2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Efficient Hierarchical Localization Method in an Omnidirectional images Memory

Jonathan Courbon*†, Youcef Mezouar*, Laurent Eck†, Philippe Martinet*‡

*LASMEA
24 Avenue des Landais
63177 AUBIERE - FRANCE

†CEA, List
18 route du Panorama, BP6
F- 92265 FONTENAY AUX ROSES - FRANCE

‡Sungkyunkwan University
Intelligent Systems Research Center
SUWON, SOUTH KOREA

Email: firstname.lastname@lasmea.univ-bpclermont.fr

Email: laurent.eck@cea.fr

***Abstract*— An efficient method for global robot localization in a memory of omnidirectional images is presented. This method is valid for indoor and outdoor environments and not restricted to mobile robots. The proposed strategy is purely vision-based and uses as reference a set of prerecorded images (visual memory). The localization consists on finding in the visual memory the image which best fits the current image. We propose a hierarchical process combining global descriptors computed onto cubic interpolation of triangular mesh and patches correlation around Harris corners. To evaluate this method, three large images data sets have been used. Results of the proposed method are compared with those obtained from state-of-the-art techniques by means of 1) accuracy, 2) amount of memorized data required per image and 3) computational cost. The proposed method shows the best compromise in term of those criteria.**

I. INTRODUCTION

Several recent publications have focused on using visual references as environment representation (*visual memory*) for mobile robot navigation [1], [2]. In this representation, images of the robot workspace are acquired during a learning stage and memorized. The data set can be topologically organized as in [1] or it can contain some additional metric information as in [3]. The first step during a navigation process is the self localization of the robot in this map of the environment. The localization consists on finding the image of the memory which best fits the current image by comparing pre-processed and on-line acquired images.

The work presented in this paper is focused on the localization onto an omnidirectional images memory. Omnidirectional vision is usefull in many robotic applications because it provides a large field-of-view of the environment. However, it exhibits some supplementary difficulties compared to conventional perspective images. Such images can be acquired by fisheye or by catadioptric cameras, which have a similar behaviour as demonstrated in [4].

The efficiency of a visual localization method can be measured by means of: 1) accuracy of the results, 2) memory needed to store data and 3) computational cost. The main objective of the work presented in this paper is to optimize the localization process under those criteria.

The methods consist on matching the current image with all the images of the memory. Two main strategies exist to match images: the image can be represented by a single descriptor (*global approaches*) [5], [6] or alternatively by a set of descriptors defined around visual features (*landmarks-based* or *local approaches*) [7], [8], [2]. Some hybrid approaches consisting on globally describing a subset of the image have also been proposed in order to be more robust to occlusions than global methods [9].

In one hand, local approaches are generally more accurate but have a high computational cost [2]. On the other hand, global descriptors speed up the matching process at the price of affecting the robustness to occlusions. One solution consists on using a hierarchical approach which combines the advantages of both methods [10]. In a first step, global descriptors allow to select only some possible images and then, if necessary, local descriptors are used to keep the best image. The proposed global descriptor is based on a cubic interpolation of the image with a triangular mesh. This descriptor has a low computational cost and provides good results with an acceptable amount of memorized data. This approach is combined with a classical patches correlation around Harris corners which gives accurate results with a low computational cost. The results of our experimental comparisons with state-of-the-art techniques show that the best compromise between computational efficiency and accuracy is obtained with this proposed method.

This paper is organized as follow. Existing approaches are presented in Section II. The proposed approach is detailled in Section III. Finally, experiments have been performed with different dense datasets. Our approach is compared with other methods by means of the given criteria in Section IV.

II. OMNIDIRECTIONAL IMAGES DESCRIPTORS

This section briefly reviews global and local descriptors for localization in a memory of omnidirectional images.

A. Global descriptors

A first solution is to globally describe the image. In that aim, omnidirectional images are mapped onto cylindrical images of size 128×32 in [5]. The image is directly described by the gray level values. In [11], a shift invariant representation is computed by rotating the cylindrical image in a reference direction. Unfortunately, this direction is not

absolute as soon as occlusions appear. In order to decrease the size of the memorized data, images can be represented by their eigen vectors using Principal Component Analysis as proposed in [12]. Unfortunately, when a new image is integrated in the memory, all eigen vectors have to be recomputed. This process is very complex and it has a very high computational cost. Moreover those methods are not robust to changes of the environment.
The histogram of the gray level values is largely employed as global signature. Its computation is efficient and it is rotation-invariant. However, histogram methods are sensitive to change of light conditions. Blaer et Allen [13] propose color histograms for outdoor scene localization. A normalization process is applied before computing the histograms in order to reduce the illumination variations.
In [6], a global descriptor based on a polar version of high-order local autocorrelation functions (PHLAC) is proposed. It is based on a set of 35 local masks applied to the image by convolution. Similarly to histogram, this descriptor is rotation-invariant.

B. Local descriptors

Global descriptor-based methods are generally less robust to occlusion compared to landmark-based methods. In those last methods, some relevant visual features are extracted from the images. A descriptor is then associated to each feature neighbourhood. The robustness of the extraction and the invariance of the descriptor are one main issue to improve the matching process. We can sort the approaches into two main categories. In the first category, the feature detection and description designed for images acquired by perspective cameras are directly employed with omnidirectional images. The second category takes the geometry of the sensor into account and thus uses operators designed for omnidirectional images. The most popular visual features used in the context of localization in an image database are projected points. However, projected lines can also be exploited as proposed in [14].

1) Perspective-based local descriptor: The Scale Invariant Feature Transform (SIFT, [15]) has been shown to give the best results in the case of images acquired with perspective cameras. The SIFT descriptor is a set of histograms of gradient orientations of the normalized (with respect to orientation and scale) Difference of Gaussian images. In view of the effectiveness of this descriptor, several extensions have been proposed. It has been used with omnidirectional images in [7]. Given that many points are detected in an omnidirectional image, Tamimi et al. [8] proposed an iterative SIFT with a lower computational cost. In [16], points are detected with a Sobel filter and described by a Modified Scale Invariant Feature Transform (M-SIFT) signature. This signature slightly takes into account the geometry of the omnidirectional sensor by rotating the patch around an interest point. In [2], the Speeded Up Robust Features (SURF) are employed as descriptors. SURF points are detected using the Hessian matrix of the image convolued with box filters and the descriptor is computed thanks to Haar-wavelet extraction. The computational cost of this descriptor is much lower than the one obtained for SIFT.
Unfortunately, those signatures describe a local neighbourhood around interest points and do not take into account the high distortions caused by the geometry of the omnidirectional cameras.

2) Descriptors adapted to omnidirectional images: In the second category, detection and description processes are specially designed to take into account those high distortions. In [17], [18], a classical Harris corner detector is proposed but the shape and the size of a patch around a feature is modified according to the position of the point and to the geometry of the catadioptric sensor. Finally, a standard 2D correlation (respectively a centered and normalized cross correlation) is applied to the patches in [17] (respectively in [18]).

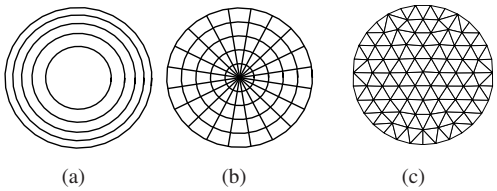
After computing the descriptors of the current and memorized images, those descriptors have to be matched. For local approaches, this step is generally based on Pyramidal matching as in [14] or on Nearest Neighbour matching as in [15]. This last algorithm considers that a matching is correct if the ratio between the distances of the first and second nearest neighbours is below a threshold.
It is possible to eliminate wrong matching through the recovery of the epipolar geometry between two views [19] at the price of higher computational cost. A full reconstruction can also be obtained with three views and the 1D trifocal tensor as proposed in [20].

C. Hybrid descriptor

Some hybrid descriptor have been designed to combine the advantages of the two previously cited categories (local and global approaches) by globally describing subsets of the image. In [9], five histograms of the first and second orders derivatives of the grey-level image are considered. Instead of the whole image, the image is decomposed into rings (refer to Fig. 1(a)). In one hand, a decomposition into few rings decreases the accuracy. On the other hand, increasing the number of rings increases the computational cost and decreases the robustness to occlusions. In [12], the image is first projected onto an englobing cylinder and a grid decomposition is then proposed. This projection step is time consuming and it implies the modification of the quality of the image which can lead to less accurate localization results. The proposed approach detailled in the next section allows to obtain equal size subregions by using triangular mesh (refer to Fig. 1(c)). In the experimental results (Section IV), an angular sector decomposition (Fig. 1(b)) has also been considered for comparison purposes.

III. Proposed approach: cubic interpolation and ZNCC around Harris corners

We propose a new hierarchical approach for localization in a database of omnidirectional images. The computational efficiency is ensured in a first step by defining a well suited global descriptor which allows to select a set of candidate



(a) (b) (c)

Fig. 1. Different subregions decomposition: (a) rings, (b) angular sectors, (c) triangular mesh

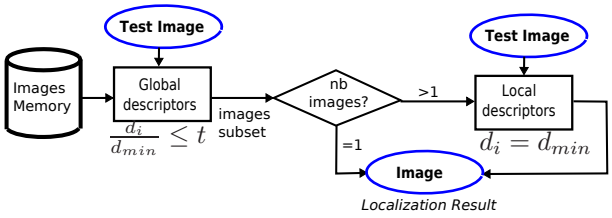


Fig. 2. The proposed hierarchical approach

images. Local descriptors are then exploited to select only the best image and thus to ensure accuracy. This principle is summarized in Fig. 2.

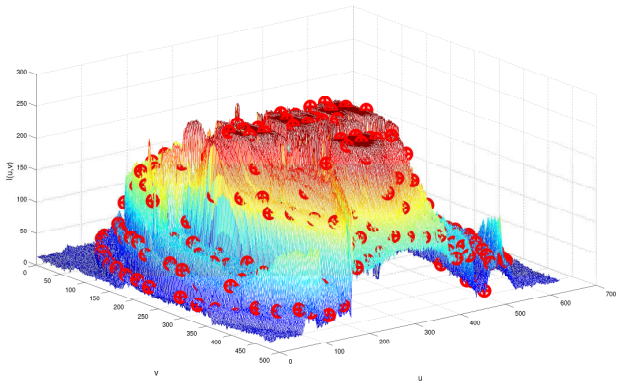
The global descriptor is based on a cubic interpolation of the image, with the nodes of the triangular mesh as control point. The local descriptor is based on the neighbourhood of Harris corners.

A. Proposed global descriptor

In order to be more robust to occlusions, we propose a new approach which can be classified as an hybrid global method. In hybrid methods, images are divided into s subregions. After this decomposition step, each subregion is globally described by a d-dimensional vector. The size of the obtained descriptor is $s \times d$. The major drawback of those approaches is that subregions are seen as independant zones while a continuous change exists between two contiguous regions. Based on this observation, we propose to use a geometrical image representation derived from surface interpolation.

Subregion decomposition: In order to have approximately the same quantity of information for each subregion (similarly to square decomposition for perspective images), a triangular mesh is employed (see Fig. 1(c)). In this decomposition, nodes are approximately equidistant one-by-one and the obtained triangles cover the same area. We use the triangular mesh generator proposed in [21]. The technique is based on the analogy between a simplex mesh and a truss structure. Meshpoints $\{P_1\ ,\ P_2\ ,\ \ldots P_p\}$ are nodes of the truss and segments between two meshpoints are bars. An appropriate force-displacement function is applied to the bars at each iteration. This function takes into account the internal force due to the bars and the external force due to the boundaries. Node locations $P_i = [x_i\ \ y_i]$ are computed by solving for equilibrium in a truss structure using piecewise linear force-displacement relations. Those nodes $\{P_1\ ,\ P_2\ ,\ \ldots P_p\}$ will be employed as control points for the surface interpolation.

(a)



(b)

Fig. 3. (a) Image and (b) same image as a surface, with the control points of the interpolated surface (small circles)

Surface interpolation: Images have first their histogram equalized in order to be more robust to illumination changes. A grey-level image can be seen as a 3D surface with the grey level as the third coordinates (refer to Fig. 3):

$$\mathbf{I} : \begin{cases} [0,1,\ldots,N] \times [0,1,\ldots,M] & \mapsto & [0,255] \\ (u,v) & \rightarrow & \mathbf{I}(u,v) \end{cases}$$

The interpolation consists on locally approximating this surface $\mathbf{I}(u,v)$ by a surface $f(s,t)$, $s \in [0;1]$, $t \in [0;1]$. Many interpolation techniques exist: linear, cubic, bicubic, nearest neighbour.... All techniques are considering *control points* (plotted in Fig. 3(b)). In order to compare descriptors of different images, it is necessary to have control points at the same positions. Moreover, regular positions ensure a better interpolation. In that aim, we propose to use the triangular mesh vertices presented previously as control points and the altitude Z of the control points of the approximated surface as descriptors. The surface is interpolated by a cubic function. The required computational cost is low and interpolation errors are small.

B. First selection

Descriptor Z_c (respectively Z_i) is computed for the current image $\mathbf{I}_c$ (respectively for the memorized image $\mathbf{I}_i$). The distance between those two images is $d_i = d(\mathbf{I}_c, \mathbf{I}_i) = \|\mathsf{Z}_i - \mathsf{Z}_c\|$. The chosen distance is the L_1 distance: $d_{(\mathsf{Z}_c,\mathsf{Z}_i)} = \sum_{k=1}^{d} |Z_{c,k} - Z_{i,k}|$ where $Z_{c,k}$ (respectively $Z_{i,k}$) corresponds to the kth element of the descriptor of the image $\mathbf{I}_c$ (respectively $\mathbf{I}_i$). The best memorized image corresponds to

the image with the minimum distance d_{min}. Kept candidate images are such that $\frac{d_i}{d_{min}} \leq t$ where the threshold $t \geq 1$ allows to not reject the images which have a distance similar to d_{min}. If the number of candidate images is upper than 1, a matching based on local descriptors is then applied for those selected images. Else, the result of the localization is the image $\mathbf{I}_k$ such that $d_k = \min_i(d_i)$.

C. Local descriptor

The two main local descriptors approaches are the SIFT [7] and the SURF [2]. As already mentionned, those approaches do not take into account the geometry of the omnidirectional sensor. If two omnidirectional images have been taken approximately at the same position, those descriptors are efficient whereas it is not the case for images taken at distant positions. As we will see in the next section, a classical local approach based on the Zero Normalized Cross Correlation (ZNCC) between patches around Harris corners has a lower computational cost than SIFT or SURF and similar accuracy when images corresponding to closed viewpoints are considered. The distance between two images is $d_i = d(\mathbf{I}_c, \mathbf{I}_i) = 1/(number\,of\,matched\,features)$. The final result of the localization is the image $\mathbf{I}_k$ such that $d_k = \min_i(d_i)$.

IV. Localization experiments

The performance of the method proposed in the last section is shown for the localization in three large sets of images by means of accuracy, amount of memorized data required per image and computational cost. Best results for a given criteria are set in bold type and worse results in italic. The results of the proposed method are given in the last raw of each table.

The cubic interpolation approach (*Cub*) is compared with PHLAC [6] (*PHLAC*) and Gonzalez [9] (*Gonz*) methods. It is also compared with three other global approaches: using the mean gray level for each angular sector (*Sect*) or for each triangular region (*Triang*), or representing each sector by the histogram of gray level values (*HistoSect*). The local approach *CorrHar* as proposed and successfully employed in [3] is compared with SIFT (*SIFT*) and SURF (*SURF*) methods. Finally, the proposed hierarchical method (*CubCorrHar*) is compared with the other approaches and with the hierarchical approach consisting of a sector decomposition *Sector* and then *CorrHar* (*SectCorrHar*).
SIFT 128-dimensional descriptors are computed with the C demo code of D. Lowe and SURF 64-dimensional descriptors with the C++ code provided by the authors. The matching is based on a Nearest Neighbour matching. For *CubHarris*, around 500 Harris points neighbourhoods are matched with the same C++ code as used in [3].

Three data sets are used: *Almere*, *UAV* and *Walking* data sets. *Almere* data set was provided for the workshop [22]. It contains images of size 1024×768 pixels acquired by a catadioptric camera pointing to the ceil and embedded onto a mobile robot navigating in a typical house environment with people walking around. As in [2], every 5th frame are extracted for the experiments: half for reference and the other half for testing. The number of test images is 978. *UAV* data set contains images of size 384×288 pixels taken by a fisheye camera embedded onto a X4-flyer UAV navigating in an indoor environment. The camera points a third to the ground, a third to the ceil and the other third forward. One frame every 5 frames is extracted for the experiments (reference) and one frame every 20 frames with an offset for testing. The number of test images is 188. Finally, *Walk* data set contains images of size 640×480 pixels taken by a fisheye camera carried by a human in different environments into or nearby our laboratory. The number of test images is 445. Contrary to the other data sets, test images come from an other walk, during an other day thus conditions are different between the training and the navigation steps. This method provide thus more realistic conditions.

A. Required memory size

Tab.I shows the needed memory for the descriptors used in the sequel. For the local methods, the dimension of a descriptor depends on the number of detected features and of the size of a patch descriptor. Excepted for the PHLAC, all global descriptors depend on the sub-region decomposition.

	Almere dataset	*UAV dataset*	*Walk dataset*
B&W image	768 Kb	108 Kb	300 Kb
SIFT	*1 100* Kb	*440* Kb	*800* Kb
SURF	240 Kb	80 Kb	150 Kb
Gonz	**0.3** Kb	**0.3** Kb	**0.3** Kb
PHLAC	0.45 Kb	0.45 Kb	0.45 Kb
Sect	0.9 Kb	0.9 Kb	0.9 Kb
CorrHar	8 Kb	8 Kb	8 Kb
Triang	3.4 Kb	3.8 Kb	3.8 Kb
Cub	2.2 Kb	2.4 Kb	2.4 Kb

TABLE I
Approximative required memory for different descriptors

Local descriptors are much bigger than global descriptors. A higher computational time is thus expected for those approaches. The SIFT descriptors are approximately 5 times bigger than the SURF descriptors. *Gonz* and *PHLAC* require small memory. The size of cubic descriptor remains reasonable.

B. Global descriptor performances

The performances of the global descriptors are detailled in Tab. II where:

- GM is the percentage of tests where the correct image is found
- GCM indicates the percentage of tests where the correct image belongs to the set of selected images. For a hierarchical approach, this is the main indicator in term of accuracy. It is compared to the GCM of the cubic approach GCM_C thanks to the ratio rg=GCM/GCM_C.

- t indicates the needed computational cost measured in seconds and it is compared with the computational cost of the cubic approach t_C by means of the ratio rt=t/t_C.
- extr. is the mean number of kept images and is also important for computational cost issue since the second step of hierarchical approaches is applied only on these images.

	GM(%)	GCM(%)	rg(%)	t(s)	rt	extr.
Almere dataset						
Sect	62.7	62.8	70.5	0.55	4.58	1
HistoSect	90.1	**94.4**	105.9	*4.58*	38.16	*1.26*
Triang	90.5	90.5	101.6	0.38	3.16	1
PHLAC	27.8	*32.7*	36.6	0.47	3.91	*1.27*
Gonz	86.9	88.5	99.3	4.09	34.08	1.07
Cub	89	89.1	100	0.12	1	1
UAV dataset						
Sect	94.1	94.1	96.7	0.1	2.5	1
HistoSect	84	94.1	96.7	*1.45*	36.25	*1.68*
Triang	96.8	**96.8**	99.1	0.11	2.75	1
PHLAC	42	*48.9*	50.2	0.12	3	1.2
Gonz	83.5	86.1	88.5	0.53	13.25	1.11
Cub	97.3	**97.3**	100	**0.04**	1	1
Walk dataset						
Sect	80.8	81.1	96.7	0.28	2	1
HistoSect	69.2	**86.9**	103.7	*4.49*	32.07	*2.82*
Triang	83.8	83.8	100	0.28	2	1
PHLAC	22.6	*28*	33.5	0.25	1.78	1.5
Cub	83.1	83.8	100	**0.14**	1	1

TABLE II
COMPARISON OF THE GLOBAL DESCRIPTORS

As detailled in Tab.II, the GCM corresponding to *PHLAC* and *Gonz* approaches are very low. *HistoSect* is the more accurate in term of GCM-indicator for *Almere* and *Walk* data sets. Unfortunately, the number of kept images extr. is relatively high. Moreover, this approach as a high t-indicator. *Triang* gives good results for *Almere* and *Walk* data sets but it has a higher computational cost than *Cub* approach. The cubic interpolation *Cub* gives the best results for *UAV* data set and it is the best compromise between accuracy, computational cost and number of kept images for *Almere* and *Walk* data sets.

C. Local descriptor performances

SURF, SIFT and *CubHarris* are considered. The two main aspects are: the percentage (GM) which measures the accuracy of the method and the computational cost t. Execution times are compared with results of *CubHarris* with the ratio: rt=t/$t_{CorrHar}$.

SIFT gives the best results for *Almere* and *Walk* data sets but the computational cost is very high (as expected). SURF is computed approximately 7 times faster than SIFT. For *UAV* and *Walk* data sets, the execution time of *SURF* and *CubHarris* is similar. *CubHarris* gives the best results in UAV data set and it is the best compromise between computational cost and accuracy.

D. All descriptors

In view of Sections IV-B and IV-C, a hierarchical approach based on a cubic interpolation in a first step and on *CubHarris* in a second step seems to be the best localization approach in terms of accuracy and efficiency.

	GM(%)	t(s)	rt
Almere dataset			
SURF	93.4	**4**	0.43
SIFT	**93.6**	*31.5*	3.46
CorrHar	*91.5*	9.1	1
UAV dataset			
SURF	91.4	1	0.83
SIFT	*90.4*	7.1	5.91
CorrHar	**96.8**	**1.2**	1
Walk dataset			
SURF	*88.7*	**8.5**	0.89
SIFT	**92.5**	*152.1*	16.01
CorrHar	91.2	9.5	1

TABLE III
COMPARISON OF THE LOCAL DESCRIPTORS.

An example of image retrieval (*Walk* dataset) is shown in Fig. 4. Only some successive images of the database are presented for clarity. The test image is given in Fig. 4(a). The expected localization result is $\mathbf{I}_{17}$ (Fig. 4(g)) and the actual results are: $\mathbf{I}_{13}$ for *SIFT*, $\mathbf{I}_{14}$ for *SURF*, $\mathbf{I}_{15}$ for *Gonz*, $\mathbf{I}_{17}$ for *CubCorrHar* and $\mathbf{I}_{25}$ for *PHLAC*.

The whole results are presented in Tab. IV where GM and t are defined in Section IV-B. Comparisons with results of the proposed hierarchical method (*CubCorrHar*) are given with the ratio: rg=GM/$GM_{CubCorrHar}$ and rt=t/$t_{CubCorrHar}$.

SIFT gives the best results for *Almere* and *Walk* data sets but it has a very high computational cost. *SURF* is a good compromise for *Almere* data set but is less accurate for the two other data sets. Considering the whole data sets, *CubCorrHar* gives the best compromise between computational time and accuracy.

V. CONCLUSION

A new efficient localization method in a memory of omnidirectional images has been proposed. It combines global descriptors computed onto cubic interpolation of triangular mesh which is computationally efficient and patches correlation around Harris corners to ensure accuracy. This method has been compared to state-of-the-art techniques. The obtained results show that the proposed method is the best compromise between accuracy, amount of memorized data and computational cost.
Future work will be devoted to combine this approach with partial 3D reconstruction technique for metric localization.

ACKNOWLEDGMENT

This work is supported by the EU-Project FP6 IST μDrones, FP6-2005-IST-6-045248.

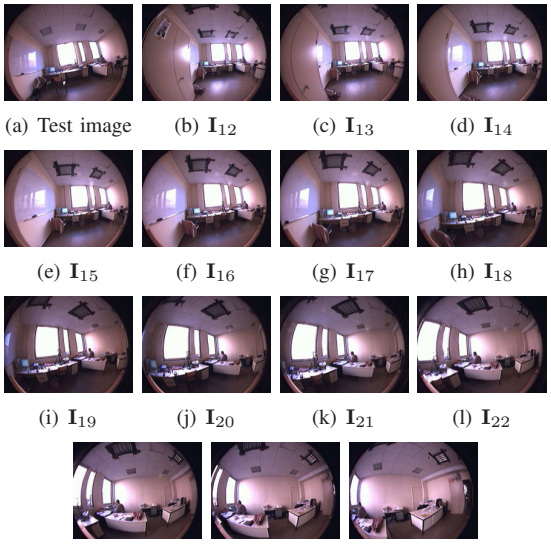
REFERENCES

[1] G. Blanc, Y. Mezouar, and P. Martinet, "Indoor navigation of a wheeled mobile robot along visual routes," in *IEEE International Conference on Robotics and Automation, ICRA'05*, Barcelona, Spain, 2005, pp. 3365–3370.

	GM	$rg(\%)$	t(s)	rt
Almere dataset				
Sect	62.7	68.5	0.55	0.11
SURF	93.4	102.1	4.06	0.87
SIFT	**93.6**	102.4	*31.5*	6.75
HistoSect	90.1	98.5	4.58	0.98
Cub	89	97.3	**0.12**	0.02
Triang	90.5	99	0.38	0.08
PHLAC	*27.8*	30.4	0.47	0.1
CorrHar	91.5	100.1	9.13	1.95
Gonz	86.9	95	4.09	0.87
SectCorrHar	85.1	93.1	2.07	0.44
CubCorrHar	91.4	100	4.66	1
UAV dataset				
Sect	94.1	97.2	0.1	0.19
SURF	91.4	94.4	1.04	2.07
SIFT	90.4	93.3	*7.1*	14.14
HistoSect	84	86.7	1.45	2.88
Cub	**97.3**	100.5	**0.04**	0.07
Triang	96.8	100	0.11	0.21
PHLAC	*42*	43.3	0.12	0.23
CorrHar	96.8	100	1.22	2.42
Gonz	83.5	86.2	0.53	1.05
SectCorrHar	95.2	98.3	0.12	0.23
CubCorrHar	96.8	100	0.5	1
Walk dataset				
Sect	80.8	88.4	0.28	0.05
SURF	88.7	97	8.59	1.69
SIFT	**92.5**	101.2	*152.1*	30.07
HistoSect	69.2	75.7	4.49	0.88
Cub	83.1	90.9	**0.14**	0.02
Triang	83.8	91.6	0.28	0.05
PHLAC	*22.6*	24.7	0.25	0.04
CorrHar	91.2	99.7	9.53	1.88
Gonz	61.1	66.8	1.56	0.3
SectCorrHar	84.9	92.8	0.54	0.1
CubCorrHar	91.4	100	5.05	1

TABLE IV

COMPARISON OF ALL THE DESCRIPTORS



(a) Test image (b) $\mathbf{I}_{12}$ (c) $\mathbf{I}_{13}$ (d) $\mathbf{I}_{14}$

(e) $\mathbf{I}_{15}$ (f) $\mathbf{I}_{16}$ (g) $\mathbf{I}_{17}$ (h) $\mathbf{I}_{18}$

(i) $\mathbf{I}_{19}$ (j) $\mathbf{I}_{20}$ (k) $\mathbf{I}_{21}$ (l) $\mathbf{I}_{22}$

(m) $\mathbf{I}_{23}$ (n) $\mathbf{I}_{24}$ (o) $\mathbf{I}_{25}$

Fig. 4. A subset of *Walk* data set.

[2] A. Murillo, J. Guerrero, and C. Sagüés, "SURF features for efficient robot localization with omnidirectional images," in *IEEE International Conference on Robotics and Automation, ICRA'07*, Rome, Italy, April 2007, pp. 3901–3907.

[3] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *International Journal of Computer Vision, special joint issue on vision and robotics*, vol. 74, pp. 237–260, 2007.

[4] J. Courbon, Y. Mezouar, L. Eck, and P. Martinet, "A generic fisheye camera model for robotic applications," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'07*, San Diego, California, USA, October 2007.

[5] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue, "Visual navigation using omnidirectional view sequence," in *International Conference on Intelligent Robots and Systems, IROS'99*, vol. 1, Korea, 1999, pp. 317–322.

[6] F. Linåker and M. Ishikawa, "Rotation invariant features from omnidirectional camera images using a polar higher-order local autocorrelation feature extractor," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'04*, vol. 4, Sendai, Japan, September 2004, pp. 4026–4031.

[7] T. Goedemé, T. Tuytelaars, G. Vanacker, M. Nuttin, L. V. Gool, and L. V. Gool, "Feature based omnidirectional sparse visual path following," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Canada, August 2005, pp. 1806–1811.

[8] A. Tamimi, H. Andreasson, A. Treptow, T. Duckett, and A. Zell, "Localization of mobile robots with omnidirectional vision using particle filter and iterative SIFT," in *2nd European Conference on Mobile Robots (ECMR)*, Ancona, Italy, September 2005, pp. 2–7.

[9] J. Gonzalez-Barbosa and S. Lacroix, "Rover localization in natural environments by indexing panoramic images," in *IEEE International Conference on Robotics and Automation, ICRA'02*, vol. 2, Washington DC, USA, May 2002, pp. 1365–1370.

[10] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro, "Hierarchical image-based localisation for mobile robots with monte-carlo localisation," in *European Conference on Mobile Robots, ECMR'03*, Warsaw, Poland, September 4-6 2003.

[11] T. Pajdla and V. Hlaváč, "Zero phase representation of panoramic images for image based localization," in *8-th International Conference on Computer Analysis of Images and Patterns*, F. Solina and A. Leonardis, Eds., no. 1689 in Lecture Notes in Computer Science. Springer Verlag, 1999, pp. 550–557.

[12] J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-based navigation and environmental representations with an omnidirectional camera," in *VisLab-TR 12/2000 - IEEE Transaction on Robotics and Automation*, vol. 16, no. 6, December 2000, pp. 890–898.

[13] P. Blaer and P. Allen, "Topological mobile robot localization using fast vision techniques," in *IEEE International Conference on Robotics and Automation*, Washington, USA, May 2002, pp. 1031–1036.

[14] A. C. Murillo, C. Sagüés, J. J. Guerrero, T. Goedemé, T. Tuytelaars, and L. V. Gool, "From omnidirectional images to hierarchical localization," *Robotics and Autonomous Systems*, vol. 55, no. 5, pp. 372–382, May 2007.

[15] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. [Online]. Available: http://www.cs.ubc.ca/~ lowe/keypoints/

[16] H. Andreasson, A. Treptow, and T. Duckett, "Localization for mobile robots using panoramic vision, local features and particle filter," in *IEEE International Conference on Robotics and Automation, ICRA'05*, Barcelona, Spain, April 2005, pp. 3348–3353.

[17] T. Svoboda and T. Pajdla, "Matching in catadioptric images with appropriate windows and outliers removal," in *9th International Conference on Computer Analysis of Images and Patterns*, Berlin, Germany, September 2001, pp. 733–740.

[18] S. Ieng, R. Benosman, and J. Devars, "An efficient dynamic multi-angular feature points matcher for catadioptric views," in *Workshop OmniVis'03, in conjunction with Computer Vision and Pattern Recognition (CVPR 2003)*, vol. 07, Wisconsin, USA, June 2003, p. 75.

[19] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence Journal*, vol. 78, pp. 87–119, October 1995.

[20] A. Murillo, J. Guerrero, and C. Sagüés, "Topological and metric robot localization through computer vision techniques," in *ICRA'07, Workshop: From features to actions - Unifying perspectives in computational and robot vision*, Rome, Italy, April 2007.

[21] P.-O. Persson and G. Strang, "A simple mesh generator in MATLAB," in *SIAM Review*, vol. 46, no. 2, June 2004, pp. 329–345.

[22] Data Set, "Workshop- from sensors to human spatial concepts- fs2hsc-data," IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'06, October 2006, last view: June 2007. [Online]. Available: http://staff.science.uva.nl/~ zivkovic/FS2HSC/dataset.html