On the Use of Disparity Maps for Robust Robot Localization under Different Illumination Conditions

Josep M. Porta **Ben Kröse**

IAS Group, Informatics Institute
University of Amsterdam
Kruislaan 403, 1098SJ, Amsterdam, The Netherlands
{porta,krose}@science.uva.nl

Abstract

In this paper, we introduce the use of disparity maps to alleviate the problem of appearance-based robot localization due to changes in illumination. We describe how is it possible to use disparity maps for appearance-based localization and we compare the results obtained using disparity maps with those obtained with other techniques commonly used to reduce the effect of illumination on images: histogram equalization and gradient-based filters. The results we present show that disparity maps are sensitive enough to rotations and translations of the robot and that they are less sensitive to changes in illumination than previously used techniques. Consequently, we show that disparity maps are a valid alternative to achieve a robust appearance-based robot localization.

1 Introduction

In the last years, much effort has been put on appearance-based robot localization [5]. In this paradigm, images obtained by the robot are compressed to get a reduced set of feature detectors. The position of the robot is determined comparing the feature detectors corresponding to the image observed by the robot with images (and, thus, feature detectors) collected at known positions.

However, images largely change due to variations on illumination and so they do the corresponding feature detectors. Consequently, illumination changes can largely degrade the performance of the appearance-based localization techniques.

One possible solution to this problem is to include in the training set images obtained in different illumination conditions. However, this is a limited solution since not all possible illumination setups can be devised when defining the training set.

A more general solution is to pre-process the images to compensate for the effect of illumination. In this line, techniques such as histogram equalization of gradient filters have been used to obtain images (and, thus, feature detectors) that are, up to a given point, illumination-independent [3, 7].

In this paper, we introduce the use of disparity maps as a source to obtain features detectors for appearance-based localization. Disparity maps are computed by matching points on two images taken by a pair of calibrated cameras. The hypothesis is that, since the two images used to define the disparity map are obtained simultaneously and, thus, with the same illumination, the resulting disparity map would be less sensitive to changes in illumination than plain images. We show that this hypothesis is valid and that a robust appearance-based localization can be performed based on disparity maps since they provide feature detectors sensitive to changes in the robot position (translations/rotations) but less sensitive to changes in illumination than the features obtained using histogram equalization and gradient filter techniques.

This paper is organized as follows. First, we formalize the appearance-based localization framework. Next, we briefly describe the three techniques to alleviate the problem of illumination we compare in this paper: histogram equalization, gradient-based filters, and disparity maps. Then, we present the results obtained with these three techniques in a real environment and, finally, we conclude summarizing our work and extracting some conclusions out of it.

2 Appearance-based Localization

Appearance-based localization departs from a training set with images $Y = (y_1, \ldots, y_m)$ taken at known positions $P = (p_1, \ldots, p_m)$. Images are linearly compressed to get a set of feature detectors for each image $F = (f_1, \ldots, f_m)$, where

$$f_i = W y_i.$$

The projection matrix W is determined off-line applying Principal Component Analysis (PCA) to find out the eigenvalues and eigenvectors[1] of the training set in the space of images. The rows of the projection matrix W are the eigenvectors corresponding to the n largest eigenvalues.

In the on-line execution, we aim at estimating the probability on the robot's position at time t $p(x_t)$. This is usually done assuming that the environment is Markovian and updating $p(x_t)$ from $p(x_{t-1})$. In this update, we use the training set as well as the feature detectors of the image taken by the robot at time t, f_t, to define the probability $p(f_t|x_t)$ that is called the *sensor model*. For instance, Vlassis *et al.* [11] introduce a nearest-neighbor approach to represent the sensor model.

Problems arise when illumination conditions in the on-line execution are different from those when obtaining the training set. This produces wrong matches (i.e., wrong nearest neighbors) of the feature detectors of the current observation with those in the training set. This leads to a wrong sensor model and, thus, to a wrong update of $p(x_t)$.

The solutions for this problem we explore next are based on processing images trying to reduce the effect of different illumination conditions in the resulting feature detectors.

3 Histogram Equalization

Histogram equalization [2] is a gray level transform that aims at producing an image with equally distributed brightness levels over the whole brightness scale. The gray value range equalization is useful for improving the image quality if the original image covers only a part of the full gray scale (what is, in most of the cases, caused by low scene illumination). The drawback of this technique is that, in case of good gray value dynamics on the input image, it can lead to quality looses in form of sharp edges.

Let $H(p)$ be the histogram of the input image with $p \in [p_0, p_k]$ one of the gray level in the image. The intention of the histogram equalization is to find a monotonic brightness transform $q = T(p)$ such that the histogram of the output image, $G(q)$, is uniform over the hole brightness scale $[q_0, q_l]$ (in many cases, $[q_0, q_l] = [0, 255]$). The histogram of a gray level image can be considered as a discrete probability density function. With this formalism, the monotonic property of the transform T implies

$$\sum_{i=0}^{j} G(q_i) = \sum_{i=0}^{j} H(p_i).$$

Since, in ideal circumstances, we want G to be uniform, we have that

$$\sum_{i=0}^{j} G(q_i) = \frac{M\,N\,(q_j - q_0)}{(q_l - q_0)}$$

for an image of size $M \times N$ pixels. Consequently, we have that the desired transformation $T(p)$ can be defined as

$$q_j = T(p_j) = q_0 + \frac{(q_l - q_0)}{M\,N} \sum_{i=0}^{j} H(p_i).$$

Since q_j should be set to the closest integer value, we do not obtain a perfectly uniform histogram G.

Figure 1 shows an image taken with the robot's cameras and the same image after equalization.

Figure 1: *Plain image (top) and the same image after histogram equalization (down).*

4 Gradient-based Filters

To apply a gradient filter we have to considerer an image as a function $I(x, y)$ where the domain of I is the set of pixel locations, and the range of I is the pixel intensity.

The gradient of the intensity image $I(x, y)$ is

$$\nabla I(x, y) = \left[\frac{\partial I(x, y)}{\partial x}, \frac{\partial I(x, y)}{\partial y} \right].$$

At each pixel the gradient is a two-dimensional vector that points in the direction of the maximum intensity increase (so, it will tend to be perpendicular to strong edges in the image).

[1]Eigenvectors are many times called eigenimages in the field of appearance-based localization.

Figure 2: *Plain image (top) and the same image after applying the Sobel filter (down).*

We can compute $\partial I(x,y)/\partial x$ and $\partial I(x,y)/\partial y$ by convolving $I(x,y)$ with carefully chosen filters that approximate derivatives. Typically, these filters blur in one direction and find pixel differences in the perpendicular direction.

A well known derivative filter is the Sobel one [10] that uses the convolution

$$\begin{array}{ccc} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{array}$$

to approximate $\partial I(x,y)/\partial x$ and

$$\begin{array}{ccc} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{array}$$

for $\partial I(x,y)/\partial y$.

After applying these convolutions, we can take the norm of the gradient vector $\nabla I(x,y)$ to define the filtered image.

Figure 2 shows the result of applying the Sobel filter to an image. As expected, the result of the filter is to enhance the edges of the image. Other gradient-based filters such as the Prewitt filter [8] or second-order edge detectors [6] yield to similar results.

5 Disparity Maps

We can determine a disparity map matching points in images taken by a pair of calibrated cameras. Given a single image, the three-dimensional location of any visible point Q must lie on the

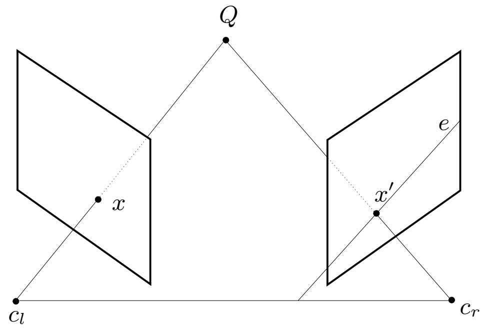

Figure 3: *Basic elements of epipolar geometry with two cameras.*

straight line that passes through the center of projection of the camera c and the image of the point x (see figure 3). The determination of the intersection of two such lines generated from two independent images is called triangulation and provides the 3-D position of that point w.r.t the cameras.

Clearly, the determination of the scene position of an object point through triangulation depends upon matching the image location of the object point in one image to the location of the same object point in the other image. At first it might seem that correspondence requires a search through the whole image, but the epipolar constraint [1] reduces this search to a single line. For instance, in figure 3, the point on the right image corresponding to the point x on the left image need to be searched only on the epipolar line e. We can go even further: if we know the range of possible depths for point Q, we can limit the search to a segment on the epipolar line.

For each pixel x in one of the images we have to search for a correspondent point along the epipolar line. Usually, the correspondence is done by comparing areas around pixel x with areas around each candidate pixel x'. The most similar pixels x and x' are assumed to correspond to different projections of the same point Q in the scene. If the images planes for the two cameras are co-planar, the distance r from the scene point Q to the cameras can be computed as

$$r = \frac{b\,f}{d_l - d_r},$$

where b is the baseline (distance between the two viewpoints), f is the focal length of the cameras, d_l is the horizontal distance from the projected point x to the center of the left image, and d_r is the same for the right image (see figure 4). The difference $d_l - d_r$ is called disparity. Since, in many cases, the baseline b and the focal length f are constant, the disparity directly relates with the distance r. For this reason, instead of working with depth maps it is enough to use disparity maps.

The stereo algorithm we use [4] applies many

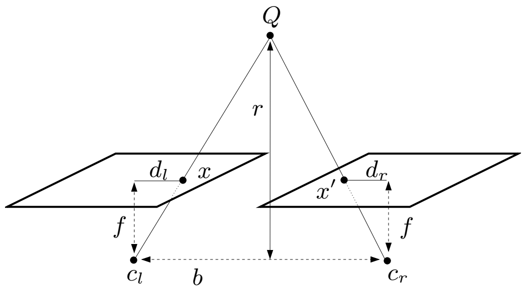

Figure 4: *Elements in the disparity computation.*

filters in the process to determine the disparity map both to speed up the process and to ensure the quality of the results. For instance, if the area around pixel x is not textured enough it would be very difficult to find a single corresponding point x': we are more likely to end up with many points x' with almost the same probability of being the corresponding point of x. For this reason, pixels on low textured areas are not even considered in the matching process.

The result of this and other filtering processes is to produce a disparity map where many pixels do not have a disparity value. To be able to apply usual PCA to disparity images we have to replace missing values with some value in the range of possible disparities. For the results presented in this paper, we just use the average value for each pixel for all the training images. We proceed in this simple way since our main objective here is to show the feasibility (and the utility) of using disparity maps for appearance-based localization. To deal with missing values in a more principled way we can, for instance, use a expectation-maximization algorithm to determine the most likely value for the missing pixels. The advantage of doing this is that it can be integrated with the eigenvector computation [9]. In any case, improving the treatment of missing values would only result in better performance that the one we report on this paper.

Image 5 shows a plain image and the corresponding disparity map. We can observe that many pixels on the image (up to 40% in many cases) have missing values (the light gray areas in the figure).

6 Experiments and Results

To test the invariance of the different methods to changes in illumination we performed the following experiment. We acquired three sets of images in the same 4×5 meters environment, but with three different lighting conditions: using tube lights, using bulb lights and using natural light (opening the curtains of the windows placed all along one wall of the lab). For each illumina-

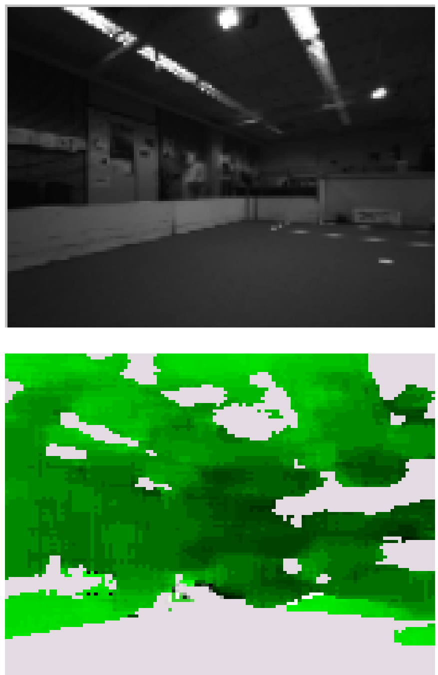

Figure 5: *Plain image (top) and the corresponding disparity map (bottom). In the disparity map, dark pixels correspond to points that are far away from the robot and light gray areas are missing values.*

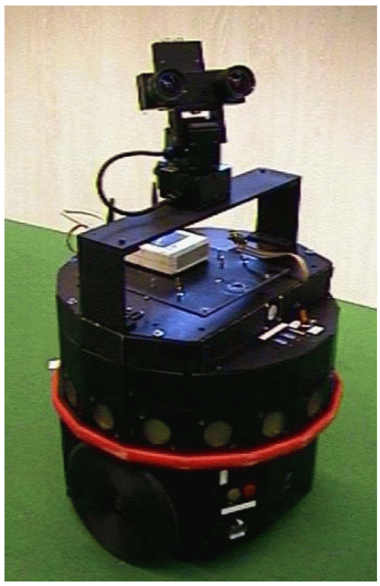

Figure 6: *Our robot with the stereo head used to take images and disparity maps.*

Image Process	Illumination Setups			Average
	Tube Lights	Bulb Lights	Natural Light	
Plain Images				
Translations	0.52 (0.23)	0.40 (0.28)	0.34 (0.24)	0.42 (0.26)
Rotations	1.47 (0.45)	0.69 (0.36)	0.70 (0.38)	0.95 (0.54)
Light change	-	1.55 (0.52)	1.58 (0.52)	1.57 (0.52)
Hist. Equalization				
Translations	0.55 (0.23)	0.70 (0.28)	0.63 (0.26)	0.63 (0.27)
Rotations	1.48 (0.44)	1.33 (0.36)	1.39 (0.46)	1.40 (0.43)
Light change	-	1.05 (0.32)	1.16 (0.28)	1.10 (0.31)
Gradient Filter				
Translations	0.77 (0.32)	0.78 (0.28)	0.69 (0.22)	0.75 (0.28)
Rotations	1.40 (0.52)	1.21 (0.35)	1.17 (0.30)	1.26 (0.42)
Light change	-	0.87 (0.18)	0.95 (0.23)	0.91 (0.21)
Disparity Map				
Translations	0.76 (0.40)	0.90 (0.44)	0.88 (0.49)	0.85 (0.45)
Rotations	1.31 (0.72)	1.42 (0.93)	1.38 (0.81)	1.37 (0.82)
Light change	-	0.55 (0.21)	0.58 (0.22)	0.56 (0.21)

Table 1: *Mean (and standard deviation in parenthesis) of the relative change of the feature detectors using different image processing techniques and in different illumination conditions.*

tion setup, we collected images every 50 cm. (both along X and Y) and every 10 degrees. This makes a total amount of about 1300 images per illumination setup. We used the set of images obtained with tube lights to determine a projection matrix W with 20 projections vectors. The two other sets of images (the one obtained with bulb lights and the one with natural light) were used as test sets. These two tests sets provide changes both in the global intensity of the images and in the distribution of light sources in the scene (that is the situation encountered in real applications).

We analyzed the sensitivity of the feature detectors to the different factors that can modify them: translations, rotations and changes in illumination. We compute the relative difference in the feature detectors as

$$d_{a,b} = \frac{\|f_a - f_b\|}{\|f_a\|},$$

with f_a and f_b the set of feature detectors corresponding to images a and b respectively.

To assess the effect of translations on the feature detectors, we evaluated the average of $d_{a,b}$ for each couple of images (a, b) taken with the same orientation and the same lighting conditions but at adjacent positions. Considering only difference in feature detectors for positions that are close each other gives us an estimation of the minimum change in feature detectors due to translations. For rotations, we computed the average of $d_{a,b}$ for each couple of images (a, b) taken at the same point and with the same illumination but with adjacent orientations. Finally, to measure differences due to illumination conditions, we computed the average of $d_{a,b}$ with a an image take with the training illumination (i.e., using tube lights) and b the image taken at the same position and orientation but with a different lighting setup. Table 1 shows the results we obtained for the experiment just described. In this table, we can see that the change in feature detectors due to rotation is larger in all cases than the variations due to translations. This is normal since a rotation of the cameras (even if it is of few degrees) produce more important changes in the images than those produced by a small translation.

Image Process	Illumination vs Translations
Plain Images	3.73
Hist. Equalization	1.75
Gradient Filter	1.21
Disparity Map	0.65

Table 2: *Ratio of average feature detector variation due to illumination vs. the changes due to translations.*

In table 2, we show the ratio of the variation of the feature detectors due to illumination changes w.r.t. the variation due to translations. This ratio gives us and idea about the importance of the feature detector change caused by illumination: the greater the ratio, the larger the effect of illumination w.r.t. the effect of translations and, thus, the larger the possible error in localization due to illumination changes. We can see that, as expected, using processed images the ratio decreases considerably compared with the ratio using plain images (images without any illumination correction). In the case of disparity maps, this ratio is the smallest one meaning that disparity is the best of the three image processing techniques we tested, as far as independence of illumination is concerned. However, we observe that, in all cases, the standard deviation of the variation of the feature detector obtained from disparity maps is larger than that using other techniques. This means that the feature detectors obtained disparity maps are usually, but not always, the best ones.

7 Conclusions

In this paper, we have surveyed three techniques that aim at reducing the effect of illumination in appearance-based robot localization. Two of them (histogram equalization and gradient filters) has been previously used for appearance-based localization and we introduce the novelty of using disparity maps.

The results we have presented show that disparity maps are a good option to achieve robust robot localization using the appearance-based framework since they provide feature detectors that are less sensible to changes in the lighting conditions than feature detectors obtained with other techniques.

Histogram equalization and gradient filters work well when we have changes in the global illumination but they do not deal properly with different distributions of light sources. On the other hand, disparity maps are more consistent over changes in the number and in the position of the light sources. This is because only reliable correspondences are taken into account when defining the disparity map and those reliable matches are likely to be detected in almost all lighting conditions.

The good results achieved using disparity maps cames at the cost of using a more complex hardware (we need not only one camera but two calibrated ones) and software (the disparity computation process is more complex than the histogram equalization and the gradient filter processes).

Acknowledgments

This work has been partially supported by the European (ITEA) project *"Ambience: Context Aware Environments for Ambient Services"*.

References

[1] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.

[2] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, 1992.

[3] M. Jogan, A. Leonardis, H. Wildenauer and H. Bischof, "Mobile Robot Localization under Varying Illumination", *Intl. Conference on Pattern Recognition ICPR02*, Quebec, Canada, 2002.

[4] K. Konolige, "Small Vision System: Hardware and Implementation", *Eighth International Symposium on Robotics Research* Japan, October, 1997.

[5] B. J. A. Kröse, N. Vlassis, R. Bunschoten and Y. Motomura, "A Probabilistic Model for Appearance-based Robot Localization", *Image and Vision Computing*, 19(6), 381-391, April, 2001.

[6] D. Marr and E. C. Hildreth, "Theory of Edge Detection", *Proc. Roy. Soc. London*, 23:441-460, 1990.

[7] H. Murase and S. K. Nayar, "Illumination Planning for Object Recognition using Parametric Eigenspaces" *IEEE Transactions PAMI*, 12:1219-1227, 1994.

[8] J. M. S. Prewitt, "Object Enhancement and Extraction", In B. S. Lipkin and A. Rosenfeld, editors, *Picture Processing and Psichopictorics*, Academic Press, 1970.

[9] S. Roweis, "EM Algorithms for PCA and SPCA", *Advances in Neural Information Processing Systems*, Vol. 10, The MIT Press, 1998.

[10] I. Sobel, "An Isotropic 3×3 images gradient operator", In H. Freeman, editor, *Machine Vision of Three-Dimensional Scenes*, pages 376-379, Academic Press, 1990.

[11] N. Vlassis, B. Terwijn and B. J. A. Kröse, "Auxiliary Particle Filter Robot Localization from High-Dimensional Sensor Observations", *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington D.C., 2002.