A Visual-Sensor Model for Mobile Robot Localisation

Matthias Fichtner Axel Grofimann

Department of Computer Science
Technische Universitat Dresden
Dresden, Germany

1 Introduction

Due to recent advances in robot hardware, there is a great demand for vision-based robot localisation techniques [DeSouza and Kak, 2002]. We present a probabilistic sensor model for camera-pose estimation in hallways and other known structured environments. Given a 3D geometrical map of the environment, we want to find an approximate measure of the probability that a given camera image has been obtained at a certain place in the robot's operating environment. Our sensor model is based on feature matching techniques that are simpler than state-of-the-art photogrammetric approaches. This allows the model to be used in probabilistic robot localisation methods, such as Monte Carlo localisation (MCL) [Dellaert *et al,* 1999]. We have combined photogrammetric techniques for feature projection with the flexibility and robustness of MCL. Moreover, our approach is sufficiently fast to allow for sensor fusion. That is, by using distance measurements from sonars and laser in addition to the visual input, we may be able to improve localisation accuracy. We have used our sensor model with MCL to track the position of a Pioneer 2 robot navigating in a hallway. Possibly, our approach can be used also for localisation in cluttered environments and for shape-based object detection.

In the next section, we briefly describe the components of the visual-sensor model. We conclude with a discussion of experimental results.

2 Sensor Model

A visual-sensor model describes the probability of obtaining a particular camera image given the camera's pose and a geometrical map of the environment. Rather than comparing camera image and the expected view at the pixel level, we gained improved robustness using image features. We decided to use line segments because they can be detected comparatively reliably under changing illumination conditions. As world model, we use a wire-frame model of the operating environment, represented in VRML. The individual processing steps are depicted in Figure 1.

Representation. Each straight-line feature is represented as a single point (ρ, θ) in the 2D Hough space given by $\rho = x\cos\theta + y\sin\theta$, where end-points are neglected. In this representation, truncated or split lines have similar coordinates.

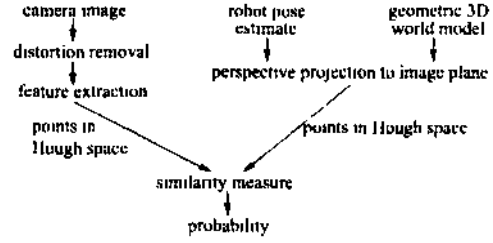

Figure 1: Processing steps of the visual-sensor model.

Similarity measures. The correspondence of image and model features is evaluated by a similarity measure. In general, such a measure may take into account the differences in orientation between corresponding line segments in image and model, or their distance and difference in length. In the following, we present two simple and efficient similarity measures. The first is solely based on the distance of line segments in the Hough space. We consider only those image features as possible matches that lie within a rectangular cell in the Hough space centred at the model feature. The matches are counted and the resulting sum is normalised. The mapping from the expectation (model features) to the measurement (image features) accounts for the fact that the measure should be invariant with respect to objects not modelled in the provided map or unexpected changes in the operating environment. Invariance of the number of visible features is obtained by normalisation. Specifically, this *centred match count* (CMC) measure s_{CMC} is defined as:

$$s_{\mathrm{CMC}} = (1/|H_e|)\,\#_{i=1}^{|H_e|} \min\left(1, \#_{j=1}^{|H_m|} m(h_{e_i}, h_{m_j})\right)$$

$$m(a,b) \equiv |\rho_a - \rho_b| < t_\rho \wedge \|\theta_a - \theta_b\| < t_\theta$$

where the predicate m defines a valid match using the distance parameters (t_ρ, t_θ) and the operator # counts the number of matches. Generally speaking, this similarity measure computes the proportion of expected model Hough points $h_{e_i} \in H_e$ that are confirmed by at least one measured image Hough point $h_{m_j} \in H_m$ falling within tolerance (t_ρ, t_θ). Note that neither endpoint coordinates nor lengths are considered here. The second measure is based on a comparison of the total length values of groups of lines, which we name *grid length match* (GLM). Split lines in the image are grouped together using a uniform discretisation of the Hough space. This method is similar to the Hough transform for straight

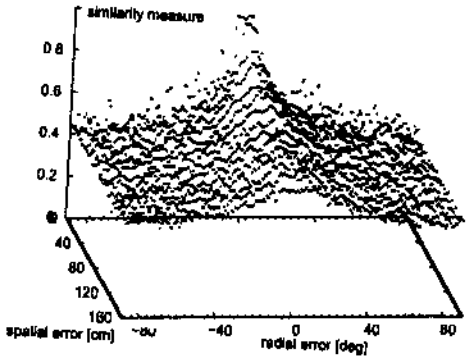

Figure 2: Performance of CMC on artificially created images.

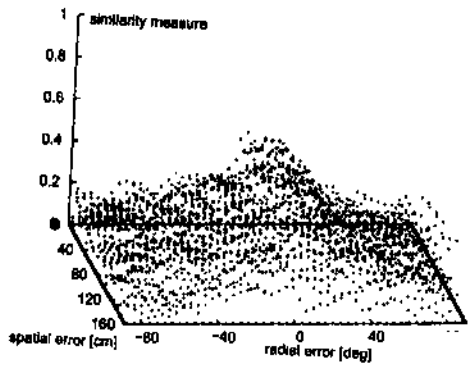

Figure 3: Performance of GLM on real images.

lines. The same is performed for line segments obtained from the 3D model. Again, the mapping is directional, i.e., the world model is used as reference, in order to obtain invariance of noise, clutter, and dynamic objects.

Obtaining probabilities. Ideally, we want a monotonically decreasing similarity measure as the pose estimate of the hypothesis departs from the actual camera pose. By averaging over a large number of different, independent situations, an abstraction from specific poses and environmental conditions is gained. To obtain commensurable values, we express the model in terms of relative robot coordinates instead of absolute world coordinates. The probability returned by the visual-sensor model is obtained from the similarity value by normalisation within the range of the pose deviation.

3 Experimental Results

We have evaluated the proposed sensor model and similarity measures in a series of experiments.

Using artificially generated images. First, we simulated the feature extraction steps from real images by duplicating the right-hand branch of Figure 1. By introducing a pose deviation Δl systematically, we investigated its influence on the similarity values. For visualisation purposes, Δx and Δy were combined into a spatial deviation Δt. For the measures CMC and GLM, over 15 million random camera poses were coupled with a random pose deviation. As shown in Figure 2, the simulation of CMC shows a unique, distinctive peak at zero deviation with monotonically decreasing similarity values as the deviation increases. For GLM, we obtained a peak that is considerably more distinctive. This conforms to our expectations since taking the length of image and model lines into account penalises incidental false matches.

Toward more realistic conditions. In order to learn the effect of distorted and noisy image data on our sensor model, we repeated the experiments above but applied distortions to the synthetically generated features before matching. The error model comprised random splitting, duplication and shifting, and addition of lines yielding typical, noisy feature maps. The results obtained for the two measures did not differ significantly from the first set of experiments.

Using real images from the hallway. As the results obtained in simulation might be questionable with respect to real-world conditions, we conducted another set of experiments using real images. We gathered images with a Pioneer 2 robot in the hallway off-line and recorded the line features. More than 3200 images were taken from 64 different (x, y) locations. Likewise as above, the sensor model was computed for pairs of poses. Results referring to the same discretized pose deviation were averaged. The result of employing the GLM measure is given in Figure 3. Again, GLM yielded are more distinctive peak compared to the curve obtained using CMC.

Application to Monte Carlo localisation. The generic interface of the sensor model allows it to be used in the correction step of Bayesian localisation methods, for example, the standard version of the MCL algorithm. In this experiment, the robot had to follow a pre-programmed route in the shape of a double loop in the corridor. On its way, it had to stop at pre-defined positions, turn to a nearby corner or open view, take an image, turn back and proceed. Upon each image capture, the weights of all samples were computed according the visual-sensor model using CMC. The preliminary results are very promising. The best hypothesis for the robot's pose was approximately at the correct pose most of the time. We noticed that rotational error has a strong influence on the degree of coincidental feature pairs. This effect corresponds to the results above, where the figures exhibit a much higher gradient along the axis of rotational deviation than along that of translational deviation. This finding can be explained by the effect of motion on features in the Hough space. Hence, the strength of our model lays at detecting rotational disagreement. This property makes it especially suitable for two-wheel driven robots like our Pioneer bearing a much higher rotational odometry error than translational error.

References

[Dellaert *et al.*, 1999] F. Dellaert, W. Burgard, et al. Using the condensation algorithm for robust, vision-based mobile robot localisation. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1999.

[DeSouza and Kak, 2002] G. N. DeSouza and A. C. Kak. Vision for mobile robot navigation: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(2):237–267, 2002.