Hypothesis Generation of Instances of Road Signs in Color Imagery Captured by Mobile Mapping Systems

A.F. Habib*, M.N.Jha
Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada - (habib, mnjha)@geomatics.ucalgary.ca

Keywords: Mobile Mapping Systems, spectral classification, region growing, geometric attributes, hypothesis generation, road signs.

ABSTRACT

Captured imagery by terrestrial mobile mapping systems proved to be an effective tool for the extraction of road and road-side features for a variety of transportation activities. Among these features, road signs are important for several applications such as navigation and driver's alert systems. Manual extraction of road signs from thousands of imagery is a very expensive and tedious process. To mitigate such a problem, this paper proposes a methodology for the generation of hypotheses regarding instances of road signs in the captured imagery. Such a process would reduce the manipulated imagery from several thousands to few hundreds. The task of object recognition involves segmenting and extracting regions and categorizing the regions into predefined classes. Inherently, this is an ill-posed problem. Therefore, a variety of attributes of the extracted regions need to be incorporated for better analysis. Spectral information is one such important attribute. In the case of color imagery, color content makes up the spectral information. In addition to the spectral information, geometric characteristics of the imaging system and the sought-after objects play a significant role in regularizing the recognition process.

More specifically, the proposed strategy will start by using the spectral information of the sought-after signs to isolate regions in the captured imagery with similar signatures. Then, the geometric characteristics of the imaging system will be used to trim some of the defined regions. In other words, the imaging configuration and the driving trajectory relative to the road will be used to define an area of interest where road signs are most probably located. Finally, geometric attributes of the defined regions (such as size, regularity of the region boundaries, and moments) will be used to generate a hypothesis regarding an instance of a road sign. The performance of the proposed methodology will be evaluated by experimental results involving real datasets captured by a terrestrial mobile mapping system.

1. INTRODUCTION

The task of model-based object recognition involves segmenting and extracting regions and categorizing the regions into predefined classes. A variety of attributes of the extracted regions need to be incorporated for better analysis. Spectral information is one such important attribute. In the case of color imagery, color content makes up the spectral information. With the advent of many imaging devices with the capability of generating and working with color images, color image processing has become an integral part of automated image analysis. Captured imagery by terrestrial mobile mapping systems proved to be an effective tool for the extraction of road signs. In addition to the spectral information, geometric characteristics of the imaging system and the sought-after objects play a significant role in regularizing the recognition process.

Color images are emerging to be very important input for automated image analysis tasks (Gonzalez, 1993). Hence, a great deal of current research is addressing the concepts of color image processing and applications of spectral information in automation (Gaurav, 1997). It is a well established fact that object recognition from imagery is an ill posed problem (Schenk, 1992). So any additional information that helps constrain the solution space is valuable. Habib et al. (1999) used color information to extract road signs and tried to identify them with corner points using generalized Hough transform. Hsu and Huang (2001) proposed a road sign detection and recognition system that uses a template-matching method and a matching pursuit filter. Kim et al. (2006) proposed an algorithm that uses line information together with color information for automatic detection of road signs. These are two input sources for the neural networks that are used to detect the location of the road signs. It is believed that the integration of spectral and geometric attributes of the sought after objects as well as the geometric characteristics of the imaging system will improve the success rate of object recognition systems using color imagery. This research is investigating the feasibility of object recognition; namely road signs, from color terrestrial imagery. In order to achieve this objective, we will be focusing on the following issues:

- The optimum reference frame to represent color information.
- Extracting the spectral attributes of the sought-after objects using training procedure.
- Developing the necessary algorithms for segmenting the input imagery according to the spectral attributes obtained from the training system.
- Use the geometric characteristics of the imaging system and the sought-after objects to extract or hypothesize instances of these objects in the input imagery.
- Design and implement a strategy combining the above-mentioned operations.

The following flow chart in figure 1 briefly describes our approach for generating hypothesis. An overview of the HSI model is given in the next section. In the second section, the training data definition and classification strategy are explained. The seed pixel generation from the classified image is also explained in this section. Region growing from the seed pixel and region attribute extraction is explained in the third section. We will also discuss how to extract spectral and geometric attributes of the segmented regions in this section.

*Corresponding author.

The fourth section addresses determining the area of interest (area where we expect to have the road signs) based on the geometric characteristics of the imaging system. The strategy to generate hypotheses about instances of the road signs in the input imagery is discussed in section five. The sixth section is dedicated to the experiment results. The case study involved color image sequence captured by a Mobile Mapping System (MMS). Finally, conclusions as well as recommendations for further research are presented in the section seven.

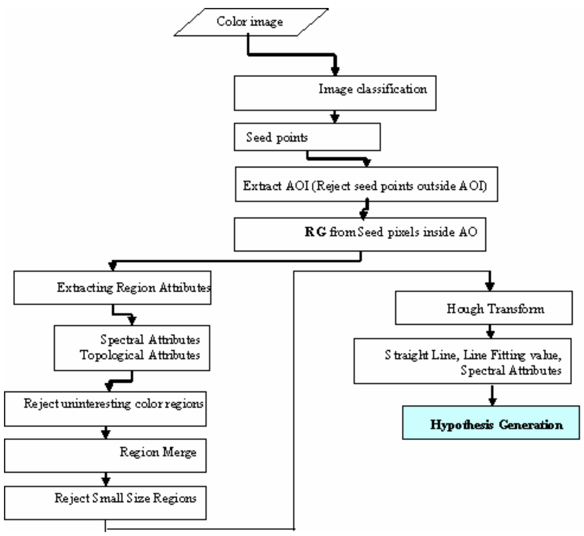


Figure 1: The flow chart for hypothesis generation

1.1 HSI (Hue-Saturation-Intensity) Model

We can describe the visual effect of any spectral distribution by three dimensions i.e. dominant wavelength (hue), saturation and intensity. This model is very useful, since it decouples color information from intensity information. This is beneficial for illumination invariant recognition systems. HSI Model is used along with the RGB model for the classification in our work.

Hue distinguishes among colors such as red, green, purple etc. It is usually defined as the dominant wavelength in a spectrum. Saturation refers to how much the color is pure from a gray of equal intensity. A completely pure color is one hundred percent saturated. Hence, saturation is expressed as either between zero and one. Intensity refers to the amount of light the color reflects.

2. TRAINING, CLASSIFICATION AND SEED PIXEL GENERATION

2.1 Defining Training Data

Training imagery should be representative of all color signs to be extracted. We should have enough spectral classes to represent the colors associated with the sought-after signs. For each color of interest (road sign colors), sample data were manually extracted (by selecting a rectangular window on the image) in homogeneous parts. By repeating this step using similar colors in the training data, the spectral statistics for this class (such as minimum, maximum, mean, median, and standard deviation) can be determined. Spectral statistics of the training data will be evaluated in both HSI and RGB domains. Eliminating the extreme values (5% for example) from both sides, after sorting the individual regions within the same class before merging would reduce the noise in the training data.

2.2 Classification Strategy

At first we examine each pixel in the image whether it lies within the boundaries of any of the given classes in the training data. The box for a class is formed using the minimum and the maximum gray value of that class in each of the RGB bands(R, G and B). If the pixel lies in only one box, then it's labeled as that class. If the pixel lies in more than a box, classification will be based on the minimum spectral distance of the pixel from the classes' means. Spectral distance will have different formulas based on the saturation. Whenever we are dealing with low saturation (a shade of gray), the distance is computed based on the intensity differences. On the other hand, for high saturation coefficients (almost pure colors) we estimate the distances based on hue differences. In case of average saturation, we use the RGB domain to calculate the spectral distance. More details about various classification techniques can be found in Lillesand and Kiefer, 1999.

2.3 Seed Pixel Generation

The classification stage outputs an indexed image, where the indices represent the color classes. Afterwards, connected component labeling is performed to connect pixels with the same index into blobs. The closest pixel to the centroid of the blob, which has homogenous neighborhood as shown in the figure 2, will be used as a seed pixel for the region growing process.

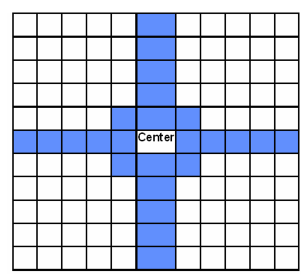


Figure 2: The neighborhood for the seed pixel qualification

3. REGION GROWING AND REGION ATTRIBUTE EXTRACTION

We grow the region from the seed pixels which lie in the AOI. The procedure to define the AOI is explained in section four. We have used 8 connected region growing. For region growing, we have to define the following threshold values:

- Mean threshold value (μ).
- Range threshold value (σ).

Now we can have two kinds of thresholds:

- Constant
- Adaptive

In this research, we have used adaptive threshold. We find out new σ and μ after adding new neighboring pixels to the region. Based on this new sigma and mean, we try to update the threshold. Initial mean is the gray value of the seed pixel. Stop condition is false if difference between the gray value of the pixel and the mean is within the standard deviation for the region growing. If stop condition is false, we will keep on growing the region. As we have multiple bands(R, G and B), there can be various stopping conditions by combining stop conditions in individual bands:

- AND: Loose Condition
  Stop(R) & Stop (G) & Stop (B) → STOP
- OR: Strict Condition
  Stop(R) || Stop (G) || Stop (B) → STOP

We have used loose condition in our work but we have one extra condition as in the 8-connected neighborhood; number of pixels satisfying this loose condition should be greater than 4.

3.1 Region Merging

Successful region merging should utilize both spectral and geometric characteristics of the regions under consideration. Regions can be considered for merging if and only if they have similar spectral properties (e.g., dominant color). This can be established by comparing the mean values of the color regions in the HSI/RGB bands as determined after region growing. On the other hand, the closeness of these regions can be evaluated by comparing the number of the boundary pixels in the two regions under consideration (the distance between these boundary pixels are smaller than a predefined threshold) and the perimeter of those regions, See Figure 3. If this ratio is larger than a certain threshold, we will assume that they originated from the same object. Thus, they will be merged together.

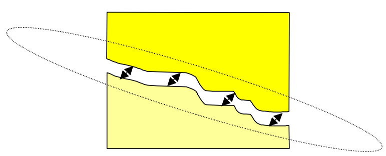


Figure 3: Spectral and spatial criteria for region merging

3.2 Region Attributes

We obtained homogeneous regions in the input imagery whose spectral properties are similar to the color classes derived from the training data. We need to extract various region attributes to generate hypothesis about the instances of the road signs. We extract spectral as well as geometric properties of the region.

- Spectral Properties:
  - Dominant Color, Saturation and Class number
- Geometric Properties:
  - Area
  - Compactness (Circularity)
  - Boundary
  - Presence of straight lines, line-fitting value

3.2.1 Boundary Detection: Boundary of the segmented region is an important geometric property. The procedure to detect the boundary of the region is explained briefly here. The starting point of the boundary pixel is the upper left point of the region. Now, we look for the pixel having same label as the central pixel in the 8-connected neighborhood. The direction for checking is in the anticlockwise direction (see figure 4). For the starting pixel the initial direction to check for the identical pixel is fixed as '3'. Then we move from direction 3 to 7 until we encounter a pixel having identical label. Once we find the similar pixel, we add that to our list of boundary pixels. Now this new found pixel is our new central pixel. We have to ensure that we don't go to previously visited pixels to check for the identical label. We update the starting direction such that previous visited pixels are avoided. We get the boundary of the segmented region at the end of this process.

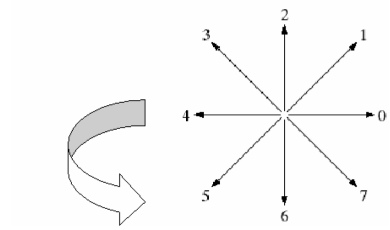


Figure 4: Directions from the central pixel to its neighbors for the 8-connectivity

3.2.1 Circularity: Circularity is calculated based on the compactness (see figure 5). Compactness= Semi Minor Axis / Semi Major Axis; i.e., ratio of the minimum and the maximum eigen values of the variance-covariance matrix of the boundary pixels.

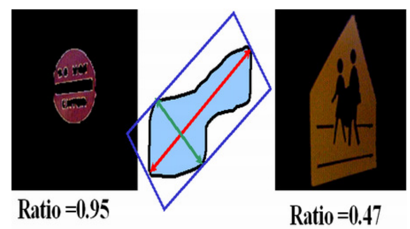


Figure 5: Compactness Ratio for two signs

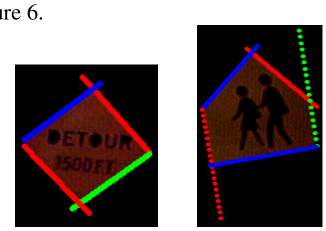
3.3 Hough Transformation

After performing region merging, boundaries of regions are inspected for the existence of straight lines by means of Hough Transform (Hough, 1962). This will be useful in improving the quality of the generated hypotheses. It is assumed that regions originating from signs will be bounded by a sequence of straight lines.
A straight line in 2-D can be represented by the following equation:

$$x \cos\theta + y \sin\theta = \rho \qquad (1)$$

We can define a parameter domain, where every line in the spatial domain will be represented by a point in that domain. We can construct an infinite number of lines going through one point in the spatial domain. These lines will be represented by a sinusoid in the parameter domain. Therefore, if we have M collinear points in the spatial domain lying on the line defined by $x \cos\theta_j + y \sin\theta_j = \rho_j$, these points will be represented by M sinusoidal curves in the parameter space that intersect at (ρ_j, θ_j). By incrementing the parameter space at all the locations along the different sinusoids, we will get a peak at the location (ρ_j, θ_j). This peak denotes detected line.

3.3.1 Line-fitting Value: Line-fitting value is defined as the ratio between the boundary elements that belong to straight lines, and the region perimeter. The higher the line-fitting value, the better is the fit. We find the normal distances between boundary pixel and all the detected lines for a given boundary. The minimum of these normal distances is compared with the threshold for a given boundary pixel. If the minimum normal distance is less than the threshold then we assume that the point belongs to the line. We do this for all the boundary pixels. Line- fitting value is given by the ratio of the boundary pixels belonging to the lines and the total number of boundary pixels; i.e., perimeter. Line-fitting values for some regions are shown in figure 6.



LF: 0.906 LF: 0.979

Figure 6: Line-fitting values for some region

4. DEFINING AREA OF INTEREST(AOI)

This is the area in the image space where we believe signs will exist. The main objective of predicting the sign location within the image is to reduce the computation time by only concentrating on the relevant areas. The expected sign location in the images depends on the internal calibration parameters of the camera as well as the geometric information describing the camera's orientation and position relative to the expected road sign locations in the object space.

4.1 Expected Road Sign Location in the Object Space

Before determining the AOI in the image space, we have to determine a three-dimensional volume in the object space that most probably includes the signs. We will define this volume relative to an object space coordinate system. A schematic diagram of the object space coordinate system can be seen in Figure 7. This system is defined as follows:

- The origin of the coordinate system is on the road, just underneath GPS antenna that is mounted on the van. If there is no GPS positioning component on board, then the origin will be located at the road level vertically below the camera perspective center.
- The X-axis coincides with the road surface and normal to the driving direction of the imaging system.
- The Y-axis coincides with the road surface and is parallel to the driving direction.
- The Z-axis is chosen in such a way that the XYZ-coordinate system is right handed. Therefore Z-axis will point upwards.

It has to be noted that the object space coordinate system changes its origin and orientation as the imaging system moves. Therefore, each image will have its own object coordinate system. The area of interest in the object space is defined by extreme values (minima and maxima) of the object coordinates of the vertices defining a three-dimensional volume as follows (see Figure 8):

- X_{min} and X_{max} will be selected based on the road width, the driving lane, and the sidewalk width.
- Y_{min} and Y_{max} will be selected based on the appropriate scale for the recognition process. One should note that as Y increases, the sign would appear smaller making the hypothesis generation less practical.
- Z_{min} and Z_{max} will be selected based on the expected sign height relative to the road surface.

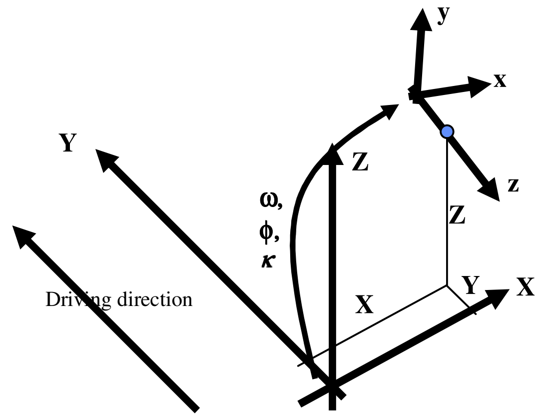


Figure 7: Relationship between the object and the image coordinate system

4.2 Expected Road Sign Location in the Image Space

In order to project the vertices of the rectangular volume into the image space, we need to know the spatial offset and the rotational relationship between the object and camera coordinate systems as shown in the figure 7. In addition, we should have the interior orientation parameters of the implemented camera. These parameters include:

- ΔX, ΔY, ΔZ: The calibrated shift between the object space and camera coordinate systems.
- ω, ϕ, κ : The calibrated rotation angels between the two coordinate systems.
- x_p, y_p, c : The calibrated interior orientation of the camera.

These parameters can be determined through a calibration procedure (Novak and Habib, 1998).

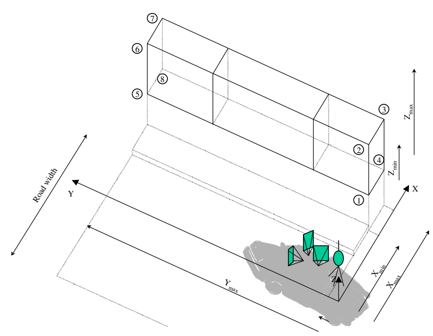


Figure 8: **AOI in the object space**

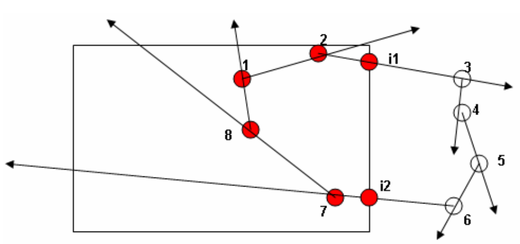
We get the image coordinates of the AOI vertices using these EOP and IOP. The collinearity equations combine the EOP and IOP together with the object coordinates of the vertices to determine the corresponding image coordinates. The collinearity equations can be described as follows:

$$\begin{bmatrix} x_i - x_p \\ y_i - y_p \\ -c \end{bmatrix} = \lambda . R^T \begin{bmatrix} X_i + \Delta X \\ Y_i + \Delta Y \\ Z_i + \Delta Z \end{bmatrix}, i = 1,..,8 \qquad (2)$$

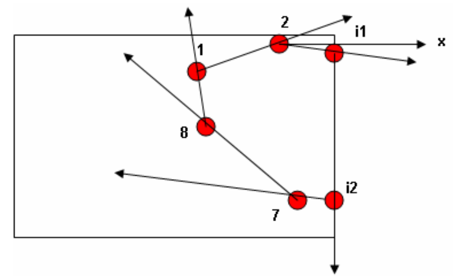
Where x_i, y_i (i=1,..8) are the corresponding image coordinates of the corners of the area of interest. The scale factor (λ) in equation can be eliminated by dividing the first two lines by the third one, leading to two equations per point. These equations can be used to compute the image coordinates of the vertices of the rectangular volume. Afterwards, we form the polygon in the image corresponding to the AOI. An example of forming polygon from the 8 vertices of the AOI is shown in the figure 9. First, we try to get the AOI vertices inside the image boundary as shown in figure 9a. We start with any two vertices say a_i and a_{i+1} of the original AOI and check for the number of intersections the vector joining these two vertices has with the polygon. There can be three cases for the intersection:

- No intersection: This means that the both the points are outside. In this case we don't do anything. For example in figure 9a, the line joining point '3' and point '4' does not intersect image boundary, and hence we do not consider them for the AOI vertices.
- One intersection: This means that either first vertex (a_i) or both the vertices lie inside the image boundary. Point a_i belongs to the modified AOI. If the intersection point lies between a_i and a_{i+1}, then the intersection point also belongs to the modified AOI. In figure 9a, the line joining point '1' and point '2' has only one intersection with the image boundary, and thus point '1' is selected as an AOI vertex.
- Two intersection: This means that starting point a_i is outside the image boundary. If the points of intersection lie between a_i and a_{i+1}, then the intersection points belong to the modified AOI. The line joining point '6' and point '7' has two intersections with the image boundary. Intersection point 'i2' lies between those points, and thus we consider point 'i2' as new AOI vertex.

We keep doing this until we reach the starting vertex again. We get the AOI vertices (1, 2, i1, i2, 7 and 8 in figure 9a) inside the image boundaries at the end of this process. The boundaries of the polygon can be tracked starting from any point with extreme coordinates and searching for directions with minimum or maximum angles relative to predefined reference directions. For example, in Figure 9b, we can start by identifying the point with the maximum y-coordinate (point '2'). We may choose the positive direction of the x-axis as the initial reference. The tracking will proceed from point '2' by looking for the point, which makes the smallest angle with the reference direction (tracking is done in clock-wise direction). The point 'i1' will be selected as the second vertex of the polygon in this example. We continue from point 'i1' while using the line connecting point '2' and 'i1' as the reference direction. The tracking continues in a clock-wise direction until we reach the starting point. At the end of this process we get the AOI inside the image as polygon.



(a)



(b)

Figure 9: a) Obtaining AOI vertices inside the image boundary b) Tracking of the polygon boundaries

5. HYPOTHESIS GENERATION

The main purpose of our work is to generate hypothesis about the existence of any of the road signs in the input imagery. Using the extracted attributes, a decision will be made to generate such a hypothesis. These attributes include:

- Spectral Properties:
  - Dominant Color and Class number
- Geometric Properties:
  - Compactness (Circularity).
  - The uniformity of the region boundaries (presence or absence of straight lines), line fitting value

By considering these factors, we will generate a hypothesis for each image in the input sequence. The number of straight lines we get from Hough transformation is the first criteria for the hypothesis. If the number of straight lines is greater then some threshold then, we look for the colors, red and blue in the region. If the color of the region is one of them then we say that region has distinct color otherwise we regard the region as Non-Specific shape. Generally signs are red or blue colored that's why we check for these colors in the region. If the number of straight lines is less than the threshold then we look for the line fitting value. As already mentioned, higher line fitting value means that more boundary pixels actually lie on the line. If the line fitting value is higher than some threshold, then we say that the region has organized shape otherwise we check for the circularity. Circularity is based on the compactness ratio. If this compactness ratio is greater than some threshold, we say that the region is circular otherwise we can not say anything about the shape of the region. The strategy for the hypothesis generation is shown in the figure 10. The various thresholds for the hypothesis generation are listed in the table 1.

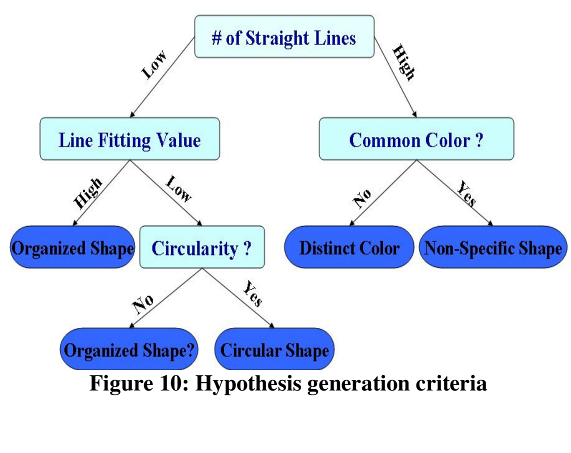


Figure 10: Hypothesis generation criteria

Table 1: Thresholds for hypothesis generation

Parameters	Threshold
Number of Lines	7
Line Fitting Value	0.8
Compactness Ratio	0.8

6. EXPERIMENTAL RESULTS

Experiments are conducted using terrestrial color imagery captured by a terrestrial MMS. The system utilizes four color cameras with 700x400 pixels in the focal plane. Intermediate results according to the suggested strategy are presented. An example of the input imagery is shown in Figure 11. This image contains an orange sign. The classified image according to the color classes determined through the training process can be seen in Figure 12. After the classification, neighboring pixels with the same labels are segmented using connected component labeling. Within each segmented region, one seed point is selected for the growing step. The seed pixel is selected as the closest pixel to the region centroid that belongs to that class. The seed pixels extracted from the classified image are shown in Figure 13. In order to reduce the computation time during the growing step, only the seed pixels within the area of interest will be considered. The area of interest is also depicted in Figure 13. Multi-band adaptive criteria is used to grow the region. Two regions are created, but they are spectrally and spatially close, and hence they are merged to form a single merged region. Following region segmentation, geometric as well as radiometric attributes are extracted. Hough transform is then used to find instances of straight lines in the boundaries of the segmented region. The presence of straight lines is another indication that these regions correspond to road signs. Extracted straight lines are shown in figure 14. The hypothesis is made based on geometric as well as spectral properties of the extracted regions as shown in figure 15. Figure 16 depicts various road signs extracted using our strategy.

Figure 11: A Terrestrial Color Image

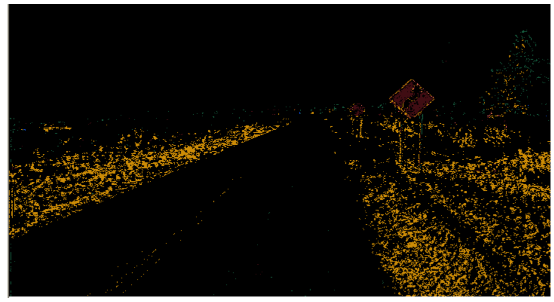


Figure 12: Classified Image.

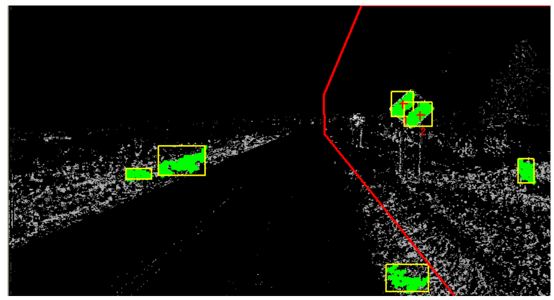


Figure 13: Seed pixels & the Area of Interest in the Classified Image.

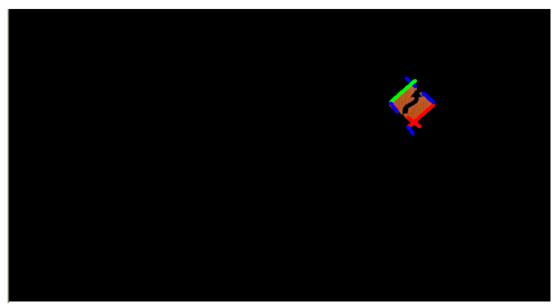


Figure 14: Straight Lines in the Boundary of the Segmented Regions

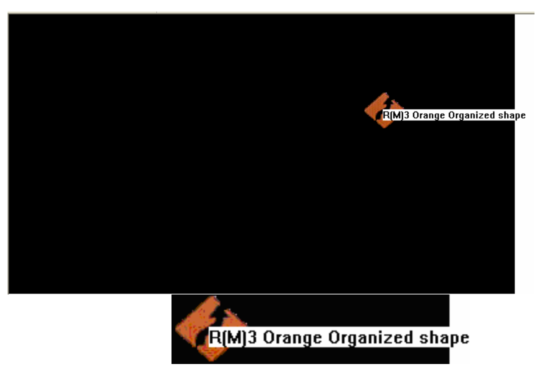


Figure 15: Hypothesis generated for the merged region, bottom (cropped region)

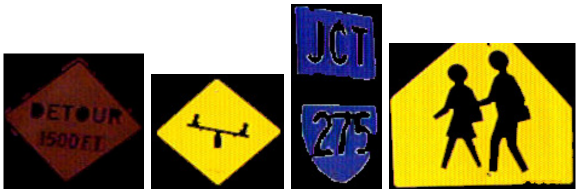


Figure 16: Example of some extracted road signs

7. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The results obtained from the suggested approach illustrate the feasibility of using spectral response patterns in extracting road signs from color imagery. This strategy is an attempt to combine both spectral as well as geometric properties of the objects in the recognition process.

So far we have generated hypotheses about the existence of road signs in color terrestrial imagery. Future work should concentrate on hypothesis verification. Comparing the extracted attributes from the input imagery and a database will be used in the verification stage. Probabilistic measure should be associated with the verified hypotheses.

8. REFERENCES

Gaurav, Shouma, H.Joel Trussel, 1997. Digital color Imaging, *IEEE Transactions on Image Processing*, Vol. 6, No.7, pp.901-932

Habib, A., Uebbing, R., Novak, K., 1999. Automatic Extraction of Road Signs from Color Terrestrial Imagery. *Journal of Photogrammetric Engineering and Remote Sensing*, Vol. 65, May 1999, pp. 597-601.

Hough, 1962, *Methods and Means for Recognizing Complex Patterns*, U.S. Patent 3,069,654.

Hsu, S. H. & Huang, C. L. (2001), Road sign detection and recognition using matching pursuit method, *Image and Vision Computing*, 19(3), 119–29.

Kim, G., Sohn, H., Song, Y. 2006.Road Infrastructure DataAcquisition Using a Vehicle-Based Mobile Mapping System. *Computer-Aided Civil and Infrastructure Engineering* 21 (5), 346–356.

Lillesand, T.M. and R.W. Kiefer, 1999. *Remote sensing and image interpretation*, Fourth Edition. John Wiley & Sons, Inc, New York.

Novak, K., Habib, A., 1998. Concepts for Positioning and Object Recognition for Multi-Sensor Data with GPS Support. *Project Report Submitted to TransMap Corporation*, Columbus, Ohio.

9. ACKNOWLEDGEMENTS

This research work has been conducted with partial funding from the GEOIDE Research Network (TDMASR37).