(less intra class distance) amongst segments in the far field as opposed to segments in the mid field. In practice, we have found that the scheme described above works quite well.

III. Experimental Setup

The experiments presented in this section are designed to explore the three main components of the approach: the segmentation algorithm, the proximity weighting model and the overall far field propagation method. We examine the system's performance under a range of parameters for both Mean Shift and graph-based segmentation and consider how the image datasets affect the performance of each segmenter. In order to evaluate the performance of our far-field labeling system we compare it to available ground/obstacle labels from Stereo and to a near-to-far Support Vector Machine (SVM) labeling approach [21][11], which uses Stereo labeled pixels from the mid field to learn to classify far field pixels.

A. Data

We tested our algorithm on six data sets, with each data set containing 100 image frames extracted from the log files recorded during live robot test runs in the DARPA LAGR program. The data sets were carefully chosen to represent different terrain and lighting conditions. The terrain varies greatly over the data sets, with ground plane varying from mulch to dirt and obstacles varying from foliage to trees and hay bales. Overall, three scenarios are considered (DS1,DS2,DS3). Each scenario is associated with two distinct data sets taken under different lighting conditions (A and B). Representative frames from the six data sets are shown in Figure 3.

We evaluated the performance of the competing algorithms by comparing their predictions with pixel level hand labeled ground truth data made available by Procopio et al. [22]. Each pixel of an image frame is labeled as one of three classes - Ground plane, Obstacle or Unknown. We compute our performance metrics only over the pixels in the far field region. This provides for a fair comparison of the far field performance of the competing algorithms.

B. Performance Metrics

To measure the performance of the competing algorithms we borrow popular information retrieval metrics of precision and recall.

Precision is the ratio of the number of relevant objects in a retrieved set to the total retrieved set. Informally, precision may be thought of as a measure of the noise in the retrieval process. It is calculated as:

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

where TP(true positives) is the number of far field pixels correctly classified as GroundPlane or Obstacle and FP is the number of far field pixels mis-classified as GroundPlane or Obstacle.

Recall is the ratio of the number of relevant objects in a retrieved set to the total number of relevant objects.Intuitively

(a) DS1A (b) DS1B

(c) DS2A (d) DS2B

(e) DS3A (f) DS3B

Fig. 3. Example frames from the six image data sets.

it is a measure of the exhaustiveness of the retrieval process. Recall is calculated using:

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

where FN is the number of far field pixels which are classified as unknown, but belong to either the Ground or Obstacle class in the ground truth data. Finally, for convenient comparison of the different algorithms we use the F_1 measure which is just the harmonic mean of precision and recall. This value is computed as follows:

$$F_1 = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{9}$$

C. Comparison Systems

a) Stereo Labels: Stereo labels are the baseline we compare against. This comparison is a basic sanity check which establishes whether there is any benefit to using the proposed algorithm in a real world scenario (Fig. 4(a)).

b) Support Vector Machines: The most common approach to far field terrain labeling is currently a near-to-far learning approach, where near or mid-field Stereo labeled samples are used to train models for labeling far-field pixels. We compare our approach to both linear and Gaussian SVMs (Figs. 4(e) and 4(f)). Both Linear [21] and Gaussian SVMs have been used for far field predictions in the past. Furthermore, Gaussian SVMs were found to be particularly well suited for this problem [11]. We use the one model per