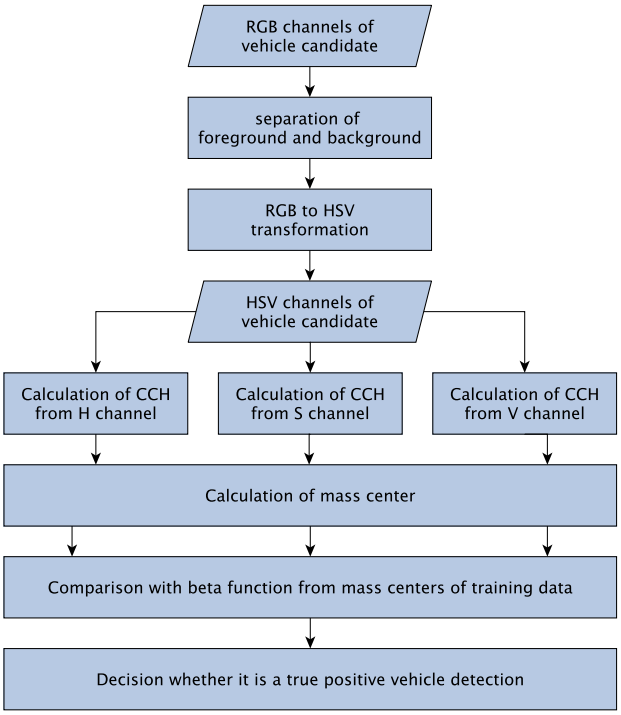


Figure 1: Workflow of the determination of areas where vehicles likely occur. The scheme is for one patch of the image only, all other patches are examined using a sliding window approach.

mask is derived from the known driving directions of relevant road segments which can be obtained from road databases. Then, color co-occurrence histograms of the remaining background are calculated for each color channel. In a final step, the mean values of the histograms are used to estimate whether requirements of a typical car are met or not. The evaluation is done with a beta function that is determined by a maximum likelihood algorithm and training samples.

2 METHOD

An overview of the proposed method for vehicle candidate validation is shown in Fig 1.

2.1 Separation of foreground and background

At first foreground and background of each car candidate are examined separately. However, the technique is identical for foreground and background areas. At the beginning all candidates are represented by images of 45×45 pixels in the RGB color space.

In order to apply a mask the orientation of the vehicles has to be known. Since the presented approach is planned to act as validation method, the orientation can be obtained from the preceding detection algorithm. Alternatively, road databases (e.g. OpenStreetMap) can be used to determine the potential driving direction of the cars. Additionally, it is assumed that a car is in the center of the examined image patch.

The principle of the fore- and background mask is depicted in Fig. 2. The size of the foreground mask is determined by the average size of a car. It is derived from the dimensions of 30 training cars. However, pixels close to the contour of the cars have been ignored. The reason is that often artifacts occur at these positions due to shadow. The major objective of the presented method is to get statistical information about color, gray value, lightness of the car and thus artifacts could adulterate the statistics.

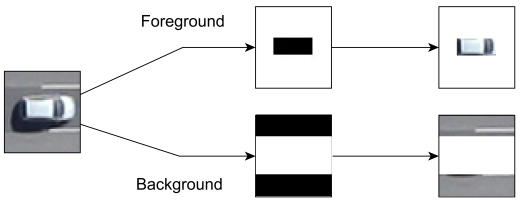


Figure 2: Foreground (potential car candidate) and background (potential street) can be set as a fixed area since the orientation of the cars is assumed to be known. The foreground is in the center of the patch while the background is restricted to the areas to the left and right of the car.

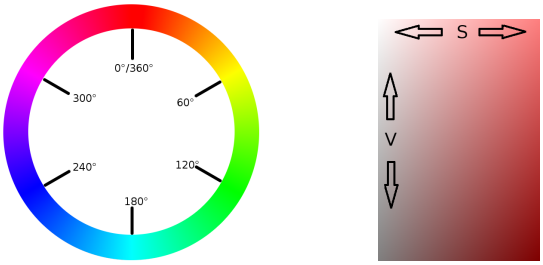


Figure 3: H channel as circle and S-V plane with sample color red $H = 0$.

The background area is represented by the remaining rectangular areas to the left and the right side of the cars. The area in front and behind a car is not used in the further process because often cars are parked in a row and other cars which disturb the process can be found at these positions.

2.2 Transformation to the HSV Color Space

After the separation a transformation into the HSV color space is performed. From that moment on the color information and the intensity can be accessed independently.

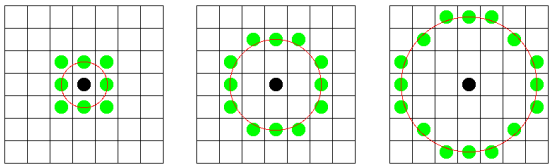
A special property of the HSV color space is the necessity of only one channel to define the color value. The H (hue) channel is not expressed as a straight line but a circle. Additionally, S (saturation) and V (value) channels are part of the HSV color space. How these channels are incorporated can be seen in Fig. 3.

The values of the S and the V channel are ranging between zero and a certain maximum value. While the Saturation expresses how much color belongs to a pixel (0 = no color, $\max$ = full color), the Value expresses the brightness of the pixel.

The transformation from RGB to HSV color space can be found in (Gonzales and Woods, 1996). After determination of the MAX and MIN values using Eq. 1 the dominant color has to be determined using Eq. 2.

$$\text{MAX} := \max(R, G, B) \qquad \text{MIN} := \min(R, G, B) \tag{1}$$

$$H := \begin{cases} 0, & |\text{MAX} = \text{MIN} \\ & \Leftrightarrow R = G = B \\ 60° \cdot \left(0 + \frac{G-B}{\text{MAX}-\text{MIN}}\right) & |\text{MAX} = R \\ 60° \cdot \left(2 + \frac{B-R}{\text{MAX}-\text{MIN}}\right) & |\text{MAX} = G \\ 60° \cdot \left(4 + \frac{R-G}{\text{MAX}-\text{MIN}}\right) & |\text{MAX} = B \end{cases} \tag{2}$$



$d = 1$ $d = 2$ $d = 3$

Figure 4: Example of a circular symmetric structure of neighborhood. This technique is used here.

The saturation S is expressed as the rescaled difference of MAX and MIN (Eq. 3), while V results directly from MAX (Eq. 4).

$$S := \begin{cases} 0, & |MAX = 0 \Leftrightarrow R = G = B = 0 \\ \frac{\text{MAX-MIN}}{\text{MAX}} & \text{else} \end{cases} \quad (3)$$

$$V := MAX \quad (4)$$

2.3 Calculation of the color co-occurrence histograms

Subsequently, color co-occurrence histograms (CCH) are calculated. Co-occurrence histograms are based on the relation of neighboring pixels and give a statement about the properties of the texture. Two ways of calculating CCHs are presented by (Liu et al., 2012). The procedure where only one direct pixel neighbor is considered has the disadvantage that rotation of the image can lead to errors. If there is a slightly rotated image the new CCH will partly differ from its former version. This problem can be avoided by using the distances between color values for the calculation instead of the orientations. There are two known alternative ways in the literature (Liu et al., 2012).

Both variants have in common that two pixels with coordinates (x_1, y_1) and (x_2, y_2) in a certain distance d are compared. The first variant uses an accumulation of neighboring pixels which are in a square around the origin pixel. The edge length of the square is $2d + 1$. The second variant, utilized for this approach, compares all pixels on a circle having the same Euclidean distance. A graphical explanation can be seen in Fig. 4. The benefit is that equal CCHs are generated also for rotated images. However, as coordinates in an image have integer values, we use a discretized distance:

$$d = \text{round}\left(\sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)}\right) \quad (5)$$

where $(x_1, y_1), (x_2, y_2)$ represent coordinate pairs of two pixels which should be compared.

In the case of original CCHs, only pairs of pixels with equal intensity I are used. Thus, the calculation of the frequencies h in a CCH can be mathematical written as (Eq. 6).

$$h(I) = \sum_x \sum_y \begin{cases} 1 & | \; I(x_1, y_1) = I(x_2, y_2) \cap d = 1 \\ 0 & \text{else} \end{cases} \quad (6)$$

The distance $d = 1$ has been chosen because experiments have not shown significant differences between $d = [1, 3, 5, 10]$. On the other hand with $d = 1$ the fewest comparisons need to be done and the calculation time can be reduced. The calculation of the CCHs is done separately for each HSV channel for foreground and for background. It results in six histograms per candidate.

category	quantity
cars	557
street	1995
vegetation	12834

Table 1: Number of training candidates for the evaluation of the mean value distribution.

2.4 Likelihood decision

A closer empirical examination of the histograms shows that often an accumulation around a certain value occurs. Hence, the implication is that the mean of the histogram is able to provide the core information. To this end, the mean of a CCH is calculated as shown in Eq. 7.

$$m = \frac{1}{\sum_I h(I)} \sum_I h(I) \cdot I \quad (7)$$

where $h(I)$ is the I-th value in the CCH.

A training set consisting of samples from all used classes is necessary for the following process. The reference data is essential for a correct classification of the candidates and the possession of an appropriate large dataset of reference data is recommended, in order to be able to make a significant statement. Furthermore, the orientation of the candidates is known and thus the mask could be optimally rotated. The utilized training samples resulted from a large set of forest areas and free highways (see Tab. 1).

Calculating the mean values of these training candidates leads to characteristic histograms for all three categories. Finally, we obtain three CCHs from every candidate and out of it three means (m_H, m_S, m_V). Subsequently, every value is compared with the values of the corresponding histogram of the three classes which we calculated from the training data. For example, the process is as follows. m_H is compared to the values of the hue-histogram of cars, roads and vegetation. We take the three corresponding (i. e. $m_\text{H} \rightarrow [h_\text{car}(m_\text{H}), h_\text{str}(m_\text{H}), h_\text{veg}(m_\text{H})]$) values $h_\text{cat}(m)$ and compare them with each other. The nine quantities ($q_\text{H,car}$, $q_\text{H,str}$, …, $q_\text{V,veg}$) stating to which distribution the mean value of a candidate belongs to are calculated using Eq. 8.

$$q_\text{chan,cat}(m_\text{chan}) = \frac{h_\text{chan,cat}(m_\text{chan})}{\sum_\text{cat} h_\text{chan,cat}(m_\text{chan})} \quad (8)$$

In the next step we multiply the quantities of the same category to get a combined value (Eq. 9).

$$k_\text{cat} = \prod_\text{chan} q_\text{chan,cat}(m_\text{chan}) \quad (9)$$

This gives us three values, named k_car, k_str and k_veg, describing the frequencies for the examined area of being car, road or vegetation. The new values k_cat are then assumed to be directly correlated to the likelihood of being such a candidate.

Based on these six (2×3) k_cat values of the foreground and the background, a decision can be made whether a candidate is a car or not. We can compare these values to each other and find scenarios which mostly show cars on streets, pure streets or others.

The following three rules describe conditions when the candidate is supposed to belong to the no car class:

- Road in foreground:
  $k_{\text{str}}(\text{foreground}) > k_{\text{veg}}(\text{foreground}) \quad \wedge$
  $k_{\text{str}}(\text{foreground}) > k_{\text{car}}(\text{foreground})$
- Vegetation in background:
  $k_{\text{veg}}(\text{background}) > k_{\text{str}}(\text{background}) \quad \wedge$
  $k_{\text{veg}}(\text{background}) > k_{\text{car}}(\text{background})$
- A too small difference between foreground and background:
  $\sum_{\text{chan}} |m_{\text{chan}}(\text{foreground}) \quad -$
  $m_{\text{chan}}(\text{background})| \leq \text{threshold}$

The threshold can be specified dependent on different light conditions and sensor properties. In our experiments, the threshold ranged from 10 to 15. When the difference was too low, foreground and background were the same category.

As an extension to this method, beta functions are calculated from the histograms of the training data using the maximum likelihood method. The candidates are than compared to these beta functions. The benefit is to obtain a continuous distribution. The experimental results in Chapter 4 is based on the distribution of the beta function.

3 CAMERA SYSTEM

The used aerial test data are acquired by the 3K+ camera system, which is composed of three off-the-shelf professional SLR digital cameras (Canon EOS 1Ds Mark III with Zeiss lenses). The nominal focal length for the 3K+ system is 50 mm. These cameras are mounted on a platform which is specifically constructed for this purpose. A picture of the cameras and the platform is shown in Fig. 5. Furthermore, a calibration was done to enable the georeferencing process. The system is designed to deliver images with a maximum recording frequency of 3 Hz. The Mark III camera delivers 21.0 MPix. Depending on the flight altitude a spatial resolution up to 13 centimeters (at 1000m altitude and nadir) is provided. For further information about the 3K+ camera system please refer to (Kurz et al., 2012).

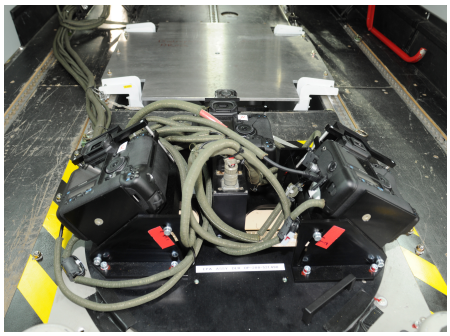


Figure 5: The 3K+ camera system. A low-cost camera system with a higher recording frequency compared to professional photographic camera systems.

4 EXPERIMENTAL RESULTS

The evaluation of the test candidates from Tab. 1 resulted in characteristic histograms for vehicle, road and vegetation. It is easy to see in Fig. 6 that many cars have a dominant blue color (Means at ≈ 60). The color of roads shows a wider spread of the distribution (Fig. 7). Most of the mean values have been in the range of 60 to 90. In contrast, vegetation mainly shows color values in the range of 25 to 55 (Fig. 8). Additionally, also clear differences in the two other channels Value and Saturation are present.

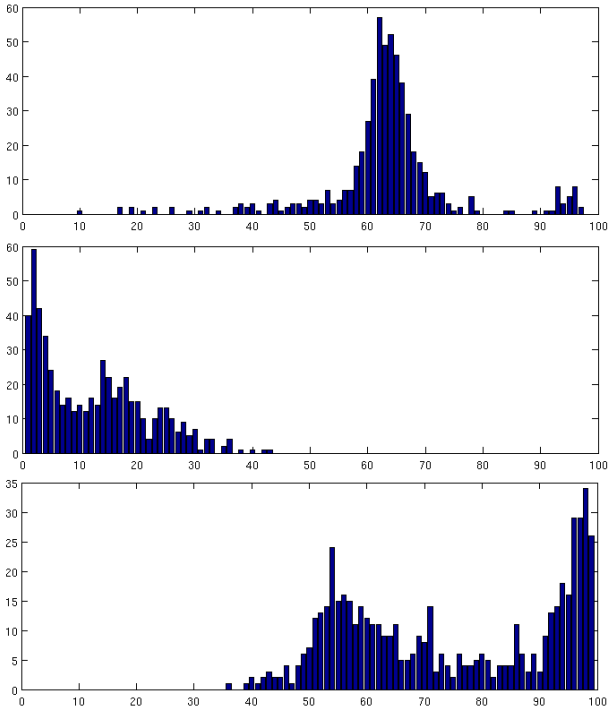


Figure 6: Histograms of the CCH mean values from 557 cars in the HSV color space. From top to bottom: H, S, V.

Class	Level of color space	a	b
Car	H	7.74	4.45
	S	1.39	9.72
	V	2.96	0.95
Road	H	4.26	2.15
	S	2.31	39.6
	V	8.47	5.13
Vegetation	H	8.59	12.43
	S	4.82	11.02
	V	3.37	9.68

Table 2: a and b of beta distribution for histograms of car, road and vegetation.

Moreover, for each of the nine histograms a beta function is calculated with the Maximum-Likelihood-method. The usage of the beta function approach is preferred over the histograms because some frequencies in the reference histograms are zero and no result could be calculated at those positions.

The calculated beta functions are presented in Tab. 2. Many histograms show more than one maximum and have to be taken with caution therefore. Nevertheless, it is shown that these values also allow reasonable decisions. The following exemplary result is based on the continuous distribution of the beta function.

Finally, the car validation strategy is tested with images from the previously described camera system. A typical car fits in a square area of 45×45 due to an image resolution of 13 cm. Consequently, the area of the foreground mask is 27×11 pixels in the center of the patch. The final result, illustrated in Fig. 9 (b), shows the image from Fig. 9 (a) classified as car, as road or as vegetation which is coded in red, in blue or in green color, respectively. The colors are mixed in areas where the decision was not clear. For instance, the yellow areas are a mixture of vegetation and road (green and blue). Target objects are vertically oriented cars.

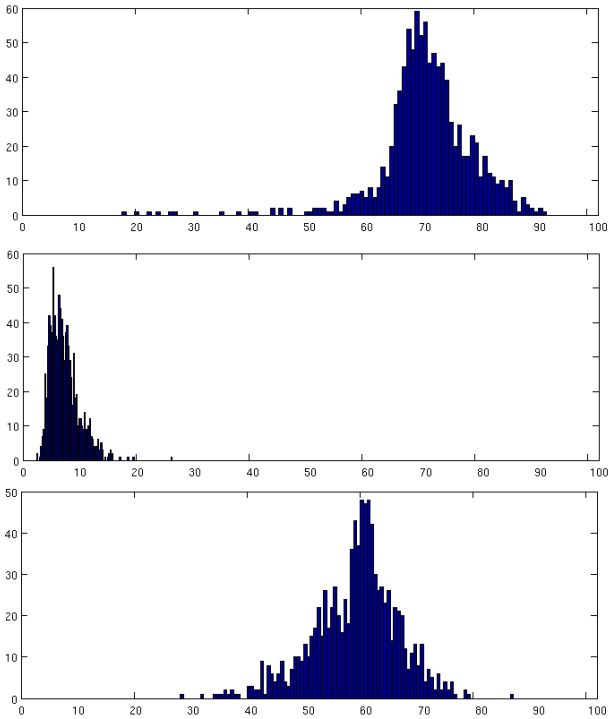


Figure 7: Histograms of the CCH mean values from 1995 road patches in the HSV color space. From top to bottom: H, S, V.

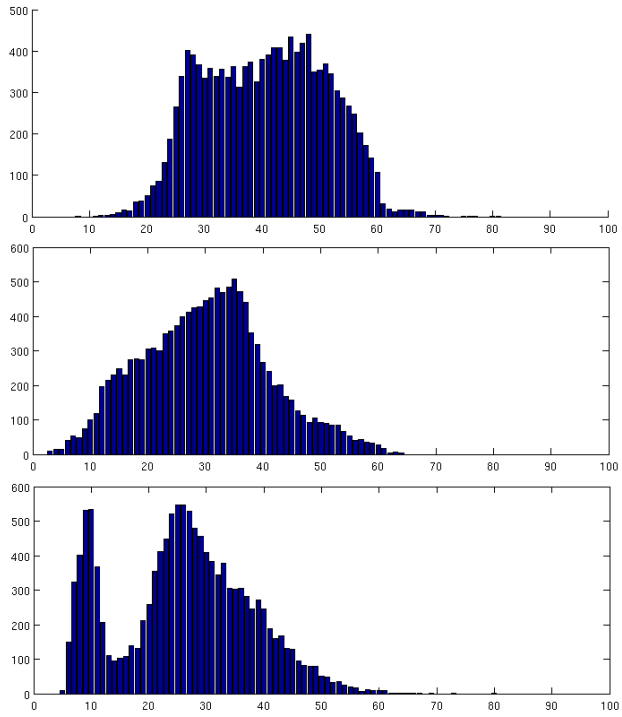
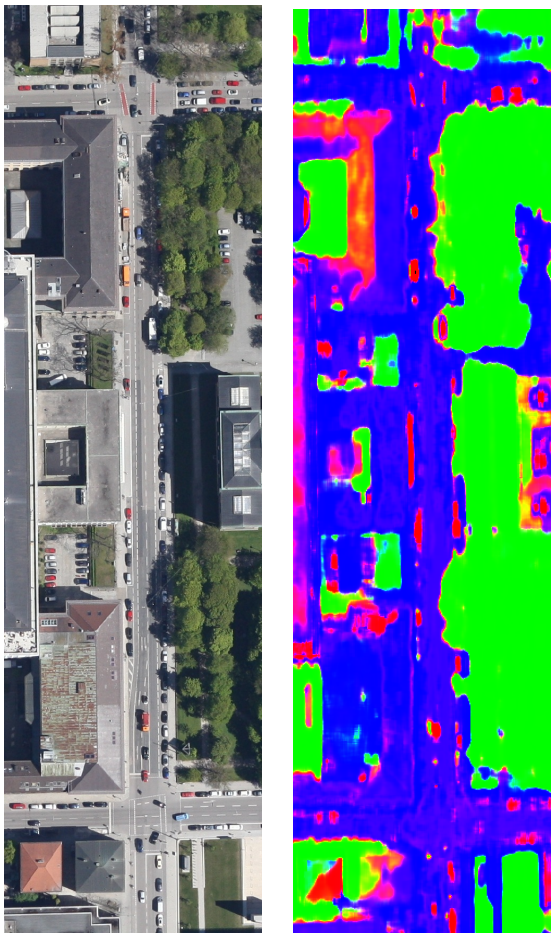


Figure 8: Histograms of the CCH mean values from 12834 vegetation patches in the HSV color space. From top to bottom: H, S, V.



(a) (b)

Figure 9: When the algorithm is applied to Fig. 9 (a) we receive a result as it is shown in Fig. 9 (b). The likelihoods k_{cand} have been mapped to the RGB color space with red, green, blue for k_{car}, $k_{\text{Vegetation}}$, k_{Road}, respectively. Other colors are a mixture of these base colors and state that the decision was not clear. Aimed objects were cars oriented in vertical direction.

5 DISCUSSION

Generally, the objective was to find a correlation between certain color features of car, road, vegetation and features only based on intensity of the gray values. It could be observed that bright green regions are often classified as cars but dark green areas are often classified as vegetation. Moreover, The CCHs of buildings have not been calculated which could be a reason why they are often associated with other categories (Fig. 9). However, the presented method is only thought as a coarse segmentation but it is not suitable for precise detection of cars.

Strengths and weaknesses of the presented strategy are shown in Fig. 9. Large areas of vegetation and roads are almost correctly classified. However, shadows are very often associated as areas of vegetation. A possible reason is the lack of training data from those areas. Also bike paths are sometimes classified as cars (see crossing in the upper part of the image) due to the rectangular contour and the width which is similar to the width of a car. The same problem occurs in the center of the image where a sidewalk does have a different length, but its width to one of the cars is very similar. Additionally, colors on the left and the right side of this sidewalk are different which could be an indicator for a car.

6 CONCLUSIONS AND FUTURE WORK

The suggested method is an approach to better investigate color in regard to car detection from aerial imagery which is a relatively seldom illuminated subfield. However, the algorithm is in an early stage and many extensions are possible. The primary reason to start with this approach was to separately observe color values in-depth instead of just passing all car features combined to a machine learning algorithm.

Furthermore, many problems are still present and further questions arose during these investigations. For instance, how sensitive are color values under changing illumination like on cloudy or sunny days? Or, how stable are color values as feature when different seasons have to be expected and snow is everywhere around in winter time or trees are without green leafs?

In this sense the method showed its suitability, especially big roads and large vegetation areas could be classified correctly. However, detailed examination of candidates is only possible if an already pre-classified group of candidates is provided. Then it can be used to extract false positives.

REFERENCES

Cao, X., Lin, R., Yan, P. and Li, X., 2012. Visual attention accelerated vehicle detection in low-altitude airborne video of urban environment. IEEE Transactions on Circuits And Systems for Video Technology 22(3), pp. 366–378.

Cheng, H.-Y., Weng, C.-C. and Chen, Y.-Y., 2012. Vehicle detection in aerial surveillance using dynamic bayesian networks. IEEE Transactions on Image Processing 21(4), pp. 2152–2159.

Divvala, S., Hoiem, D., Hays, J., Efros, A. and Hebert, M., 2009. An empirical study of context in object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Gonzales and Woods, 1996. Digital image processing. Reading, Massachusetts: Addison-Wesley.

Heitz, G. and Koller, D., 2008. Learning spatial context: Using stuff to find things. In: Computer Vision ECCV 2008, pp. 30–43.

Hinz, S., 2004. Detection of vehicles and vehicle queues in high resolution aerial images. Photogrammetrie - Fernerkundung - Geoinformation (PFG) 3/04, pp. 201–213.

Hinz, S., Bamler, R. and Stilla, U., 2006. Editorial theme issue: Airborne und spaceborne traffic monitoring. ISPRS Journal of Photogrammetry and Remote Sensing 61(3-4), pp. 135–136.

Holt, A. C., Seto, E. Y. W., Rivard, T. and Peng, G., 2009. Object-based detection and classification of vehicles from high-resolution aerial photography. Photogrammetric Engineering and Remote Sensing 75(7), pp. 871–880.

Kembhavi, A., Harwood, D. and Davis, L., 2011. Vehicle detection using partial least squares. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(6), pp. 1250–1265.

Kirchhof, M. and Stilla, U., 2006. Detection of moving objects in airborne thermal videos. ISPRS Journal of Photogrammetry and Remote Sensing 61(Issues 3-4), pp. 187–196.

Kozempel, K. and Reulke, R., 2009. Fast vehicle detection and tracking in aerial image bursts. In: U. Stilla, F. Rottensteiner and N. Paparoditis (eds), CMRT09, Vol. 38, 3/W4, ISPRS, pp. 175–180.

Kurz, F., Tuermer, S., Meynberg, O., Rosenbaum, D., Leitloff, J., Runge, H. and Reinartz, P., 2012. Low-cost optical camera systems for real time mapping applications. Photogrammetrie, Fernerkundung, Geoinformation (PFG) 2012(2), pp. 0159–0176.

Leitloff, J., Hinz, S. and Stilla, U., 2010. Vehicle extraction from very high resolution satellite images of city areas. IEEE Transactions on Geoscience and Remote Sensing 48(7), pp. 2795–2806.

Lenhart, D., Hinz, S., Leitloff, J. and Stilla, U., 2008. Automatic traffic monitoring based on aerial image sequences. Pattern Recognition and Image Analysis 18(3), pp. 400–405.

Liu, L., Fieguth, P., Clausi, D. and Kuang, G., 2012. Sorted random projections for robust rotation-invariant texture classification. Pattern Recognition 1, pp. 2405–2418.

Schofield, C., 2009. Smart routing. Traffic Technology International June/July, pp. 28–31.

Zhou, J., Gao, D. and Zhang, D., 2007. Moving vehicle detection for automatic traffic monitoring. IEEE Transactions on Vehicular Technology 56(1), pp. 51–59.