Terrain Surface Classification with Control Mode Update Rule using 2D Laser Stripe-Based Structured Light Sensor

Liang Lu\textsuperscript{a}, Camilo Ordonez\textsuperscript{b}, Emmanuel G. Collins, Jr.\textsuperscript{b}, Eric Coyle\textsuperscript{b}, Dushyant Palejiya\textsuperscript{b}

\textsuperscript{a}GE (China) Research and Development Center Co. Ltd, Shanghai, China  
Liang.Lu@ge.com
\textsuperscript{b}Department of Mechanical Engineering, FAMU-FSU College of Engineering, 2525 Pottsdamer Street, Tallahassee, FL 32301  
camilor@eng.fsu.edu, ecollins@eng.fsu.edu, ejc06c@fsu.edu, dp09h@fsu.edu

Abstract

It is necessary for autonomous ground vehicles operating on outdoor terrains to identify and adapt to different terrains in order to improve their mobility and safety. This work presents a classification scheme to identify outdoor terrains and an update rule to reduce the possibility of implementing control modes based on classification inaccuracies. A laser stripe-based structured light sensor, which has a laser and infrared camera component, is used to sense terrains directly in front of the vehicle (<1m). Use of this infrared vision system allows sensing at night unlike many previous vision-based approaches that rely on stand-alone cameras. Also, unlike many previous results the classification algorithm presented here does not rely on measures of color which are subject to illumination and weather conditions. Instead, the method presented here relies on spatial relationships which are captured in two quantities: spatial frequency and texture. The presented terrain classification scheme uses a probabilistic neural network classifier to exploit the spatial differences in four terrains: asphalt, grass, gravel and sand. This approach yields empirical results that report a greater than 97% classification accuracy when both spatial frequency and texture features are used. Color robustness and lighting robustness is then shown through additional experiments. Furthermore, instead of implementing control modes directly from the identified terrains, it is shown that the use of current and past terrain detections allows for the rejection of misclassifications with minimal effect on the rate at which a new control mode can be implemented.

Keywords: Terrain Classification, Terrain-Dependent Control, Computer Vision

1. INTRODUCTION

The use of autonomous ground vehicles (AGVs) is increasing for outdoor tasks in unstructured environments. These tasks include agricultural applications, search and rescue missions, as well as military missions. It follows that AGVs will be required to traverse a variety of terrain surfaces. To enable safe and efficient traversal, a vehicle’s control system may be tuned to a given surface. An example of such a control system is the Terrain Response system available on many Land Rovers [1]. This system has modes for everyday driving, grass/gravel/snow, mud and ruts, sand, and rock crawls. Each mode has pre-defined settings that change vehicle parameters such as anti-lock braking, throttle response, and differential locking. Terrain-dependent AGV guidance can also involve limits on turn radius and speed, tire pressure adjustments, rut following, etc. In manned systems, the driver must recognize the terrain and then make these necessary adjustments. However, implementing terrain-dependent control adjustments in AGVs requires automated terrain classification.

Terrain surface classification methods for mobile robots have two main categories, those based on proprioceptive (i.e., vibration and slip) sensors [2, 3, 4, 5, 6, 7, 8, 9] and those based on vision sensors [10, 11, 12], corresponding respectively to “feeling” and “seeing.” Unlike proprioceptive sensors, which are often associated with an inertial measurement unit, measurements from vision sensors are essentially independent of the speed and load of the vehicle. However, vision sensors can lead to misclassifications when the ground has a superficial covering (e.g., leaves, dry grass, or a small amount of water) or when the environment has reduced visibility due to smoke, fog or other precipitation. They may also fail to discern between surfaces that have a similar appearance, but are very distinct from a control perspective, such as dry and wet sand. Hence, what is needed is a synergy of the two basic approaches to terrain classification. Two methods to combine the classification results from the two basic approaches are given in [11], ultimately showing that the synergy of vision and vibration sensors can produce improved detection results for the Mars surface.

Most prior classification research based on the use of optical or laser-based sensors [13, 14, 15, 16, 17, 18, 19, 20] has focused on classifying terrain to determine traversability or non-traversability. Traversability classification in [13] uses cameras and is approached by characterizing the terrain in terms of roughness, slope, discontinuities, and hardness. The research of [14] used a multi-spectral sensor to distinguish between pliable vegetation and true mobility obstacles. The research of
used a 2D laser scanner to discriminate between obstacles and traversable terrain. Additional research classifies terrains into broad categories. For example, the research of [16] and [17] used three far range sensing ladars (≥ 10 m), Z+F LARA 21400, the LADAR used on the experimental unmanned vehicle (XUV) and Acuity AccuRange 4000, to distinguish ground and vegetation. The research of [19], which also used three far range sensing ladars, SICK laser LMS 291, Z+F LARA 21400 and Riegl LMS-Z210, performed 3D data segmentation for classifying terrain into three categories: “scatter” (porous volumes such as grass or tree canopy), “linear” (thin objects) and “surface” (solid objects). In [19] terrain traversability decisions are made based on geometry and color data from IR range sensors and stereo cameras. However, authors of [20] use only stereo vision sensors to classify terrain as ground plane (traversable) or obstacle (non-traversable) near-field and then apply online learning technique to estimate far-field traversability.

This paper focuses exclusively on surface classification for traversable terrains using vision sensors. Only a few vision-based approaches to this problem have appeared in the literature to date. The research of [10] used a long range camera and classified large terrain patches using average red, average color, 3D color histogram, and texture features. Parallel research in [11] used cameras at short range to detect terrains in close proximity to planetary rovers on Mars-like environments. In addition, short-wave, mid-wave and long-wave infrared cameras, which work at night, were used in [12] to classify mud. Additional research presented in [21], proposed an image-retrieval approach that utilizes wavelet signatures to estimate the size of fine particles in mineral ore. This research has potential to be used in the domain of surface classification.

The terrains classified in this research are similar to those classified in [10] and [11]. However, the research presented in this paper uses a laser linestriper [22], which is used here as shorthand for a laser stripe-based structured light sensor [23]. It is a high resolution, low proximity, hybrid sensor, consisting of a single infrared camera and a single laser. An advantage that this sensor has over a stand-alone camera is that it is an active sensor, due to the laser, and is hence capable of classifying terrains at night. The laser linestriper captures the intersection line between the laser plane and the terrain. However, compared with a standard laser scanner with a 180° or 360° field of view, the laser linestriper has only a 30° field of view. However, a 30° field of view is sufficient for this research as it is only desired to detect a small patch of terrain at close range (<1m), directly in front of the vehicle. The laser linestriper has high resolution: 0.05° angular resolution and 4 mm range resolution.

Since color may change under different illumination and weather conditions and also certain terrains may have more than one color, e.g., grass may be green or yellow and sand may be red or white, the methodology used for classification, unlike that in [10] and [11], does not rely on color; instead it relies only on spatial relationships, specifically spatial frequency response and texture, which captures spatial relationships between different gray levels. This work expands on the work of [24] where terrain classification with each of these features separately is conducted using a probabilistic neural network (PNN) [25], similar to its use in [6, 7, 8, 9]. Classification results achieved by combining both of these features are presented here and compared with the results that can be achieved individually by these features [24]. This paper also considers the use of the classification results in updating the vehicle control mode.

There is always a possibility that the results of the terrain classification scheme will be inaccurate, which means that misclassifications could be used to update terrain-dependent control modes, adversely impacting vehicle mobility. For example, if the classifier inaccurately classifies sand as a paved surface, then the AGV could lose mobility due to the use of a paved road control mode on sandy terrain. An update rule is proposed that considers prior classification results in order to avoid inaccurate control mode usage, utilizing the method proposed in [26] which implements an update rule for vibration-based terrain classification.

This paper is organized as follows: Section 2 presents the proposed approach to terrain classification and the update rule. Section 3 describes the experimental setup and presents the results of the experiments conducted on four different outdoor terrains. Section 4 presents concluding remarks and directions for future research.

2. PROPOSED APPROACH

The proposed terrain classification method is composed of four stages: data collection, feature extraction, classification and update rule as described below.

2.1. Data Collection

The laser linestriper consists of an infrared laser and an infrared camera. The laser light is fanned out in about 60° to form a plane of light. If an object or the ground intersects the laser plane, the reflected light can then be observed by a camera positioned at some vertical offset from the laser. The camera also has an angular position β which allows the camera to view the ground/laser plane intersection. The camera then captures grayscale images, in a resolution of 640 × 480, that allow triangulation of the laser line in the image. This sensor setup is shown in Figure 1. It should be noted that the size of the terrain patch observed by this sensor will be dependent on the vertical position of the camera, with higher positioning corresponding to a larger field of view and lower resolution. Sample images from four terrains, asphalt, grass, gravel and sand were collected and are shown in the first column of Figure 2. This particular mounting of the laser and camera results in a terrain patch of 45 cm × 35 cm and has a resolution of less than a millimeter (∼ 0.8 mm).

2.2. Feature Extraction

Two different features, spatial frequency response and texture, were used separately and in combination to characterize the terrain. Although this discussion on the determination

---

1 The results were based on actual camera data from Mars.
Figure 2: Raw laser line images (left column), extracted terrain profiles (middle column) and terrain profiles after slope and DC component elimination (right column) for asphalt (a),(b),(c); grass (d),(e),(f); gravel (g),(h),(i); and sand (j),(k),(l).
of spatial frequency response and texture considers four common terrains (asphalt, grass, gravel and sand) the classification methodology is applicable to a greater number of terrains.

2.2.1. Spatial frequency response

Unlike texture, the spatial frequency response does not take advantage of the intensity data provided by the camera in the laser line stripe. However, an accurate frequency response results from the high spatial resolution provided by the laser line stripe. This feature also provides a direct link between vision-based classification and vibration-based classification [6, 7, 27]. In fact, as first discussed in [7] and later in [27], the terrain signatures in the frequency response of the vehicle vibrations are a direct result of the spatial frequency response terrain signatures shown by the results of this paper.

To compute the spatial frequency response feature, a homogeneous transformation must first be obtained by calibrating the laser line stripe [22]. The points on the laser line are then extracted from the image and transformed from the image coordinates to the laser plane coordinates \((x, y)\). The resulting set of points in the laser plane coordinate system is used to represent the terrain profile. Figure 2 shows the raw images of four different terrains captured by the line stripe and the resulting terrain profiles.

Before finding the frequency contents of the terrain profiles, it is important to remove information that is not inherent to the terrain type like the zero frequency (DC) component and the average slope of the profile. The DC component represents the average height of the profile and the slope characterizes the average incline/decline of the terrain surface. The slope and DC components are removed by fitting a line in the least squares sense to the profile, and subtracting the corresponding predicted value obtained using the approximating line from the \(y\) (vertical) component of each point on the profile. The third column of Figure 2 presents the terrain profiles after the elimination of the DC component and slope.

The new terrain profile \(y\)-values obtained after the elimination of the DC component and slope form a discrete sequence \(y[n]\) with length equal to the number of laser points. To find the frequency components of \(y[n]\), the spatial Fourier transform of \(y[n]\) is obtained by using the Fast Fourier Transform (FFT). The magnitude of the frequency response is then organized into the frequency based feature vector \(x_f\) defined by

\[
x_f = \left[|y(j\omega)|\right]^T.
\]

Figure 3 shows the average frequency domain representation of the four different terrain profiles being considered. Notice that asphalt has the flattest response because asphalt has the flattest surface among these four terrains. Sand possesses substantial low frequency content due to the low frequency undulations that naturally occur in sand. Grass and gravel have more high frequency content than either sand or asphalt. Although the average response makes it appear that the spatial signature of grass is substantially different from that of gravel, the large amount of variability in the signatures of these two terrains can sometimes make it difficult to distinguish between the two terrains.

Once the spatial Fourier transform of \(y[n]\) is obtained, it is observed through Principal Component Analysis (PCA) [28] that the first 10 principal components capture more than 90% of Fourier transform energy. This leads to the use of the projection of the spatial frequency response onto the space spanned by the eigenvectors corresponding to the first ten principal components. Hence, the resulting feature vector, \(x_f\), is of dimension 10.

2.2.2. Texture

Although a formal definition of texture does not exists, intuitively this descriptor provides measures of properties such as smoothness, coarseness, and regularity [29]. To characterize texture it is necessary to characterize the gray level primitive properties as well as the spatial relationships between them [30].
In this paper, a statistical approach based on the Gray-Level Co-occurrence Matrix (GLCM) [29] is chosen to describe the texture of the laser line striper images. To quantify the texture content of the laser line striper images, a set of descriptors from the GLCM are obtained. The GLCM can be specified in a matrix of relative frequencies \( p(i, j) \) in which two adjacent pixels occur on the image, one with gray level \( i \) and the other with gray level \( j \) [30]. The following textural features are chosen to quantify the texture content of the images:

1. **Contrast**: Measures the local variations in the GLCM, given by
   \[
   \sum_{i,j} |i - j|^2 p(i, j). \tag{2}
   \]

2. **Correlation**: Measures how correlated a pixel is to its neighbor over the whole image, which is mathematically described by
   \[
   \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}, \tag{3}
   \]
   where
   \[
   \mu_i = \sum_i \sum_j p(i, j), \quad \mu_j = \sum_j \sum_i p(i, j), \tag{4}
   \]
   \[
   \sigma_i = \sum_i (i - \mu_i)^2 \sum_j p(i, j), \tag{5}
   \]
   \[
   \sigma_j = \sum_j (j - \mu_j)^2 \sum_i p(i, j). \tag{6}
   \]

3. **Energy**: Provides a measure of uniformity, given by
   \[
   \sum_{i,j} p^2(i, j) \tag{7}
   \]

4. **Homogeneity**: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal, mathematically described by
   \[
   \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \tag{8}
   \]

The texture feature vector \( \mathbf{x}_t \) is then chosen as the 4 dimensional vector,
\[
\mathbf{x}_t = \begin{bmatrix} \text{Contrast} & \text{Correlation} & \text{Energy} & \text{Homogeneity} \end{bmatrix}^T, \tag{9}
\]
such that each component represents a texture metric. Figure 4 shows texture descriptors for a representative sample of each of the four terrains. (Note that for ease of visualization, the values of contrast are 100 times larger than the actual values.) Three of the descriptors, correlation, energy and homogeneity are dominant in magnitude and have distinct patterns for most terrains. However, asphalt and gravel have similar patterns of these three descriptors, leading to the need to also include contrast values.

![Figure 4: Texture descriptors for representative samples of four terrains.](image)

2.3. **Classification**

A probabilistic neural network (PNN) was selected to classify the feature vector \( \mathbf{x}_f \) or \( \mathbf{x}_t \) as a particular terrain. It was chosen because of its simplicity, robustness to noise and fast training speed [25].

The PNN is based on Bayesian classifiers and uses a supervised training set to develop distribution functions. These functions are used to estimate the likelihood of an input feature vector being part of a learned category or class. The learned patterns can also be combined or weighted with a priori probability of each category to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally likely and the determination of category is solely based on the closeness of the input feature vector to the distribution function of a class [31].

![Figure 5: Architectural structure of probabilistic neural network.](image)

The network structure of the PNN is shown in Figure 5. The network has an input layer, a pattern layer, a summation layer, and an output layer [32]. The input layer buffers an arbitrary input feature vector \( \mathbf{x} \) to the neurons in the pattern layer. Then, the neuron \( x_j \) computes its output using...
\[
\phi_j(x) = \frac{1}{(2\pi)^{d/2}} \exp \left( \frac{-(x-x_j)^T(x-x_j)}{2\sigma_j^2} \right),
\]

where \( i \) is the class, in this research the terrain type, \( j \) is the training sample in class \( i \), \( \sigma \) is a smoothing factor, and \( l \) is the dimension of the feature vector \( x \). The summation layer uses a Gaussian distribution as the activation function [28], to compute the probability \( P(C_l|x) \) of a given input \( x \) belonging to a class \( C_i \). In particular,

\[
P(C_l|x) = \frac{1}{(2\pi)^{d/2}} \sum_{j=1}^{m_i} \exp \left( \frac{-(x-x_j)^T(x-x_j)}{2\sigma_j^2} \right),
\]

where \( m_i \) denotes the total number of samples in class \( C_i \). The output layer uses the calculation of the probability distribution function from the summation layer, and applies the decision rule of (12) to select the class with the highest probability.

The vector \( x \) is said to belong to a particular class \( C_i \) if

\[
P(C_l|x) > P(C_l|x_j), \quad \forall j = 1, 2, \ldots, m, \quad j \neq i.
\]

2.5. Update Rule

As the classification result, regardless of whether it is a true or false positive, will be used to implement terrain-dependent control modes, a process is needed which prevents or minimizes control mode usage based on misclassifications. The research of [26] utilizes such a process, called an update rule, by considering both the current and past terrain detection. The motivation behind the update rule is to utilize inherent consistency in terrain data that results from the fact that terrain changes are not random. This allows for misclassifications to be rejected while minimizing the delay in switching control modes when a true terrain transition is experienced.

The update rule has two design parameters:

1. \( n \): the width of the history window.
2. \( \eta \): the minimum fraction of samples required to be of a single terrain in the history window before the control mode will be updated. (It can be viewed as a threshold required to switch control modes.)

When a new terrain is classified, the update rule considers the new classification and the \( n-1 \) prior classifications. The update rule then calculates the percentage of each terrain in this history window. If the highest percentage \( \tilde{p} \) satisfies \( \tilde{p} \geq 100\eta \), then the control mode will be updated. In the case where \( \tilde{p} < 100\eta \), the robot will continue to use the last control mode specified by the update rule. For more information on the update rule, see [26].

Update rule performance is measured in terms of robustness to misclassification and sensitivity to terrain transition. Robustness is expressed as the ratio of the number of accurate control modes used to the total number of control modes used and sensitivity is defined as the average delay in updating control modes when a terrain transition is experienced. A smaller delay (measured in number of samples) implies higher sensitivity. Due to the inherent trade-off in robustness and sensitivity, values of \( n \) and \( \eta \) that simultaneously maximize robustness and sensitivity may not exist. Therefore, in general a Pareto optimal choice of \( n \) and \( \eta \) is sought. A pair \((n, \eta)\) is Pareto optimal if there is no way to increase the sensitivity without reducing the robustness and vice versa. Determining the Pareto optimal values of \( n \) and \( \eta \) begins by defining finite sets of values of \( n \) and \( \eta \) to be considered: \( N = \{n_1, n_2, \ldots \} \) and \( H = \{\eta_1, \eta_2, \ldots \} \). Update rules with all possible combinations of \( n \in N \) and \( \eta \in H \) are then applied to the classification results obtained using the classifiers discussed in Subsections 2.3 and 2.4. For each update rule, sensitivity is computed along with the robustness. Figure 6 shows an example of the computed sensitivity and robustness for \( N = \{1, 2, 3, \ldots, 20\} \) and \( H = \{0.5, 0.525, 0.55, \ldots, 1.0\} \) when applied to the classifier of Subsection 2.4 which uses both spatial frequency and texture features. The set of Pareto optimal pairs \((n, \eta)\) can then be determined. Selection of one pair from this set of Pareto optimal values is dependent on user’s preference between sensitivity and robustness.
3. EXPERIMENTS AND RESULTS

3.1. Experimental Setup

The robotic platform used in the experiments is the ATRV Jr, shown in Figure 7. This vehicle is a skid steered robot that weighs 50 kg and has a maximum translational speed of 1400 mm/sec.

Mounted on the ATRV Jr. is a laser line striper sensor provided by the Carnegie Mellon University Robotics Institute (See Figure 8). This particular sensor will hence be subsequently referred to as the CMU Line Striper. The laser is mounted 37 cm above the ground and is pointed 25° downward with respect to the horizontal plane. The camera is mounted 53 cm above the ground and is pointed 48° downward with respect to the horizontal plane. The images, which correspond to a 45cm x 25cm terrain patch, are then collected at a rate of 15 Hz. This sensor has a unique set of features that make it useful for this terrain classification research.

1. **Illumination Robustness**: This sensor is designed to work under different illumination conditions (e.g., any time of day or night). To make this possible, the system suppresses background from the sun by employing an IR (GaAs 900 nm) pulsed laser, a fast camera shutter (1/100,000 second) and a narrow filter (25 nm) placed between the camera lens, and the CCD sensor.

2. **High Angular Resolution**: The sensor provides high angular resolution. The current set up uses a 30° field of view camera, which achieves an angular resolution of 0.05°.

3. **Real-Time Operation**: This feature is obviously important for real-time terrain classification, an ultimate aim of this research.

4. **Provides Range and Intensity Information**: Due to the usage of both a laser and a camera, both range and intensity information are provided by the system, which respectively provide the ability to compute spatial frequency responses and texture metrics.

5. **Small Size**: This sensor is suitable for use with relatively small robots.

6. **Low Cost**: If mass produced, this sensor is expected to cost on the order of $100.

7. **Eye Safe**: This is a critical feature for the numerous applications in which a robot will operate in the presence of humans.

3.2. Experimental Procedure

A set of outdoor experiments were conducted on the four common terrains shown in Figure 9: asphalt, grass, gravel and sand. For each type of terrain the robot was commanded to travel for 10 sec at speeds of 100 mm/sec to 1000 mm/sec in increments of 100 mm/sec. During each run, the CMU Line Striper images were collected at a rate of 15 Hz. These images constitute the perception data employed to classify the different terrains. Because vision-based terrain classification is less dependent on speed, data collected at different speeds were combined to create a set of feature vectors. This set was then partitioned into three subsets: one set for training; the second for testing and the third for validation purpose. Each of these three sets has 2000 images consisting of 500 images for each of the four terrains.

A smaller set of indoor experiments were also conducted to prove that the laser line striper can be effective at night (i.e., in the dark). Figure 10 shows the three terrains that were brought indoors to enable testing in a controlled light environment. During these experiments, the light luminance was changed to imitate evening and night and the luminance was measured using the Extech EasyView digital light meter EA30. The room lighting originated from fluorescent ceiling lights. The classification
experiments were run under four different light conditions: 1) All lights on. (The luminance was 653 lux.) 2) Half of the lights on. (The luminance was 338 lux.) 3) The other half of the lights on. (The luminance was 293 lux.) 4) All lights off. (The luminance was 0.24 lux.) The data collected under lighting condition 1 was used as training data, while the data collected under the other lighting conditions was used as testing data.

In previous experiments, the robot only traversed on white sand. Hence, another small set of experiments was conducted on red sand in order to check the classification procedure for color robustness. For this data collection the robot ran at speeds of 0.2 m/sec, 0.4 m/sec and 0.6 m/sec. The algorithm was then tested on the feature vectors from the red sand experiments using classifiers trained with only white sand data. The visual differences in white and red sand can be seen in Figure 11.

3.3. Classification Results

This section presents terrain classification results obtained using spatial frequency and texture features as described in Subsection 2.2 as well as the results achieved from the combined use of spatial frequency and texture features. This section also includes an improvement to the classification results by applying the update rule of Subsection 2.5. Results from additional experiments to show illumination robustness and color robustness of sand are discussed as well. All results presented here correspond to testing of images in the previously described test set.

3.3.1. Spatial frequency response

Table 1 shows classification results using the spatial frequency response method. The results show that spatial frequency response is an efficient terrain signature. The algorithm classified the four terrains with greater than 86.0% accuracy. Interestingly, grass and gravel are the two terrains that were most often confused because of significant variation in their spatial frequency responses. Sand has the largest classification variability, i.e., can be misclassified as any of the other terrains and conversely all of the other terrains are sometimes misclassified as sand. This is because sand terrains are sometimes as flat as asphalt and sometimes display coarser patterns that may be confused with gravel or grass.

3.3.2. Texture

Table 2 shows classification results using the textural features. This table shows that texture features are also efficient

<table>
<thead>
<tr>
<th>Tested Terrain</th>
<th>Asphalt</th>
<th>Grass</th>
<th>Gravel</th>
<th>Sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected</td>
<td>Asphalt</td>
<td>98.8%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Terrain</td>
<td>Grass</td>
<td>0.0%</td>
<td>86.0%</td>
<td>4.8%</td>
</tr>
<tr>
<td></td>
<td>Gravel</td>
<td>0.0%</td>
<td>12.8%</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>Sand</td>
<td>1.2%</td>
<td>1.2%</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

Figure 9: Outdoor terrains.

Figure 10: Small terrain patches for indoor experiments.

Figure 11: The comparison of white sand and red sand.

Figure 12: Classification accuracy comparison.
Table 2: Texture classification results

<table>
<thead>
<tr>
<th>Detected Terrain</th>
<th>Asphalt</th>
<th>Grass</th>
<th>Gravel</th>
<th>Sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt</td>
<td>99.2%</td>
<td>0.0%</td>
<td>1.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Grass</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Gravel</td>
<td>0.8%</td>
<td>0.0%</td>
<td>89.0%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Sand</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.4%</td>
<td>92.6%</td>
</tr>
</tbody>
</table>

Table 3: Combined features classification results

<table>
<thead>
<tr>
<th>Detected Terrain</th>
<th>Asphalt</th>
<th>Grass</th>
<th>Gravel</th>
<th>Sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt</td>
<td>99.4%</td>
<td>0.0%</td>
<td>1.2%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Grass</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Gravel</td>
<td>0.6%</td>
<td>0.0%</td>
<td>97.4%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Sand</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1.2%</td>
<td>97.0%</td>
</tr>
</tbody>
</table>

at identifying terrains, as all four terrains are shown to be identified at a rate of 89.0% or higher. Unlike spatial frequency response all grass samples were classified correctly. Additionally, gravel is now more easily confused with asphalt and sand since 10.6% of gravel classified as asphalt and 5.0% of sand classified as gravel. This corresponds with Figure 4, which shows that asphalt, gravel and sand have relatively close values of texture descriptors, especially asphalt and gravel.

3.3.3. Spatial Frequency and Texture Combination

Table 3 shows classification results when the feature vector consists of both spatial frequency and texture features as described in Subsection 2.4. All four terrains were classified with more than 97% accuracy, which shows that the combination of both features produces highly accurate results. Only 0.6% of asphalt was misclassified (as gravel) and grass was always classified correctly due to its significantly different texture properties compared to other terrains. Gravel is misclassified as asphalt at a rate (1.2%) which is significantly better than texture (10%) but not as good as spatial frequency (0%). This was as expected considering that spatial frequency features or texture features alone were confused by different combinations of terrains.

3.3.4. Results comparison

Figure 12 shows the results comparison for all three classification features: spatial frequency \( x_f \), texture \( x_t \) and the combination of spatial frequency and texture \( x_c \) (see (13)). All features produced satisfactory classification results, as the accuracy for each type of terrain was above 85.0%. For asphalt and sand \( x_f \) and \( x_t \) yielded almost the same results. For grass, \( x_t \) provided 14.0% better classification accuracy than \( x_f \), while the use of \( x_f \) was 4.4% more accurate than \( x_t \) for gravel. The use of \( x_t \) produced the same or higher classification accuracies, for all four terrains, than the use of \( x_f \) or \( x_t \). The terrain that yielded the most improvement by the use of \( x_t \) is gravel; the classification accuracy was improved by 8.4% over the use of \( x_t \). The use of \( x_t \) also yields higher classification accuracy for asphalt than either \( x_f \) or \( x_t \), which produced results that were above 98% accurate. These improvements can be attributed to increased characterization of the spatial terrain properties than can be obtained by either set of features individually.

3.3.5. Update Rule Results

To determine an update rule for \( x_f \), \( x_t \) and \( x_c \), the set of considered values for \( n \) was defined as \( N = \{1, 2, 3, \ldots, 20\} \), while the set of considered values of \( \eta \) was defined as \( H = \{0.5, 0.525, 0.55, \ldots, 1.0\} \). This led to obtainable values of sensitivity and robustness, such as shown in Figure 6, which shows the results corresponding to \( x_c \). Interestingly, regardless of the choice of \( x_f \), \( x_t \) or \( x_c \), there exists a pair \((n, \eta)\) with \( n \in N \) and \( \eta \in H \) that leads to perfect robustness. Since it is decided to favor robustness over sensitivity, the pair \((n, \eta)\) which corresponds to perfect robustness and maximum sensitivity is chosen for update rule implementation. For \( x_f \) and \( x_t \), this corresponds to \( n = 5 \) and \( \eta = 0.625 \), and for \( x_c \), this corresponds to \( n = 3 \) and \( \eta = 0.675 \).

To evaluate the performance of the corresponding update rules on test data, a sequence of three terrain-transitions was formed by concatenating 500 samples of the four terrains. This sequence is shown in the top bar graph (labeled true terrain) of Figure 13, which shows bar graphs of three stages of the terrain perception process. The middle stage gives the classifier results, and the bottom stage gives the results of the update rule when applied to the classifier results. All inaccuracies seen in the classifier results are removed by the update rule with a maximum delay of only 3 samples during terrain transition. Additionally, the terrain sequence produced by the update rule has an initial \( n - 1 \) sample delay, caused by the update rule waiting for \( n \) classifications, which are necessary to build the first history window. However, this delay may be negligible since the line stripper detects the terrain just ahead of the vehicle. That is, if the speed of the vehicle is slow enough, the control mode will be updated by the time the AGV begins to traverse the terrain transition.

The results shown in Figure 13 were achieved with combined spatial frequency and texture features \((x_c)\). Figures 14 and 15 show results produced with spatial frequency \((x_f)\) and texture features \((x_t)\) respectively. The update rule results are uniformly accurate for all three classifiers even though each classifier misclassified different test samples. This is a potentially important outcome as it means the choice of features may not be limited to the method that produces the highest classification accuracy. Instead, the choice of classification method can be based on other classification criteria like computational efficiency, execution time or terrain transition delay, as long as an update rule is properly determined and implemented.

3.3.6. Indoor experiments result

Figure 16 shows the result of the indoor experiments. Similar to the outdoor experiments, the accuracy for each type of terrain was above 85.0%, but interestingly the indoor experiments yielded higher classification accuracies than the outdoor experiments. The main reason is that many of the influences...
of outdoors terrain were eliminated, in particular, the non-homogeneity of the terrains and terrain roughness. For example, outdoor gravel is often not pure gravel. It frequently contains elements of grass, sand and other objects. Hence, when gravel data is collected, it is corrupted with non-gravel information. In contrast, the terrains used for the indoor experiments were more homogeneous and smoother than the outdoor terrains yielding more pure terrain signatures. In contrast to the outdoor experiment, texture provides a higher classification accuracy than spatial frequency response not only for grass but also for gravel. This indicates that texture signatures are more corrupted from the inhomogeneous nature of the outdoor terrains.

3.3.7. Red sand experiments result

Table 4 shows classification results of the experiments conducted on red sand. While the testing data originates from red sand, the terrain classification algorithm was trained with only white sand information. Features based on spatial frequency response are able to classify red sand perfectly as sand, even without training knowledge of red sand. Therefore, spatial frequency features have high color robustness. Texture classified red sand with a high accuracy of 87.3%, which is also considered a relatively high level of color robustness.
4. CONCLUSIONS

This paper presents a new method for vision-based terrain classification using close range sensing. Unlike previous research, this classification relies on spatial relationships. Two types of features were developed, spatial frequency response of the terrain profile and texture, which presents the spatial relationships between different gray levels. The range information proved to be useful in determining the spatial frequency content of the terrain profiles, while high resolution of the laser line striping allowed localized texture analysis of the terrain images. Outdoor experimental results on four different terrains show the effectiveness of the proposed method. The combined use of spatial frequency and texture features produces an overall classification accuracy of more than 97%. The results also show that while each feature has advantages in classifying some terrains and disadvantages in classifying other terrains, combining both features yields significant improvements. In addition, the indoor, light-controlled and color experiments show that the spatial frequency response and texture features have luminance robustness and color robustness.

It is also shown that the few classifier inaccuracies that occur with these types of features are easily ignored by applying an update rule before implementing control modes. A method to determine optimal values of update rule parameters using training data is also discussed. Application of this update rule not only rejects all inaccuracies for all three classifiers considered but also does it with a maximum delay of three samples between terrain transitions.

All the experiments were conducted on a low speed robot. The classification speed and sensing technology may limit the method’s application to a high speed vehicle. In the future, some experiments need to be conducted to find the maximum speed boundary for this method. In addition, the proposed approach was tested with four relatively homogeneous terrains. Therefore, more challenging experiments involving mixed terrains, surfaces with anisotropic features (e.g., concrete roadways to grooves in predefined directions) and other outdoor terrains like mud should be conducted to determine the range of applicability of the proposed approach.
5. ACKNOWLEDGMENTS

The authors thank Christoph Mertz and Martial Hebert from the Carnegie Mellon University Robotics Institute for providing the laser line stripers used in this research.

This research and development project was conducted by Florida State University and is made possible by an assistance agreement that was awarded and administered by the U.S. Army Medical Research & Materiel Command (USAMRMC) and the Telemedicine & Advanced Technology Research Center (TATRC), under Contract Number W81XWH-10-2-0024.

Also, this research was supported by the U.S. Army Research Laboratory under the Robotics Collaborative Technology Alliance program, Cooperative Agreement W911NF-10-2-0016.

References


