Toward image-based localization for AIBO using wavelet transform

Alberto Pretto*, Emanuele Menegatti*, Enrico Pagello*†,
Yoshiaki Jitsukawa‡, Ryuichi Ueda‡and Tamio Arai‡

*Dept. of Information Engineering, The University of Padua, Italy
†*also with* ISIB-CNR corso Stati Uniti, Padua, Italy
‡The University of Tokyo, Tokyo, Japan

Abstract. This paper describes a similarity measure for images to be used in image-based localization for autonomous robots with low computational resources. We propose a novel signature to be extracted from the image and to be stored in memory. The proposed signature allows, at the same time, memory saving and fast similarity calculation. The signature is based on the calculation of the 2D Haar Wavelet Transform of the gray-level image. We present experiments showing the effectiveness of the proposed image similarity measure. The used images were collected using the AIBOs ERS-7 of the RoboCup Team Araibo of the University of Tokyo on a RoboCup field, however, the proposed image similarity measure does not use any information on the structure of the environment and do not exploit the peculiar features of the RoboCup environment.

1 Introduction

Three are the main problems to be solved by any techniques of image-based localization one can develop: (i) how to reduce the number of images necessary to fully describe the environment in which the robot is working; (ii) how to efficiently store a large data set of reference images without filling-up the robot's memory (it is common to have several hundred reference images for typical environments); (iii) how to calculate in a fast and efficient way the similarity of the input image against all the reference images in the data set.

Several works have been published that use the image-based localization approach (among the others [14][5]). Each work tried to solve these problem in a different way. One of the most effective approaches to reduce the number of images needed to describe the environment is to mount an omnidirectional camera on the robot. In fact, an omnidirectional camera can acquire a complete view of the surroundings in one shot avoiding the need to shot at different gazing directions. The most popular technique, to reduce the memory consumption of the reference data set, is to extract a set of eigenimages from the set of reference images and to project the images into eigenspaces. The drawback of such systems is that they need to further preprocess the panoramic cylinder images they created from the omnidirectional image in order to obtain the rotational invariance, as

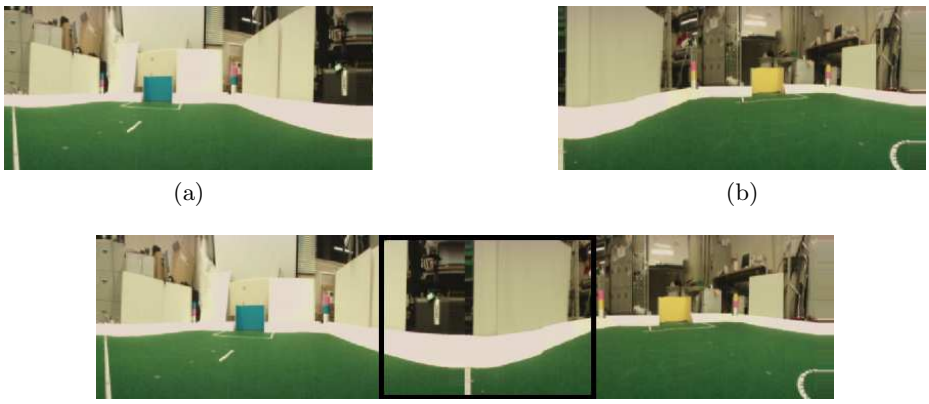
in [1], in [10] and in [7] or to constrain the heading of the sensor as in [11]. An approach that exploit the natural rotational invariance of the omnidirectional images is to create a signature for the image based on the colour histograms of vertical sub-windows of the panoramic image, as in [8] or in [6]. However, this approach based on colours might not be very effective in a general environment with poor color information. An alternative approach to preserve the rotational invariance of omnidirectional image is the one presented in [12], which exploit the properties of the Fourier signature of the omnidirectional images.

Despite the effectiveness of the approaches based on the omnidirectional cameras, it is not always possible to mount an omnidirectional camera on the robot. A solution can be to constrain the movements of the robot in order to keep the camera pointing at the same location [2], but this greatly limits the motion of the robot. An alternative solution can be to extract from the perspective images some features that reduce the amount of required memory while retaining a rich description of the image. A good example of this is reported in [17], where 936 images were stored in less than 4MB by extracting features invariant to translation and to some amount of scaling. However, to extract such a large amount of images is time consuming, even if an automatic procedure is available.

2 Image-Based Localization for ERS-7 AIBO

In order to minimize the reference images to be stored, our idea is to keep as reference images two 180 degree panoramic views of the environment at every reference location; whose two images fully captures the appearance of the environment at a reference location. How to store in a memory-saving way the reference images and how to efficiently compare them with the input images is particularly important when using a robot with limited storage memory and limited computational resources, as the AIBO ERS-7 used in our experiments. Nowadays, a very popular approach used also for image-based localization is the SIFT approach proposed in [15]. However this approach is computationally expensive and focuses on local features in the single images, rather than on the global appearance of the environment. We developed an algorithm that allows the ERS-7 robot to autonomously build two 180 degrees panorama images using its standard camera and to stitch them together.

At the running stage, the 208×160 pixels image grabbed by the AIBO is matched with subwindows of the 360 deg. panoramic image (the black sliding window in Fig. 1(c). This can be done in an extremely efficient way, by effectively exploiting properties of the wavelet signature we developed. The matching returns similarity values that can be used to localize the robot. The fundamental assumption in this matching strategy is the head and the neck of the ERS-7 are always in the same configuration: same height and parallel to the ground. Image grabbed in different situations simply are not used in the robot localization process.



(a) (b)

(a)

Fig. 1. In (a) and (b) two reference images, both taken by the ERS-7 at the same reference position, but with opposite heading. In (c) their composed 360 deg. panoramic image, with depicted the horizontal-sliding window used for input image matching. The position of the window is related to the returned bearing angle.

3 Discrete Wavelet Transform Image Signature

For an effective vision-based localization strategy, we need to store the visual memory of the environment (i.e., the reference images) in a compact and effective way. Images should be represented using specific *signatures* that characterize the content and some useful features of each image in the database. Signatures must have very small size compared with image sizes. The signature of a query image will be directly compared with the signatures of the reference images through a specific metric called *similarity measurement*. In the image-base localization context, a small similarity value between two images means that the two images have been grabbed one close to the other.

A tool for non-stationary signal analysis (whose frequency response varies in time, like in the images) is the Wavelet Transform [16]: it gives information about which frequency components exist and where these components appear. Wavelet features are successfully exploited in the image coding algorithms; for instance, the upcoming still image compression standard JPEG-2000 [3] is based on Wavelet Transform. As well, wavelet signatures are successfully used in image retrieval algorithms, e.g. [9, 13], and texture retrieval algorithm, e.g. [4]. We exploited these properties of the Wavelet Transform using the Discrete Wavelet Transform (DWT) coefficients in order to represent images in a compact way, without losing information about location of the image discontinuity, shapes and texture [13].

Discrete Wavelet Transform are used to analyze signals at different scale, $scale = 1/frequency$. In single level discrete 1-D Wavelet Transform, the signal is decomposed into a coarse approximation and a detail information (Eq. 3,3).

Decomposition is performed convolving the input signal with a low-pass filter and an high-pass filter. After filtering, according to the Nyquist's rule, it is possible to eliminate half of the samples. $g()$ and $h()$ low and high-pass filter depend on chosen wavelet type.

$$y_{low}(k) = \sum_n x(n) * g(2k - n) \quad (1) \qquad y_{high}(k) = \sum_n x(n) * h(2k - n) \quad (2)$$

The single-level discrete Wavelet Transform can be recursively repeated for further decomposition of the previously y_{low}. In the 2-D case, the 1-D Wavelet Transform is applied first on each row of the image. The process results in two new matrices with half columns than the input image. A further 1-D DWT is applied to the columns of the resulting matrices. At the end of the one-level 2-D decomposition, $m \times m$ input matrix is decomposed in 4 $m/2 \times m/2$ matrices. In Fig. 2 is shown a multilevel 2-D Wavelet decomposition: I is the input image, C_i are the approximation coefficients, H_i,V_i and D_i are respectively the horizontal, vertical and diagonal detailed coefficients, $i = 1, 2, \ldots, n$ represent the recursion level of wavelet decomposition.

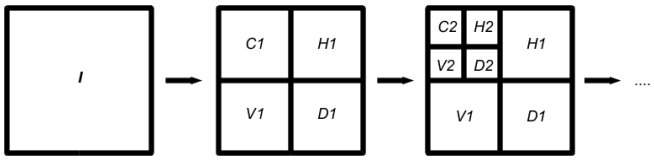


Fig. 2. Multilevel 2-D Wavelet decomposition.

3.1 The Proposed Signature

We use as image signature a 2-D Haar Wavelet Transform of the grey-level values of the image. We decide to stop at 4-th level decomposition, and to characterize images only by the detailed coefficients (horizontal, vertical and diagonal) of this level. Thanks to the great properties of frequency localization given by the DWT, it is possible to store only a few of Discrete Wavelet Signatures for the references panoramic images: the subset of coefficients required for similarity measurement can be extracted using a simple sliding-window strategy.

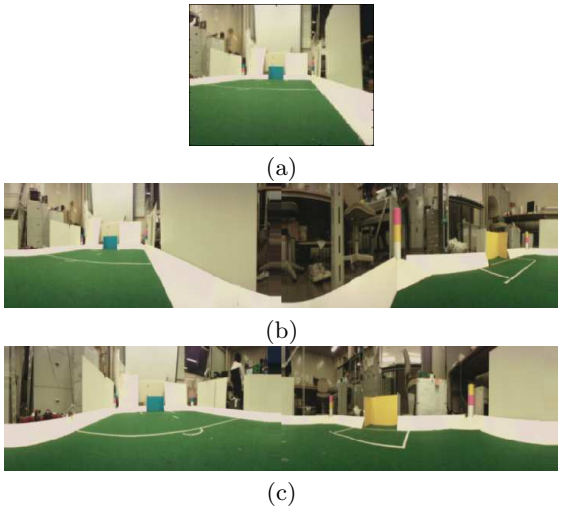
Haar Wavelet is chosen as wavelet type because of it is very effective in detecting the exact instants when a signal changes: image discontinuity are one of the most important features chosen in image-based localization. Haar Wavelet can be easily implemented and they have very fast to compute. If one is interested in image reconstruction phase, the Haar Wavelets are not the good choice, because they tend to produce a lot of squared artifacts in the reconstructed im-

age. However, we are not interested in the reconstruction phase, we exploit the DWT coefficients to calculate the similarity.

Other wavelet type was taken into account: Daubechies' Wavelets [16] family, commonly used in image coding, were tested. Surprisingly, growing the vanishing moments of the wavelets (i.e. the Daubechies' Wavelets order) performances decade. Those Wavelets are better suitable than Haar to detect a rupture in high-order derivative, but we are interested on detecting discontinuity and features directly in the signal.

Coupling the 180-degree references images, we obtain panoramic 720×160 pixels images. By applying recursively 2-D Haar Wavelet Transform, we can reduce a lot the signature size. On the other hand, high level failed on represent effectiveness feature of images, as edge and texture, useful for environment characterization. Choice of decomposition level 4 is a trade off between a compactness representation and a reliability similarity computation.

Coefficient of Haar Wavelet at level 4 pertains to 16×16 pixels square. In order to achieve horizontal 1-degree accuracy, we need to calculate only 8 global Discrete Wavelet Signature for every panoramic reference image, starting from pixels with $x = 0, 2, \ldots$ to 14. Given the bearing angle of an input image, the DWT coefficients can be effectively extracted from the right Wavelet table (from the eight precomputed, the index of the table depending on orientation) with a simple horizontal sliding-window strategy.



(a)

(b)

(c)

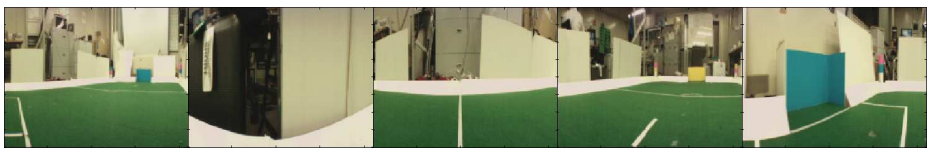
Fig. 3. (a) is an input image; (b) is the references image with the best match (100%); (c) is the second best match (61%)using the proposed DWT signature.

In our experience, the approximation coefficients are not well suitable for image similarity computation: considering Haar Wavelet, those coefficients rep-

resent only the mean of the intensity of the pixels composing the macro-squares (16×16 pixels in our case). On the other hand, detailed coefficients can be used to well detected and highlight image discontinuities, shapes and patterns. Our image signature is based on those coefficients computed at level 4: approximation coefficients are simply discarded. As shown in [9], a coarse quantization of these coefficients doesn't affect the effectiveness of the Haar Wavelet coefficients in the image retrieval field. We tested a similar approach for our scope obtaining very good experimental results. We simply represent detailed coefficients d_i as -1 if $d_i < 0$ and as 1 if $d_i \geq 0$. In this way, it is possible to storage every detailed coefficient in only a single bit. Given the signature of an input images and the right subset of coefficients of a reference image, we compute our similarity measure as:

$$Sim = w_h * \sum_m \sum_n |H_i(m,n) - H_r(m,n)| + w_v * \sum_m \sum_n |V_i(m,n) - V_r(m,n)| + w_d * \sum_m \sum_n |D_i(m,n) - D_r(m,n)| \quad (3)$$

Where m, n represent rows and columns of the detailed coefficients matrices, H_i and H_r, V_i and V_r, D_i and D_r represent the horizontal, vertical and diagonal detailed coefficients respectively of the input and reference image. w_h, w_v, w_d are weights usefull to move importance through the three different set of coefficients in the localization process. Our default value are $w_h = 0.5, w_v = 0.25, w_d = 0.25$, because of the large amount of vertical characteristic in indoor environment. The memory saving of our approach is considerable: a gray-scale omnidirectional reference image $720 \times 160 = 115.2$ Kbyte of memory can be represented by our DWT signature with only 1.3 Kbyte. Experimental results will show the reliability of our approach.



(a) (b) (c) (d) (e)

Fig. 4. Input images used for Fig. reffig:likelihoods.

4 Experiments

We tested the system in a RoboCup Four-legged League 540×360 cm soccer field, using a grid of 13 by 9 reference images. The images have been grabbed in known

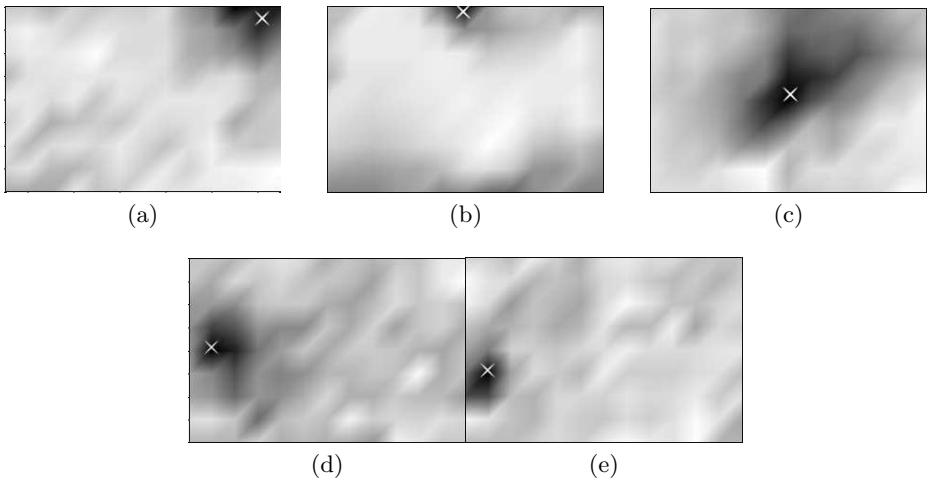


Fig. 5. Similarity values for all possible poses of the robot in the field, given the input images of Fig. 4. Darker areas correspond to a higher similarity. The cross represents the ground-truth robot position.

poses regularly distributed all over the field. For every reference position two 180 deg. panoramic images were collected, using the technique explained in the previous sections. A set of input images, taken in distinct known positions and at different rotations, was used to test the proposed image similarity measure. The ground-truth position of the AIBO for every input image was measured by hand with an error less than 0.3 cm. In Fig. 4, five input images are depicted. In Fig. 5, the corresponding similarity values against the reference images are plotted. The similarity values have been interpolated to obtain a similarity value for every possible pose of the robot in the field. In the plot the darker areas correspond to a higher similarity. The white cross represents the actual pose of the robot.

5 Conclusion and Future Works

We presented a new way to calculate the similarity between images to be used in the image-based localization approach on autonomous robot with low computational resources. The proposed technique exploit the properties of the Haar Wavelet Transform. We presented a technique that enables a quick set-up of the robot and of its localization system without requiring any previous knowledge on the environment. Successful experiments on the calculation of the image similarity of real images grabbed by a AIBO ERS-7 robot have been presented. These results encourage the creation of a Monte-Carlo localization system that uses this approach and make us feel more confident on exploring of the use of this similarity measure to develop a visual topological SLAM strategy.

References

1. H. Aihara, N. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localisation using omnidirectional images. In Anil K. Jain, Svetha Venkatesh, and Brian C. Lovell, editors, *Proc. of the 14th International Conference on Pattern Recognition*, volume vol. I, pages pp. 1799–1803, 1998.
2. R. Cassinis, D. Duina, S. Inelli, and A. Rizzi. Unsupervised matching of visual landmarks for robotic homing using Fourier-Mellin transform. *Robotics and Autonomous Systems*, 40(2-3), August 2002.
3. JPEG Committee. Jpeg home page. http://www.jpeg.org.
4. M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Tr. Im. Proc.*, 11(2):146–158, February 2002.
5. G. Dudek and D. Jugessur. Robust place recognition using local appearance based methods. *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, 2, 2000.
6. Emanuele Frontoni and Primo Zingaretti. An efficient similarity metric for omnidirectional vision sensors. *Robotics and Autonomous Systems*, 54(9):750–757, 2006.
7. Jos Gaspar, Niall Winters, and Jos Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transaction on Robotics and Automation*, Vol 16(number 6), December 2000.
8. H.-M. Gross, A. Koenig, Ch. Schroeter, and H.-J. Boehme. Omnivision-based probabilistic self-localization for a mobile shopping assistant continued. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS 2003)*, pages pp. 1505–1511, October 2003, Las Vegas USA.
9. Charles E. Jacobs, Adam Finkelstein, and David H. Salesin. Fast multiresolution image querying. In *Proceedings of SIGGRAPH 95 (Los Angeles, California, August 6–11, 1995)*, New York, 1995.
10. M. Jogan and A. Leonardis. Robust localization using panoramic view-based recognition. In *Proc. of the 15th Int.Conference on Pattern Recognition (ICPR00)*, volume 4, pages pp. 136–139. IEEE Computer Society, September 2000.
11. B.J.A. Krse, N. Vlassis, R. Bunschoten, and Y. Motomura. A probabilistic model for appareance-based robot localization. *Image and Vision Computing*, vol. 19(6):pp. 381–391, April 2001.
12. Emanuele Menegatti, Takeshi Maeda, and Hiroshi Ishiguro. Image-based memory for robot navigation using properties of the omnidirectional images. *Robotics and Autonomous Systems, Elsevier*, 47(4):pp. 251–267, July 2004.
13. Apostol Natsev, Rajeev Rastogi, and Kyuseok Shim. Walrus: a similarity retrieval algorithm for image databases. *SIGMOD Rec.*, 28(2):395–406, 1999.
14. SK Nayar, H. Murase, and SA Nene. Learning, positioning, and tracking visual appearance. *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pages 3237–3244, 1994.
15. S. Se, D. Lowe, and J. Little. Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks. *The International Journal of Robotics Research*, 21(8):735, 2002.
16. M. Vetterli and J Kovacevic. *Wavelets and Subband Coding*. Signal Processing Series. 1995.
17. J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization by combining an image retrieval system with monte carlo localization. *IEEE Transactions on Robotics*, 21(2):208–216, 2005.