

## USING COLOR IMAGE FEATURES IN DISCRETE SELF-LOCALIZATION OF A MOBILE ROBOT\*

WŁODZIMIERZ KASPRZAK<sup>†</sup>, WOJCIECH SZYNKIEWICZ<sup>†</sup>

<sup>†</sup>Institute of Control and Computation Engineering, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland, Email: [W.Kasprzak,W.Szynkiewicz]@ia.pw.edu.pl, Fax: (22) 825 3719

**Abstract.** Natural landmarks are assumed to exist in the environment. Global color image features are extracted from sensor data to feed the robot's self-localization approach. The color features correspond to natural landmarks, that are learned by the navigation sub-system. During the localization process, which is a Bayes filtering of a Markov environment, the posterior probability density over possible discrete robot locations (the belief) is recursively computed. The approach was tested to provide robust results under varying scene brightness conditions and small measurement errors.

**Key Words.** autonomous navigation, Bayes filtering, visual sensor.

### 1. INTRODUCTION

The localization process of an autonomous robot takes as input a previously acquired map, an estimate of the robot's current pose, and a set of sensor data acquired in current pose, and it produces as output a new estimate of the robot's pose [1, 4]. Obviously, any input data for the localization process may be incomplete and distorted by noise or errors. In generally, pose means the position and orientation of the robot in the world coordinates or global map.

The vision data is acquired by a passive sensor, i.e. a camera does not influences the environment by its measurement process. This kind of sensor is especially applicable for indoor navigation in environments, that are populated by humans, i.e. offices, hospitals, museums, etc. [3]. Additionally, image processing methods can rely on natural landmarks, whereas this case for the active sensor devices has started to be studied only recently [2]. The use of image analysis methods in robot navigation has been intensively studied over the past 30 years [6, 9, 8, 10]. In this paper we focus on general image features, that could be relatively

insensitive to changing lighting conditions, but at the same time, can be relatively easy computed, to be obtained in real-time by a simple processing unit.

The general scheme of our self-localization approach is described in an accompanying paper [7]. In theoretical terms the localization process is equivalent to a Bayes filtering of a finite environment satisfying the Markov condition, i.e. past and future data are conditionally independent if one knows the current state. During the localization process the posterior probability density over possible discrete robot locations (the belief) is recursively computed. The use of different global features of monochromatic images was tested to provide robust results if the brightness of observed scene is constant or it can be compensated [7].

In this paper, we describe a detailed algorithm for the discrete self-localization scheme and we propose and test a more robust set of image features, based on color information. In sec. 2 the detailed algorithm for self-localization is presented. Sec. 3 defines two alternative sets of global color image features, that serve as "visual" measurements for the localization process. Test results are given in sec. 4.

---

\*) This work was supported by a Research Grant of the Faculty of Electronics and Information Technology at Warsaw University of Technology, Poland.

## 2. THE SELF-LOCALIZATION ALGORITHM

The general discrete self-localization scheme [7], based on Bayes filtering of a Markovian environment, is also called *state condensation* or *particle filtering* [5]. It assumes, that the number of states can be limited to a finite number. Only then is computationally feasible to estimate the probability distribution over states.

By *belief* we denote the pdf of states upon the condition of a sequence of observations (measurements  $m_t$ ):

$$\forall s^k : Bel_t(s^k) = p(s_t^k | m_t, m_{t-1}, \dots, m_{t-n}) \quad (1)$$

In the learning phase the system should acquire two a priori pdf's:

1. The a priori conditional pdf of measurement upon state, i.e. for each discrete state  $s \in S$  and possible measurement vector  $m$  to determine the pdf:  $p(m|s)$ ;
2. The a priori pdf of state transition

$$p(s_{t+1}^k | s_t^l, \dots, s_0^i) = p(s^k | s^l) \quad (2)$$

where  $s_t^l, \dots, s_0^i$  is the history of past best belief states. In autonomous navigation the action performed by the vehicle or camera are usually known, due to the odometry. Hence, this knowledge can be incorporated into the state condensation scheme - for each pair of states  $s^k, s^j$  and each possible action  $a$  to determine the pdf of state transition with respect to action:  $p(s^k | s^j, a)$ .

The discrete self-localization algorithm consists of the initialization step and a main iterative belief "refinement" step with sub-steps of: *belief prediction, stochastic diffusion, measurement and modification of belief* (the reaction onto the measurement).

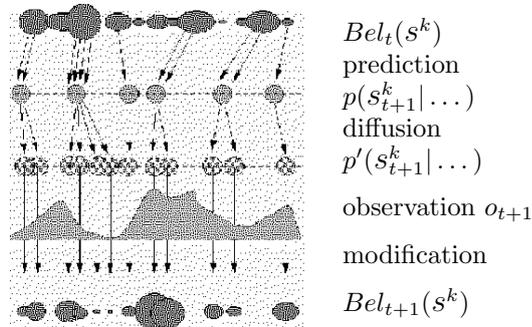


Fig. 1. One iteration of belief refinement.

### The algorithm of the self-localization process

1. Get the goal state.

2. Initialization of a default belief state at  $t = 0$  (for example by a uniformly distributed pdf)  
 $Bel_0(s^k) = p(s_0^k | H_0)$ .

3. REPEAT until the goal state is not reached:

- (a)  $t = t + 1$ ;
- (b) find the current best state:  
 $s_{t-1}^* = \operatorname{argmax} p(s_{t-1} | H_{t-1})$ , where  
 $H_{t-1} = (s_{t-1}, m_{t-1}, s_{t-2}, m_{t-2}, \dots, s_0, m_0)$   
 is the history of past belief states and measurements;
- (c) determine and perform the next action resulting from minimization of the distance between current best state and the goal state;
- (d) as the current action  $a_t$  and the a priori pdf  $p(s_t | s_{t-1}, a_t)$  are known the predicted belief state at time  $t$  can be computed

$$\widehat{Bel}_t(s^k) = \sum_s [p(s_t^k | s_{t-1}, a_t) p(s_{t-1} | H_{t-1})]$$

- (e) acquire the measurement  $m_t$  at new position.
- (f) with the a priori pdf  $p(m_t | s_t)$  modify the belief state at time  $t$ :

$$Bel_t(s^k) = p(s_t^k | H_t) = c_t p(m_t | s_t) \widehat{Bel}_t(s),$$

where  $c_t$  is the current normalization coefficient (the sum of belief state distribution should be equal to 1).

In our experiments the camera was fixed mounted on the mobile platform, which was moved along two directions only, without performing a rotation. Hence only two degrees of freedom of the camera were allowed: a translation along the X and Z axes by unit steps. A single image corresponds to one particular view of the scene (see fig. 2). The measurement system in this case is performed by an image analysis system, which detects a measurement vector for each image.

## 3. GLOBAL IMAGE FEATURES

In a previous paper ([7]) we have implemented three types of global measurements in monochromatic and color images: (MeanVar) the three mean and three standard deviation values of the R, G and B channels of the image; (FFT6) the modules of first 6 components of a Fourier transform of the intensity image; (Hist6) the three dominating color components in the image with their density values. They are of simple nature and usually they do not provide unique values for all possible states (views). But they are of sufficient quality (under assumption of constant scene illumination) to be used in a localization process, that takes a sequence of belief refinement steps.

### 3.1. The color feature vector

Every pixel in an RGB-color image can be mapped to a point in the 3-D color space. The measure-

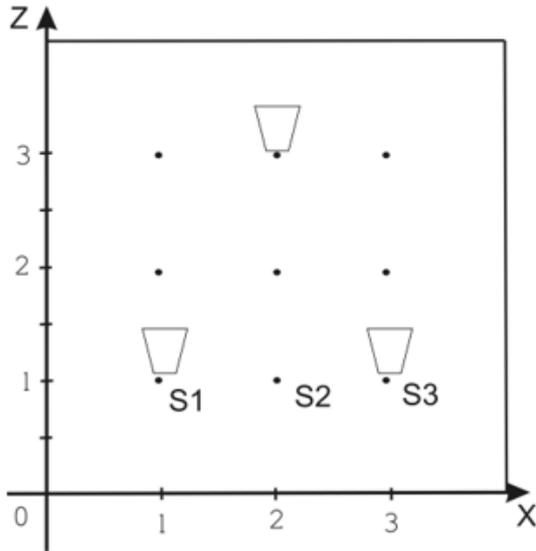


Fig. 2. The distribution of pose localizations (top view): there is a finite number of positions on the floor expressed in the OXZ coordinate system.

ment process for each image needs first to obtain a 2-D histogram of relative color densities. The individual cell of this histogram is indexed by the relation pair  $[R(p)/B(p), G(p)/B(p)]$ . In addition to the density of each "relative color" region we compute its enlargement, i.e. the boundary box of regions of such color, and its mass center position in the image.

The feature vector consists of these data items for some number (e.g.  $N = 6$ ) of pixel colors, that are selected from the 2-D histogram on the base of their largest densities in the image. Especially in this paper, we consider two sets of features:

- (set C1) only the  $N$  color indices ( $R/G, G/B$ ) and their largest densities are considered;
- (set C2) additionally to the  $N$  color indices with largest densities, the feature vector contains the boundary boxes of regions with every selected color and the mass centers of such regions.

The results of the measurement process are stored in a matrix rather than a vector. The feature matrix consists of  $N$  vectors of following  $p = 7$  values:

$$(R/B_i, G/B_i, \rho_i, \mu X_i, \mu Y_i, \Delta X_i, \Delta Y_i), i = 1, \dots, N$$

where  $(R/B_i, G/B_i)$  is the color with the  $i$ -highest density,  $\rho_i$  - its density in the image,  $(\mu X_i, \mu Y_i)$  - the center and  $\Delta X_i, \Delta Y_i$  the enlargement of image regions consisting of such pixels. The set C1 is a subset of the C2 set (i.e. for C2 the parameter  $p = 3$ ).

### 3.2. Learning the a priori pdf

The a priori pdf  $p(m|s)$  should be computed during the learning phase. But the number of possible

measurement vectors is very large, even if we restrict the measurement items to be discrete valued only. In practice we compute this pdf during the active phase of the localization process.

In the learning phase we only compute and store the feature vectors associated with each discrete state.

During the active localization phase the feature vector of current view is detected.

The a priori pdf  $p(m_{k+1}|s)$  is defined according to the difference of both measurement vectors: the current measurement  $m_{k+1}$  at time  $k + 1$  and the stored measurement  $m_s$  for  $\forall s \in s$ . The conditional probability density is modelled by a 1-D Gaussian normal distribution, with its mid point corresponding to the zero value of a weighted difference

$$\sum_{i=1}^{N \times p} w_i |m_{k+1}^i - m_s^i|^2$$

(where  $w$  is a weighting vector that scales the expected ranges of particular feature elements to some common level).

## 4. TEST RESULTS

Several laboratory scene were available for testing - two of them are shown in figure 3. From the point of view of applied general intensity-based or color-based features, the structure of the scene seems not to be so crucial as it may be for specialized features. A scene which contains nearly a single red table on the wall (fig. 4(a)) (in fact a single artificial marker) should allow more precisely to determine the state than a scene with more structure (fig. 4(b)), but only if one knows what kind of specialized feature to search for. In our localization approach both types of scenes are well managed. Below we report some quantitative results obtained for the localization process in the less structured laboratory scene.

Besides the new two color feature sets C1 and C2 we also tested the previous feature schemas of *MeanVar*, *FFT6* and *Hist6* for these scenes.

Table 1. The minimum, maximum and average 'p-values' obtained for all pairs of states during the learning phase on base of the original scene images.

p-val	C1	C2	M.Var	Hist6	FFT6
min	0.166	0.092	0.014	0.006	0.017
aver	0.658	0.704	0.462	0.595	0.471
max	1.000	0.998	0.999	0.999	0.998

The illumination conditions of the scene have been



Fig. 3. Examples of three views in different robot locations of two real laboratory scenes.

changed between the learning process and the active localization process. Additional problem was caused by the fact, that we are able to estimate robot's positions by odometry with finite accuracy only. This causes the measurement of images, that are shifted by several pixels if compared to appropriate images, applied for state distribution learning.

To achieve a statistically sufficient amount of results, many repetitions of the self-localization process have been run with changing start and goal states.

#### 4.1. Statistics of the feature vectors

Let us first verify the statistical characteristics of the proposed image feature vectors.

The *p-value* of two distributions expresses the correctness of a hypothesis, that both distributions are statistically equivalent. If the *p-value* is equal to zero, then the above hypothesis is wrong and both features can be treated as being different. This definition can be extended to more than two distributions.

Table 1 summarizes the results of 'p-val' computed for all pairs of the measurement vector (pairs of states of the original scene). The values of 'p-val' near zero indicate that the features are distinct. From this point of view both color feature sets *C1*, *C2* are less distinctive than the *MeanVar*, *Hist6*

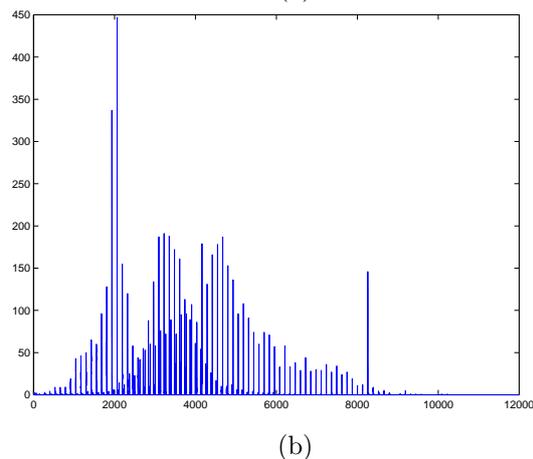
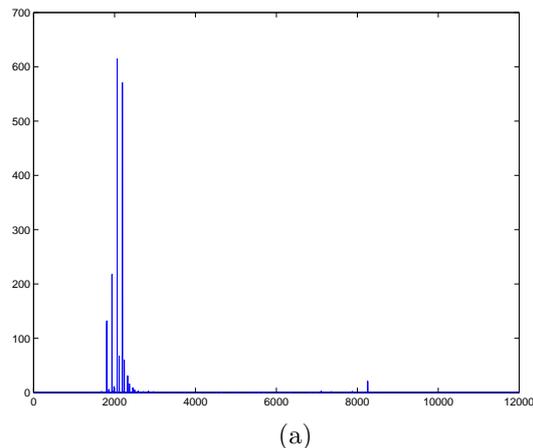


Fig. 4. The color histograms of two scenes: (a) with single red marker, (b) with more structure.

and *FFT6* sets. At the first view this may require a longer navigation sequence if we use the color features *C1* or *C2*. But the advantage of proposed color features is their insensitivity to small illumination and positional changes. This was not the case with the other three sets of features, where we required a global scene illumination sensor to compensate the changes of scene intensity.

Table 2 summarizes the relative errors between learned features and corresponding measured features, when disturbances of the measurement process appeared:  $|m_k - m_s|/|m_s|$ . The first feature vector of every pair corresponds to the state of the original scene, and the second vector - to the compatible state in the real scene. The real scene views may be shifted against-the-other by few pixels, and also the color can be nonlinearly changed. A zero-valued error means, that the features for learned view and real view are the same. This table documents, that the feature set *C2* can compensate small measurement errors of both kind, i.e., they perform better in this matter than the global intensity based feature sets. Only slightly worse if compared to *C2* was the performance of the other color set *C1*.

Table 2. The minimum, maximum and average relative feature vector errors obtained for the C2 set in case of measurement disturbances: (8,8) means position error in pixels and B10 means an increase of the blue color component by 10%.

Relative error	(0,0) +B0	(4,4) +B0	(8,8) +B0	(8,8) +B10
minimum	0.0000	0.0136	0.0176	0.1233
average	0.0000	0.1361	0.1221	0.2946
maximum	0.0000	0.4091	0.3447	0.6550
Relative error	(0,0) +B10	(0,0) +B25	(0,0) +B50	(8,8) +B25
minimum	0.0768	0.0375	0.0760	0.1041
average	0.2756	0.2644	0.2930	0.2795
maximum	0.6509	0.4799	0.6247	0.4803

Table 3. The rate of successful runs of the self-localization process (for 100 tests in total) and the average path length, under different measurement errors; i.e. (8,8) means position error in pixels and B10 means an increase of the blue color component by 10%.

	(0,0) +B25	(0,0) +B50	(8,8) +B0	(8,8) +B10	(8,8) +B25
Success rate	100	75	100	100	75
Aver. path	4.25	5.0	4.25	4.5	5.2

#### 4.2. The quality of self-localization

For each measurement method we have run the self-localization process 100 times, with randomly chosen start and goal states. A particular self-localization process is illustrated in figure 5. At the start point the belief state distribution is an uniform distribution. After 2-4 steps the appropriate state that corresponds to the real position can already be selected - the belief state value for such state dominates already the beliefs of remaining states.

In table 3 we provide the data that illustrates the correctness (quality) of self-localization tests if the C2 set is applied under different measurement disturbances. From this table it can be concluded, that up to a 25 % error of the blue component or up to a position error of (8 x 8) pixels a 100 % success of the self-localization process was achieved.

#### 5. SUMMARY

Two methods for global color feature detection in color images were proposed and its use as the measurement step in a discrete self-localization pro-

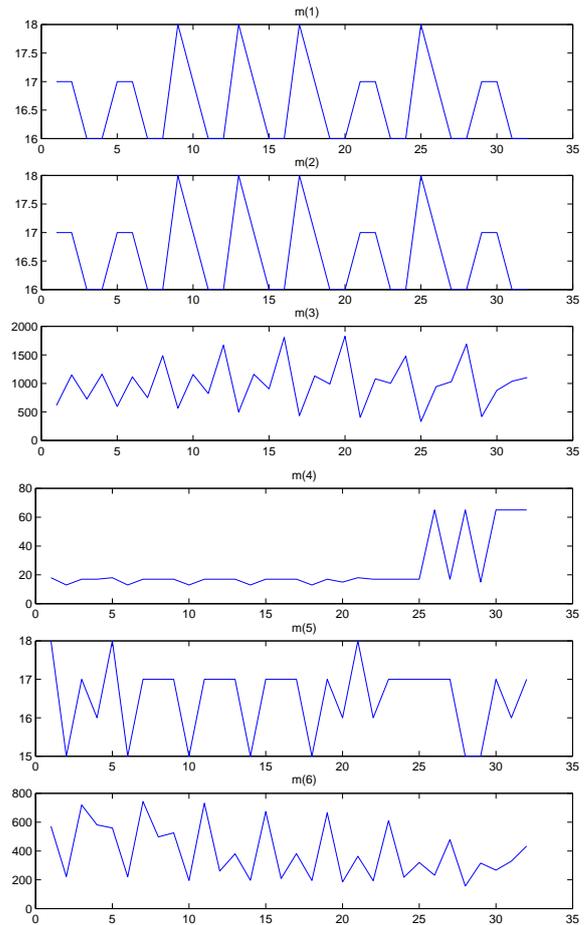


Fig. 5. The distribution of the 6 measurement features of the C1-set corresponding to all 32 discrete states of the test scene.

cess was experimentally tested.

It was shown that even for natural scenes with changing illuminations and small perturbations of the odometry data, the use of even a small set of image features, expressing only global information of a particular view, is sufficiently robust.

In more complex scenes the detection of discrete image features is preferred, although this is out of scope of this paper.

#### 6. REFERENCES

1. J. Borenstein, H. Everett, L. Feng: *Navigating Mobile Robots*. Wesley, Mass., 1996.
2. Brown R.G., Donald B.R.: Mobile Robot Self-Localization without Explicit Landmarks, *Algorithmica*, Springer Publ., vol. 26 (2000), 515–559.
3. W. Burgard, A. Cremers, D. Fox, D. Hahnel, G.Lakemeyer, D. Schulz, W. Steiner, S. Thrun: Experiences with an Interactive Museum Tour-Guide Robot. *Artificial Intelligence*, vol. 114 (1999), No. 1-2, 3-55.
4. Fox D., Burgard W., Thrun S.: Markov Lo-

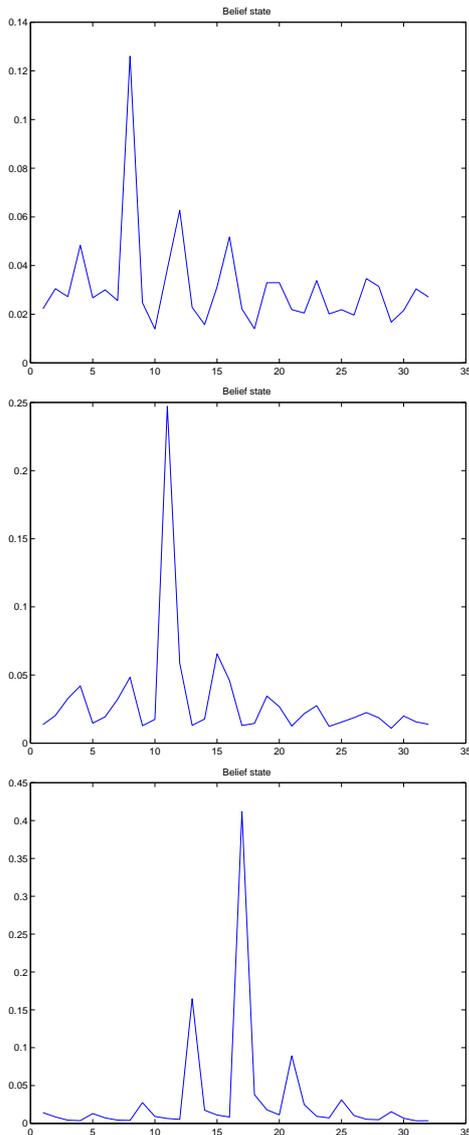


Fig. 6. Illustration of the self-localization process: after the first update (the belief of best state = 0.14), after the 2nd (belief = 0.25) and the final belief state distribution (the best belief is 0.42).

calization for Mobile Robots in Dynamic Environments, *Journal of Artificial Intelligence Research*, vol. 11(1999), 391-427.

5. Fox D., Thrun S., Burgard W., Dellaert F.: Particle Filters for Mobile Robot Localization. In: Doucet A., DeFreitas N., Gordon N. (Eds)- *Sequential Monte Carlo Methods in Practice*. Springer Publ., Berlin, etc., 2000.
6. Kasprzak W.: Adaptive computation methods in image sequence analysis. *Prace Naukowe - Elektronika*, No. 127 (2000), Warsaw Univ. of Technology Press.
7. Kasprzak W., Szykiewicz W.: A method for discrete self-localization using image analysis. *Proceedings of ROMOCO'2002*, Poznan Univ. of Technology Press, 2002, 644-649.
8. Thorpe C. (ed.): *Vision and Navigation: The*

*Carnegie Mellon Navlab*. Kluwer Academic Publ., Norwell, Mass., 1990, 25-38.

9. Masaki I.: *Vision-based Vehicle Guidance*. Springer, New York, Berlin-Heidelberg etc. 1992.
10. Maurer M., Behringer R., Thomanek F., Dickmanns E.D.: A compact vision system for road vehicle guidance. *13th Int. IAPR Conference on Pattern Recognition*, Vienna, Aug. 1996, 313-317.