

COMPARING GLOBAL MEASURES OF IMAGE SIMILARITY FOR USE IN TOPOLOGICAL LOCALIZATION OF MOBILE ROBOTS

Syeda Nusrat Ferdous, Andrew Vardy, George Mann and Ray Gosine*

Intelligent Systems Lab, Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, NL, Canada
*nusrat@enr.mun.ca

ABSTRACT

In this paper, two global measures of image similarity are compared for use in topological localization of mobile robots. A catadioptric vision system was used as the sensor. Our localization system employs a discrete Bayes filter. Experimental results with real images of an office environment indicate the robot can localize itself from the memorized training images even with the modification of the environment.

Index Terms— Localization, Bayes filter, omnidirectional images, mobile robot.

1. INTRODUCTION

In this paper, we compare two global techniques of image similarity by performing image-based localization on real images of an office environment. Image-based localization consists of matching the current image captured by the mobile robot with the training images stored in the memory of the robot. Visual sensors, in comparison with other sensors used in mobile robot localization (e.g. sonar sensors, laser range finders, GPS), are highly informative as well as inexpensive.

We have performed image-based localization using omnidirectional camera as the sensor. The most important advantage of the omnidirectional camera over a normal perspective camera is rotational invariance i.e. the same area of the environment is captured independent of the camera's orientation [1]. Another advantage is the large field of view; this also makes the system robust against small changes in the environment.

Image-based mobile robot localization can be broadly categorized into two areas: global image comparison techniques and feature-based techniques. In global image comparison techniques, characteristics of the whole image are collectively used to describe a view. On the other hand, feature-based techniques identify visually salient features in the image. The primary advantages of global techniques over feature-based

techniques are that global techniques are simple and computationally fast. In this paper, we have used two different global techniques for image comparison, namely histograms and fourier transform.

We intend to use these global techniques in topological localization. Topological representation can be thought of as a coarse graph-like representation, where nodes correspond to significant places in the environment; these nodes are interconnected by traversable paths [2]. Building a topological map representation for the image-based mobile robot localization system is the next step in our research work. For the time being, we have used a sequence of images for the experiments instead of a topological map representation.

The rest of the paper is organized as follows: the next section gives an overview of the related work. The image comparison methods are described in section 3, section 4 describes the image-based localization technique, section 5 provides the experimental results and section 6 concludes the paper.

2. RELATED WORK

Histogram-based matching is a well known method among the global techniques of image-based mobile robot localization. Ulrich and Nourbakhsh used histogram-based matching technique for the purpose of place recognition for topological localization in [1]. The eigenspace decomposition of the training images was used in [9]. Krose et al. used principal component analysis for image comparison in [5]. Stricker et al. proposed a method based on the Fourier-Mellin transform to compare images in [11]. Another popular global technique for image-based localization is Fourier transforms which was used in [3, 12].

Feature-based techniques for image-based localization can be described based on the detection method of the local features. Stereo cameras were used to calculate the range and bearings to the image features in [13]. Single camera-based featured position estimation was performed in [14, 15]. David Lowe proposed a method for using scale-invariant image fea-

tures for localization in [16]. In this paper, we chose to use global techniques for image-based localization; we intend to use feature-based techniques in our future work.

3. IMAGE COMPARISON

A sound image comparison technique is very important for an image-based localization system. The aim is to determine a similarity or dissimilarity measure for each pair of images, which gives an observation of how visually similar the two images are. In this paper, we have used two global techniques for image comparison, namely histograms and Fourier transform.

3.1. Using histograms

For an image-based localization system, one of the major problems is to store a large number of images in the memory database, which takes a large amount of space. Histograms are good solutions for this problem. In our experiments, they require very little memory space. As a result, the amount of memory required to store an image can be reduced. Image histograms are rotationally invariant i.e. images captured at a certain location of the environment with different orientations should have identical histograms. In order to determine how well two image histograms match, three histogram matching techniques have been used in this paper, they are given below:

- Jeffrey divergence method
- χ^2 statistics method
- Sum of absolute difference method

A good overview of different histogram matching techniques is given in [4]. The Jeffrey divergence method is numerically stable and robust with respect to size of histogram bins [17]. It is defined as:

$$d_J(H_i, H_j) = \sum_k (h_{ik} \log \frac{h_{ik}}{m_k} + h_{jk} \log \frac{h_{jk}}{m_k}) \quad (1)$$

where, $m_k = \frac{h_{ik} + h_{jk}}{2}$ and h_{ik} and h_{jk} are the histogram entries of the two image histograms H_i and H_j respectively.

The χ^2 statistics method was used in [1, 4]. A detailed description of this method can be found in [18]. It is defined as:

$$d_{\chi^2}(H_i, H_j) = \sum_k \frac{(h_{ik} - m_k)^2}{m_k} \quad (2)$$

where again, $m_k = \frac{h_{ik} + h_{jk}}{2}$ and h_{ik} and h_{jk} are the histogram entries of the two image histograms H_i and H_j respectively.

The last image comparison method is a straightforward one, the dissimilarity measure is obtained by the sum of the

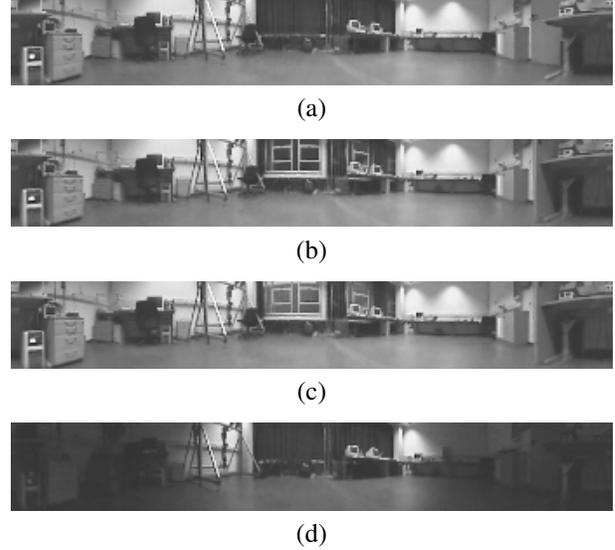


Fig. 1. Sample images from 4 image databases: (a) Original, (b) Night, (c) Twilight, (d) Winlit. Image position (1,2).

absolute differences of the two image histogram entries:

$$d_S(H_i, H_j) = \sum_k |h_{ik} - h_{jk}| \quad (3)$$

where, h_{ik} and h_{jk} are the histogram entries of the two image histograms H_i and H_j respectively.

3.2. Using Fourier transform

Similar to image histograms, the magnitude of the Fourier transform of an image is invariant to the rotation of the image around the optical axis. The panoramic image is transformed row by row via the Fourier transform.

$$d_F(I_i, I_j) = \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} |F_{ik}(l) - F_{jk}(l)| \quad (4)$$

where I_i and I_j are the two panoramic images, each having m rows. $F_{ik}(l)$ and $F_{jk}(l)$ are the Fourier coefficients of the l^{th} frequency of the k^{th} row of images I_i and I_j respectively. This method was also used in [12].

The Fourier coefficients of the low frequency components of the panoramic image are stored to represent the image. In our experiments, we took the first 30 frequency components, because the later frequency components have very small values and thus can be neglected in the calculation of the dissimilarity measure.

4. IMAGE-BASED LOCALIZATION

Image-based localization consists of matching the current image experienced by the mobile robot with training images

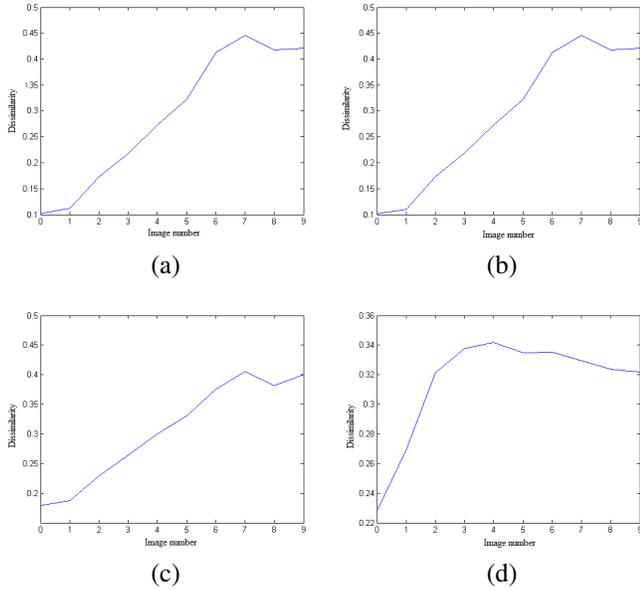


Fig. 2. Image comparison with (a) Jeffrey divergence (b) χ^2 statistics (c) Sum of absolute difference method (d) Fourier transform. The test image is the first image of the image database *original* positioned at (0,0) and the training images are the first 10 images from the image database *original*.

stored in the memory of the robot. In a new environment, the mobile robot is lead along a route and training images are captured. Then if a new test image is captured, it is compared with all the training images and an hypothesis is formed about the current location of the mobile robot. This hypothesis is refined using the discrete Bayes filter as soon as the robot starts to move and new test images are captured. So the output of image-based localization system is a location which refers to one of the training images.

In this paper, our experiments were done using two types of localization problems, namely local localization and global localization. When a mobile robot first starts to localize itself in the environment, it has no knowledge of its location in the environment; this is known as global localization. In case of local localization, the initial location of the robot is known by the mobile robot. Global localization is more difficult than local localization, because at the beginning there is no knowledge about the location of the robot, so the algorithm starts with equal probability given to each training image.

Our image-based localization system is able to perform both types of localization. A probabilistic approach is used in this paper i.e. we will represent the robot's belief of its location as a probability distribution. There are a number of ways to represent probability distributions: continuous or discrete, single or multiple hypothesis. In this paper, we used a discrete Bayes filter, with probability distribution approximated by an array of possible locations (i.e. training images).

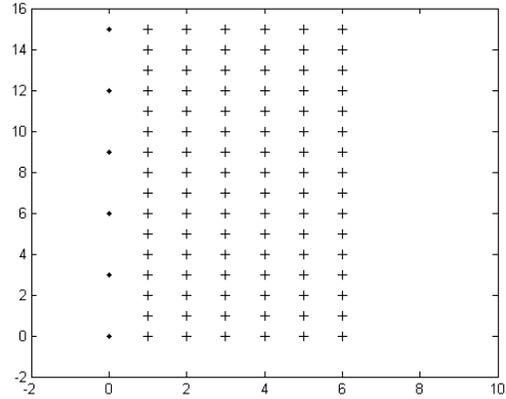


Fig. 3. Experimental image setup for image-based localization. Black dots refer to training images (90 cm apart) and black plus signs refer to test images (30 cm apart). The image sequences are 30 cm apart from each other.

4.1. Bayes filter

The general form of the Bayes filter is given below:

$$\overline{bel}(x_t) = \int p(x_t|u_t, x_{t-1})bel(x_{t-1})dx_{t-1} \quad (5)$$

$$bel(x_t) = \eta p(z_t|x_t)\overline{bel}(x_t) \quad (6)$$

where, x_t is the robot's state at time t , u_t is the control input at time t , z_t is the sensor measurement at time t and η is a normalizing factor. $bel(x_t)$ is the robot's belief that it is in some state x_t . A belief distribution assigns a probability to each possible hypothesis with regards to the true state [10]. Belief distributions are posterior probabilities over state variables conditioned on the available data. $\overline{bel}(x_t)$ is the predicted belief. It predicts the state at time t based on the previous state posterior, before incorporating the measurement at time t .

Equation (5) updates the belief to account for the robot's motion. This generates the **prediction** $\overline{bel}(x_t)$. Equation (6) achieves the **measurement update**. It incorporates the sensor values and combines this information with the prediction to update the belief.

In this paper, we used a discrete Bayes filter which is given by equations (5) and (6), with the exception that the integration is replaced by summation. In order to localize the mobile robot, a probability distribution is maintained over all the training images. So the output of the discrete Bayes filter is a probability distribution over all the training images. The training image with the highest probability value refers to the probable location of the mobile robot in the environment.

4.2. Motion Model

The motion model $p(x_t|u_t, x_{t-1})$ gives the probability of a transition from position x_{t-1} to x_t .

Generally motion models are based on odometry information. The motion model of the differentially driven robot was used in [8].

If a mobile robot is at a certain location in the environment and it makes a forward motion, it is very probable that it will move to a neighbor location in the next time instant; the probability of moving to a place far from its current location is very low. As a result the motion model can be represented using a Gaussian probability distribution, as used in [7]. The motion model is defined as:

$$p(x_t|u_t, x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{dist(x_t, x_{t-1})}{\sigma_x^2}} \quad (7)$$

In the above equation, the function $dist(x_t, x_{t-1})$ refers to a measurement of the distance between the two places x_{t-1} and x_t ; and σ_x is the standard deviation of the distances. In our experiment, the robot moves one step forward in each time instant, we assumed the distance between two adjacent places to be 1. In our experiments, the robot can either move one step forward or stay at the same place in case of no movement; thus a value of $\sigma_x = 1$ is reasonable.

4.3. Measurement Model

The measurement model $p(z_t|x_t)$ gives the probability of acquiring sensory observation z_t under the assumption that the robot is positioned at x_t .

In mobile robotics, different types of sensors are used to acquire sensory observations, such as laser range finders, ultrasonic sensors, camera, etc. The measurement model in [6] is composed of some discrete and continuous measurements: node degree, node equidistance, edge travel distance and feature map landmark location. Goedeme *et al.* [7] used a Gaussian probability distribution to represent the measurement model and we have adopted this approach. As mentioned in [7], there exists a low probability of acquiring an image at a certain location that differs substantially from the training image taken at that location. The measurement model is defined as:

$$p(z_t|x_t) = \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{diff(h_1, h_2)}{\sigma_z^2}} \quad (8)$$

In the above equation, the function $diff(h_1, h_2)$ is obtained by image comparison methods explained in the previous section, and σ_z is the standard deviation measured on the data.

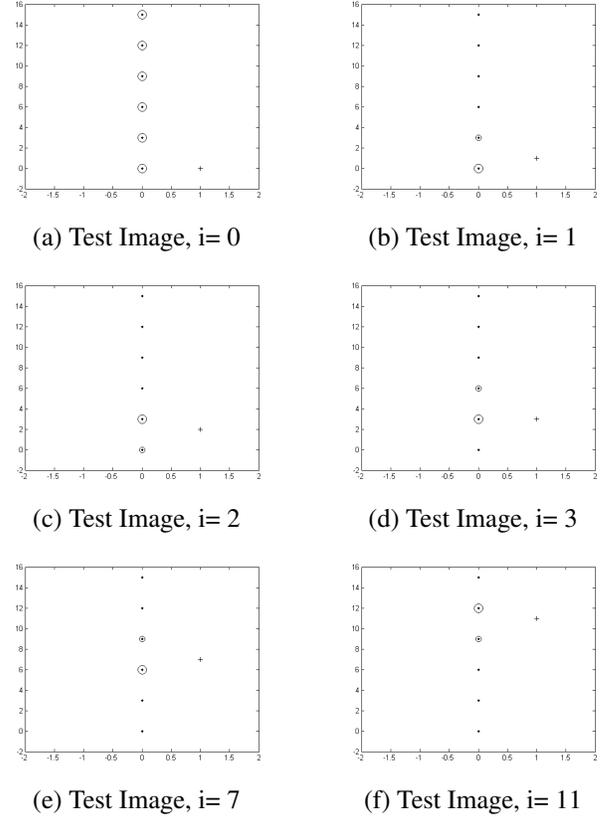


Fig. 4. Demonstration of the image-based global localization.

5. EXPERIMENTAL RESULTS

5.1. Image database

A database of images was collected in the robotics laboratory of Bielefeld University, the database is publicly available at www.ti.uni-bielefeld.de/html/research/avardy. Images were collected by a camera mounted on a pioneer mobile robot. The camera was a catadioptric system consisting of an upward looking camera with a hyperbolic mirror mounted over it. The hyperbolic mirror expanded the camera's field of view to allow the capture of omnidirectional images. A detailed description of the image databases and the catadioptric vision system can be found in [2]. Figure 1 shows sample images from the image databases: *original*, *night*, *twilight* and *winlit*. Image database *original* refers to the standard or default condition of the room, with the curtains and door closed. Images of the database *night* were captured at night with the curtains and door open. Images of the database *twilight* were captured just after the sunset, at that time the room was still receiving plenty of daylight, the curtains and door were kept open. In the images of the database *winlit*, only the two lights near the window were kept on.

Table 1. Image-based Mobile Robot Localization Results (With distant training and test image databases):

Training image database	Test image database	Jeffrey divergence	χ^2 statistics	Sum of difference	Fourier Transform
Original_0	Original_0	100%	100%	100%	100%
Original_0	Original_1	100%	100%	100%	100%
Original_0	Original_2	93.75%	81.25%	75%	81.25%
Original_0	Original_3	81.25%	75%	56.25%	75%
Original_0	Original_4	75%	56.25%	37.5%	62.5%
Original_0	Original_5	25%	18.75%	18.75%	56.25%
Original_0	Original_6	12.5%	12.5%	12.5%	31.25%

Table 2. Image-based Mobile Robot Localization Results (With modified environment):

Training image database	Test image database	Jeffrey divergence	χ^2 statistics	Sum of difference	Fourier Transform
Original_0	Original_1	100%	100%	100%	100%
Original_0	Night_1	93.75%	87.5%	81.25%	100%
Original_0	Twilight_1	93.75%	81.25%	75%	93.75%
Original_0	Winlit_1	18.75%	6.25%	6.25%	37.5%

5.2. Results

5.2.1. Image Comparison Results with different methods

Figure 2 shows the dissimilarity measure using both histogram and Fourier transform methods. For all the image comparison methods, the test image is the first image of the image database *original* positioned at (0,0), and the training images are the first 10 images from the database *original*. The dissimilarity function behaves as expected in all four cases, the dissimilarity of the test image is smallest for the first training image positioned at (0,0), and then it increases for other images. In other words, the dissimilarity function increases with spatial distance over larger area.

5.2.2. Results of the image-based localization system

Figure 3 shows the experimental image setup for image-based localization system. Black dots refer to training images (90 cm apart) and black plus signs refer to test images (30 cm apart). There are 6 images in the sequence of training images and 16 images in the sequence of test images, as can be seen from figure 3. The image sequences are 30 cm apart from each other.

Our image-based localization system is able to perform both local and global localization. Global localization is performed by initializing the system with uniform probability distribution; while for local localization, the initial location of the mobile robot was given. The task is to determine which training image the robot is closer to for a certain test image. If, for example, the location where the test image is captured is closer to the second training image, then the second training image should have the highest probability value. Figure 4 demonstrates the image-based global localization. Black dots refer to training images (90 cm apart) and the black plus sign refers to the current image (i.e. test image). In figure 4(a), the

system is initialized with uniform probability distribution, so each training image has the same probability value (depicted by equal sized circles around the black dots). When the robot moves forward, the first training image obtains the highest probability value, as this training image is the closest to the current image; the training image with the highest probability value is depicted by a large circle and the training image with the second highest probability value is depicted by a small circle. Our localization experiments were done with four different image databases. The comparison is based on the ratio of successful localizations; for example, if the number of test images is 16 and the number of successful localization is 15, then the result is 93.75%.

For the first experiment, the distances between the sequence of training images and that of test images were 0, 30, 60, 90, 120, 150 and 180 cm respectively. The results are summarized in table 1. It can be observed from the table that the ratio of successful localizations is almost 80% even when there is 60 cm distance between the training and test images (for training database *original_0* and test database *original_2*); afterward, the ratio decreases as the distance between the training and test images is increased. The Jeffrey divergence method and the Fourier transform method performed well in this experiment.

For the second experiment, the sequence of training images and that of test images is 30 cm apart; but different test image databases were used; the results are summarized in table 2. It can be observed from the table that our image-based localization system performs well even with modified environments; the Fourier transform method works better than the histogram methods for this experiment. It should be noted that no method performed well when the test images were taken from the image database *winlit*; because only two lights in the room were kept on for the images in the database *win-*

lit, so the room was really dark as can be seen from Figure 1(d). This image database has significant illumination change from all the other databases, we intend to look further into this illumination change problem in our future works.

6. CONCLUSION AND FUTURE WORK

In conclusion, we can say that we have successfully compared two global methods of image comparison for topological localization of mobile robots. The only sensor used is an omnidirectional camera. A probabilistic approach using the discrete Bayes filter was taken. The experiments were done using four image databases; our image-based localization system performed well with distant training and test image databases, it worked well even with modified conditions of the same environment. The Jeffrey divergence method and the Fourier transform method performed well in our experiments. Future extensions of this work include developing a topological map for image-based localization and implementing the system with feature-based techniques like SIFT. We intend to test our system in a larger and more complex environment.

7. REFERENCES

- [1] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in *Proceeding of the IEEE International Conference on Robotics and Automation*, pp. 1023-1029, April 2000.
- [2] A. Vardy and R. Moller, "Biologically plausible visual homing methods based on optical flow techniques," *Connecton Science*, vol. 17, pp. 47-89, 2005.
- [3] H. Aihara, N. Iwasa, N. Yokoya and H. Takemura, "Memory-based self-localization using omnidirectional images," in *Proceeding of the 14th International Conference on Pattern Recognition*, pp. 1799-1803, 1998.
- [4] Y. Rubner, C. Tomasi and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," STAN-CS-TN-98-86, Stanford University, 1998.
- [5] B. J. A. Krose, N. Vlassis and R. Bunchoten, "A probabilistic model for appearance-based robot localization," *Image and Vision Computing*, vol. 12, no. 6, pp. 381-391, April 2001.
- [6] S. Tully, H. Moon, D. Morales, G. Kantor and H. Choset, "Hybrid localization using the hierarchical atlas," in *Proceeding of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2857-2864, November 2007.
- [7] T. Goedeme, M. Nuttin, T. Tuytelaars and L.V. Gool, "Omnidirectional vision based topological navigation," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 219-236, 2007.
- [8] D. Amarasinghe, G. Mann and R. Gosine, "Integrated laser and camera sensor for simultaneous localization and mapping (SLAM)," in *Proceeding of the 16th Annual Newfoundland Electrical and Computer Engineering Conference*, November 2006.
- [9] M. Jogan and A. Leonardis, "Panoramic eigenimages for spatial localization," in *Proceeding of the 8th International Conference on Computer Analysis of Images and Patterns*, pp. 558-567, 1999.
- [10] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics*, The MIT Press, 2005.
- [11] D. Stricker and P. Daehne, "Design and development issues for archeoguide: an augmented reality based cultural heritage on-site guide," in *Proceeding of the International Conference on Augmented, Virtual Environments and 3D Imaging*, 2001.
- [12] E. Menegatti, M. Zoccarato, E. Pagello and H. Ishiguro, "Hierarchical image-based localization for mobile robots with monte-carlo localization," in *Proceeding of the European Conference on Mobile Robots*, 2003.
- [13] A. J. Davison, "Mobile robot navigation using active vision," Ph.D. dissertation, University of Oxford, June 1998.
- [14] A. J. Davison, "Real-time simultaneous localization and mapping with a single camera," in *Proceeding of the International Conference on Computer Vision*, pp. 1403-1410, 2003.
- [15] J. M. M. Montiel, J. Civera and A. J. Davison, "Unified inverse depth parametrization for monocular slam," in *Proceeding of the Robotics: Science and Systems Conference*, 2006.
- [16] D. Lowe, "Object recognition from local scale-invariant features," in *Proceeding of the International Conference on Computer Vision*, pp. 1150-1157, 1999.
- [17] J. Puzicha, T. Hofmann and J. Buhmann, "Non-parametric similarity measures for unsupervised texture segmentation and image retrieval," in *Proceeding of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 267-272, 1997.
- [18] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, Brooks/Cole Publishing Company, 1991.