sures available in the literature [20]. In this paper, we use the sum of minimum bin measure, where the distance between two segment histograms is defined as:

$$D(H_1, H_2) = \sum_{i=1}^{N_{bins}} I(min\{h_1(i), h_2(i)\}) \qquad (1)$$

where i indexes the bins of the histograms of the two segments. It is implicitly assumed that H_1 and H_2 have been normalized by the total number of pixels in the corresponding segments.

D. Data Modeling

We model each class as a mixture of components, where each component is the histogram(30 bins) of a contiguous stereo labeled region. The ground plane class is modeled as a mixture of histograms of contiguous regions made up of the pixels belonging to set S_g while the obstacle class is modeled as a mixture of histograms of regions made up of the elements of S_o. Figure 2 provides a pictorial depiction of the class modeling. Formally, our model may be expressed as a mixture model:

$$H_s = \sum_{k=1}^{K} \eta_{sk} H_k \qquad (2)$$

where k is the number of components and $\eta_{\mathbf{s}} = \{\eta_{s1}, \eta_{s2} ... \eta_{sk}\}$ represents the mixing weights of the different components. It is important to note that the histogram mixture is indexed by $s \in \Omega$. This implies that we can have a different model representing the same class depending on the particular s being classified. The histogram mixtures differ by having different component histograms mix according to input dependent mixing proportions η_{sk}, where $\eta_{sk} \geq 0$ and $\sum_{k=1}^{K} \eta_{sk} = 1$. Furthermore, our model employs semi parametric histograms as opposed to parametric models (such as Gaussian) for modeling the components. This allows us to abstain from making assumptions about the data generation process.

Ideally, the mixing weights would be learnt through a Expectation Maximization algorithm. However, such an iterative algorithm would have to be run for each segment which is not desirable for a near real time algorithm. Instead, we take a simpler route and assign the mixing weights in a heuristic fashion. We experiment with three different schemes for assigning the mixing weights $\eta_{\mathbf{s}}$. The first model which will be referred to as the **Global** model assumes that all components are equally important and thus allots equal weights to all components.

$$\eta_{sk}^{global} = \frac{1}{K}; \quad k = 1..K \quad \forall s \in \Omega \qquad (3)$$

At the other end of the spectrum we have the **Local** model. In this model only the nearest component to s has a weight of 1, with all other components set to zero.

$$\eta_{sk}^{local} = \begin{cases} 1 & \text{if } k \text{ is the (spatially) nearest component to } s \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

Finally, we have the **Semi-Local** model, which is a compromise between the local and the global models. The local model gives absolute importance to the closest component, while completely disregarding all other components. The global model on the other hand does not take spatial proximity into account at all. The semi-local model reconciles the two models by fitting an exponential distribution[2] on $\eta_{\mathbf{s}}$. The weight of a component decays exponentially with its spatial distance δ from the segment to be classified.

$$\eta_{sk}^{semilocal} = \frac{\frac{1}{\delta_{sk}}}{\sum_{k=1}^{K} \frac{1}{\delta_{sk}}}; \quad k = 1..K \quad \forall s \in \Omega \qquad (5)$$

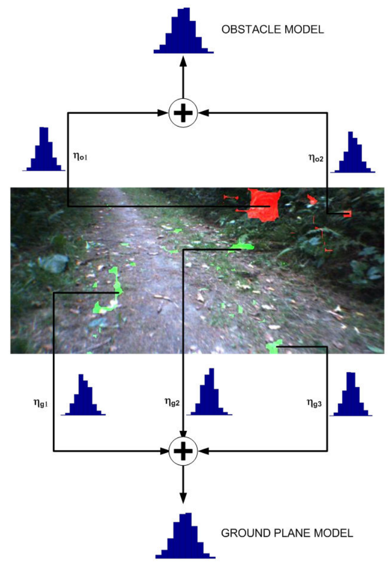

Fig. 2. Data modeling illustration.

E. Threshold Selection

The final issue of interest in the algorithm is that of determining the appropriate thresholds d_g and d_o. We select these values by segmenting the mid field region of each image frame. Segments which transgress class boundaries (as determined by stereo labels) are split such that each segment belongs solely to one class. Next, the histogram distance between all segments of the same class is computed. Further, the mean ($\bar{d}$) and standard deviation (σ) of the intra class distances is computed. Finally, the distance threshold is computed as:

$$d_{th} = \bar{d} - \sigma \qquad (6)$$

where d_{th} corresponds to either d_g or d_o.

We compute the thresholds over the mid field region because the availability of stereo labels here is considerably larger than in the far field, thereby making the above intra class distance statistics meaningful. The motivation behind choosing this threshold is that images smooth out in the far field. This smoothing causes there to be less discriminability

[2]In our experiments, we set the rate parameter of the distribution λ to 1.