Evaluation of Local Descriptors for Vision-Based Localization of Humanoid Robots

Noé G. Aldana-Murillo(✉), Jean-Bernard Hayet, and Héctor M. Becerra

Centro de Investigación en Matemáticas (CIMAT),
C.P. 36240 Guanajuato, GTO, Mexico
{noe.aldana,jbhayet,hector.becerra}@cimat.mx

Abstract. In this paper, we address the problem of appearance-based localization of humanoid robots in the context of robot navigation using a visual memory. This problem consists in determining the most similar image belonging to a previously acquired set of key images (visual memory) to the current view of the monocular camera carried by the robot. The robot is initially kidnapped and the current image has to be compared with the visual memory. We tackle the problem by using a hierarchical visual bag of words approach. The main contribution of the paper is a comparative evaluation of local descriptors to represent the images. Real-valued, binary and color descriptors are compared using real datasets captured by a small-size humanoid robot. A specific visual vocabulary is proposed to deal with issues generated by the humanoid locomotion: blurring and rotation around the optical axis.

Keywords: Vision-based localization · Humanoid robots · Local descriptors comparison · Visual bag of words

1 Introduction

Recently, the problem of navigation of humanoid robots based only on monocular vision has raised much interest. Many research has been reported for this problem in the context of wheeled mobile robots. In particular, the visual memory approach [1] has been largely studied. It mimics the human behavior of remembering key visual information when moving in unknown environments, to make the future navigation easier.

Robot navigation based on a visual memory consists of two stages [1]. First, in a learning stage, the robot creates a representation of an unknown environment by means of a set of key images that forms the so-called visual memory. Then, in an autonomous navigation stage, the robot has to reach a location associated to a desired key image by following a visual path. That path is defined by a subset of images of the visual memory that topologically connects the most similar key image compared with the current view of the robot with the target image.

Few work has been done for humanoids navigation based on a visual memory [2,3]. In both works, the robot is not initially kidnapped but it starts the navigation from a known position. In this context, the main interest in this paper

J.A. Carrasco-Ochoa et al. (Eds.): MCPR 2015, LNCS 9116, pp. 179–189, 2015.
DOI: 10.1007/978-3-319-19264-2_18

is to solve an appearance-based localization problem, where the current image is matched to a known location only by comparing images [4]. In particular, we address the localization of humanoid robots using only monocular vision.

This paper addresses the problem of determining the key image in a visual memory that is the most similar in appearance to the current view of the robot (input image). Figure 1 presents a general diagram of the problem. Consider that the visual memory consists of n ordered key images $(\mathcal{I}_1^*, \mathcal{I}_2^*, ..., \mathcal{I}_n^*)$. The robot is initially kidnapped and the current view $\mathcal{I}$ has to be compared with the n key images and the method should give as an output the most similar key image $\mathcal{I}_o^*$ within the visual memory.

Since a naive comparison might take too much time depending on the size of the visual memory, we propose to take advantage of a method that compresses the visual memory into a compact, efficient to access representation: the visual bag of words (VBoW) [5]. A bag of words is a structure that represents an image as a numerical vector, allowing fast images comparisons. In robotics, the VBoW approach has been used in particular for loop-closure in simultaneous localization and mapping (SLAM) [6,7], where re-visited places have to be recognized.

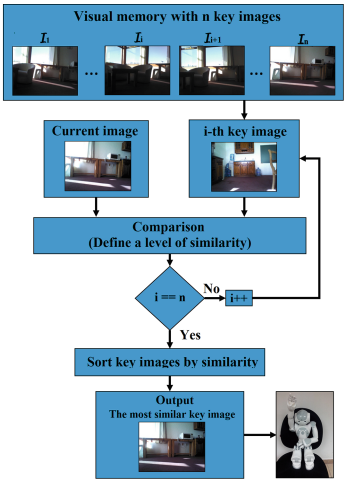

Fig. 1. General diagram of the appearance-based localization from a visual memory.

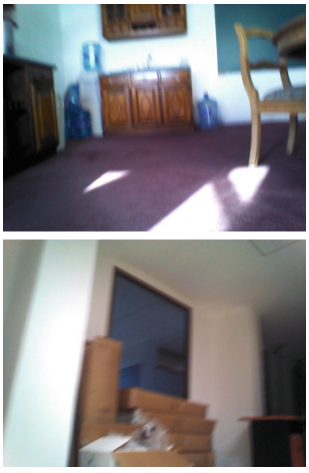

Fig. 2. Example of images from an onboard camera of a humanoid NAO robot.

In this paper, a quantitative evaluation of the VBoW approach using different local descriptors is carried out. In particular, we evaluate the approach on real datasets captured by a camera mounted in the head of a small-size humanoid robot. The images are affected by issues related to the sway motion introduced by the humanoid locomotion: blurring and rotation around the optical axis. A specific visual vocabulary is proposed to tackle those issues. Figure 2 shows two

examples of images captured from our experimental platform: a NAO humanoid robot. These images are of 640×480 pixels.

The paper is organized as follows. Section 2 introduces the local descriptors included in our evaluation. Section 3 details the VBoW approach as implemented. Section 4 presents the results of the experimental evaluation and Sect. 5 gives some conclusions.

2 Local Descriptors

Local features describe regions of interest of an image through descriptor vectors. In the context of image comparison, groups of local features should be robust against occlusions and changes in view point, in contrast to global methods. From the existing local detectors/descriptors, we wish to select the best option for the specific task of appearance-based humanoids localization. Hereafter, we introduce the local descriptors selected for a comparative evaluation.

2.1 Real-Valued Descriptors

A popular keypoint detector/descriptor is SURF (Speeded Up Robust Features) [8]. It has good properties of invariance to scale and rotation. SURF keypoints can be computed and compared much faster than their previous competitors. Thus, we selected SURF as a real-valued descriptor to be compared in our localization framework. The detection is based on the Hessian matrix and uses integral images to reduce the computation time. The descriptor combines Haar-wavelet responses within the interest point neighborhood and exploits integral images to increase speed. In our evaluation, the standard implementation of SURF (vector of dimension 64) included in the OpenCV library is used.

2.2 Binary Descriptors

Binary descriptors represent image features by binary strings instead of floating-point vectors. Thus, the extracted information is very compact, occupies less memory and can be compared faster. Two popular binary descriptors have been selected for our evaluation: Binary Robust Independent Elementary Features (BRIEF [9]) and Oriented FAST and Rotated BRIEF (ORB [10]). Both use variants of FAST (Features from Accelerated Segment Tests) [11], i.e. they detect keypoints by comparing the gray levels along a circle of radius 3 to the gray level of the circle center. In average, most pixels can be discarded soon, hence the detection is fast. BRIEF uses the standard FAST keypoints while ORB uses oFAST keypoints, an improved version of FAST including an adequate orientation component. The BRIEF descriptor is a binary vector of user-choice length where each bit results from an intensity comparison between some pairs of pixels within a patch around keypoints. The patches are previously smoothed with a Gaussian kernel to reduce noise. They do not include information of rotation or scale, so they are hardly invariant to them. This issue can be overcome by

using the rotation-aware BRIEF descriptor (ORB), which computes a dominant orientation between the center of the keypoint and the intensity centroid of its patch. The BRIEF comparison pattern is rotated to obtain a descriptor that should not vary when the image is rotated in the plane. In our evaluation, we use oFAST keypoints given by the ORB detection method as implemented in OpenCV along with BRIEF with size of patches 48 and descriptor length 256. The ORB implementation is the one of OpenCV with descriptors of 256 bits.

2.3 Color Descriptors

We also evaluate the image comparison approach by using only color information. To do so, we use rectangular patches and a color histogram is associated to each patch as a descriptor. We select the color space HSL (Hue-Saturation-Lightness) because its three components are more natural to interpret and less correlated than in other color spaces. Also, only the H and S channels are used, in order to achieve robustness against illumination changes. The color descriptor of each rectangular patch is formed by a two-dimensional histogram of hue and saturation and the length of the descriptor was set experimentally to 64 bits. Three different alternatives are evaluated using color descriptors:

- Random patches: 500 patches of size 48×64, randomly selected. This option is referred to as Color-Random.
- Uniform grid: A uniform grid of 19×19 patches covering the image, with patches overlapped a half of their size. This option is referred to as Color-Whole.
- Uniform grid on half of the image: Instead of using the whole image, only the upper half is used. This is because the inferior parts, when taken by the humanoid robot, are mainly projections of the floor and do not discrimante well for localization purposes. This option is referred to as Color-Half.

3 Visual Bag of Words for Humanoid Localization

As mentioned above, this work relies on the hierarchical visual bag of words approach [12] to combine the high descriptive power of local descriptors with the versatility and robustness of histograms. In Sect. 3.1, we recall the main characteristics of [12], and then in Sect. 3.2, we introduce a novel use of the BRIEF descriptor suited within a VBoW approach in the context of humanoid robots.

3.1 Hierarchical Visual Bag of Words Approach

The visual bag of words approach first discretizes the local descriptors space in a series of words, i.e., clusters in the local descriptors space. Here, we followed the strategy of Nister et al. [12], who perform this step in a hierarchical way: in the set of n key images $\mathcal{I}_1^*, \mathcal{I}_2^*, ..., \mathcal{I}_n^*$ forming the visual memory, a pool of D

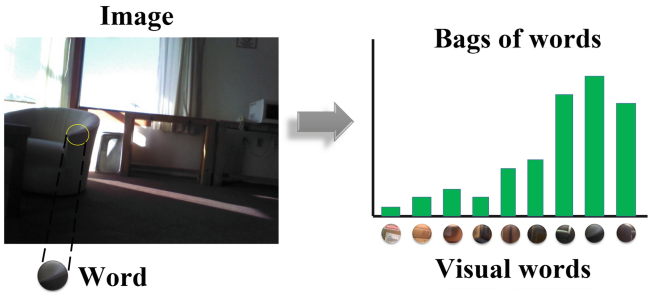

Fig. 3. Representation of an image in visual words.

local descriptors is detected, as illustrated in Fig. 3, left. The local descriptors can be extracted by any of the methods mentioned before. Given a branch factor k (a small integer, typically 3 or 4), the idea is to form k clusters among the D descriptors by using the kmeans++ algorithm. Then, the sets of descriptors associated to these k clusters are recursively clustered into other k clusters, and so on, up to a maximum depth of L levels. At each level, the formed clusters are associated to a representative descriptor chosen randomly (by the kmeans++ algorithm) that will be compared with new descriptors. The leaves of this tree of recursively refined clusters correspond to the visual words, i.e., the clusters in the local descriptors space. The advantage is that, when faced with descriptors found in new images, it is computationally efficient to associate them to a visual word, namely with kL distance computations, i.e., k at each level. Since we obtain $W = k^L$ leaves (i.e., words), characterizing a descriptor as a word is done in $O(k \log_k W)$ operations, where W is the number of words, instead of the W computations with a naive approach. This principle is illustrated in Fig. 4.

When handling a new image $\mathcal{I}$, d descriptors are extracted, and each of these is associated to a visual word as explained. This way, we obtain an empirical distribution of the visual words in $\mathcal{I}$, in the form of a histogram of visual words $v(\mathcal{I})$ (see Fig. 3, right). Now the content of $\mathcal{I}$ can be compared with the one of any of the

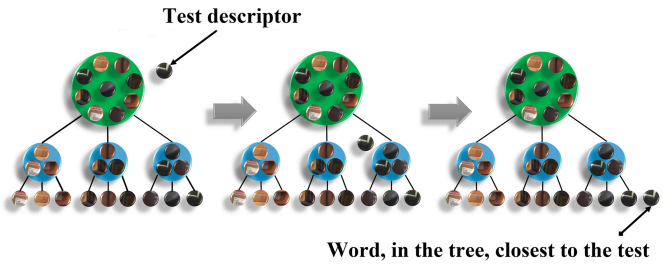

Fig. 4. Hierarchical approach of visual words search: When a new descriptor is found in some image $\mathcal{I}$, it is recursively compared to representatives of each cluster.

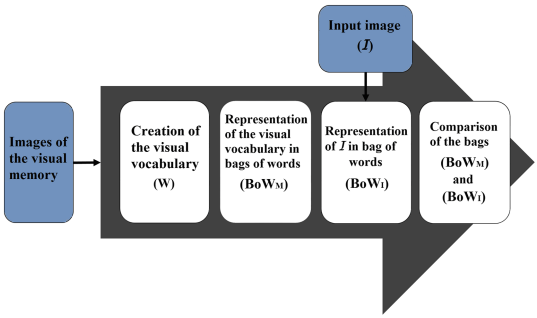

Fig. 5. Complete method of image comparison based on visual bag of words.

key images $\mathcal{I}_i^*$ by comparing their histograms. Of course, because n may be very high, it is out of question to compare the histogram of $\mathcal{I}$ with the n histograms of the key images. That is why an important element in this representation is the notion of inverse dictionary: for each visual word, one stores the list of images containing this word. Then, on a new image, we can easily determine, for each visual word it contains, the list of key images also containing this word. To limit the comparisons, we restrain the search for the most similar images to the subset of key images having at least 5 visual words in common. For an image $\mathcal{I}$, each histogram entry v_i (where i refers to the visual word) is defined as:

$$v_i(\mathcal{I}) = \frac{c_i(\mathcal{I})}{c(\mathcal{I})} \log(\frac{n}{n_i})$$

where $c(\mathcal{I})$ is the total number of descriptors present in $\mathcal{I}$, $c_i(\mathcal{I})$ the numbers of descriptors in $\mathcal{I}$ classified as word i, and n_i the number of key images where the word i has been found. The log term allows to weight the frequency of word i in $\mathcal{I}$ in function of its overall presence: When a word is present everywhere in the database, then the information of its presence is not that pertinent.

Last, we should choose how to compare histograms. After intensive comparisons made among the most popular metrics for histograms, we have chosen the χ^2 distance, that compares two histograms v and w through:

$$\chi^2(v, w) = \sum_{i=0}^{W} \frac{(\hat{v}_i - \hat{w}_i)^2}{\hat{v}_i + \hat{w}_i},$$

where $\hat{v} = \frac{1}{\|v\|_1} v$. Fig. 5 sums up the whole methodology.

3.2 A BRIEF-based Vocabulary for Humanoids Localization

We introduce a novel use of the BRIEF descriptor suited within a VBoW approach in the context of humanoid robots. This is a specific vocabulary that we called BRIEFROT, which deals with the issues generated by the humanoid locomotion. BRIEFROT possesses three independent internal vocabularies, two of

which are rotated a fixed angle: one anti-clockwise, the other clockwise. Through experimentation, we found that for the NAO humanoid platform a suitable value for the rotation of the vocabularies is 10 degrees. These rotated vocabularies were implemented with the idea of settling the slight variations in the rotation caused by the locomotion of these robotic systems. The rotated vocabularies represent to the images of the visual memory as rotated images. The third vocabulary is identical to the normal BRIEF. The idea of using three vocabularies is that if the input image is rotated with respect to any image of the visual memory, then the image is detected by any of the rotated vocabularies; if the input image is not rotated with respect to any image of the visual memory, then it is detected with the vocabulary without rotation. Additionally, the detected local patches are smoothed with a Gaussian kernel to reduce the blur effect.

4 Experimental Evaluation

We evaluated the local descriptors mentioned in Sect. 2 on 4 datasets. We used three datasets in indoor environments (CIMAT-NAO-A, CIMAT-NAO-B and Bicocca) and one outdoors (New College). The tests were done in a laptop using Ubuntu 12.04 with 4 Gb of RAM and 1.30 GHz processor.

4.1 Description of the Evaluation Datasets

The *CIMAT-NAO-A* dataset was acquired with a NAO humanoid robot inside CIMAT. This dataset contains 640×480 images of good quality but also blurry ones. Some images are affected by rotations introduced by the humanoid locomotion or by changes of lighting. We used 187 images, hand-selected, as a visual memory and 258 images for testing. The *CIMAT-NAO-B* dataset was also captured indoors at CIMAT with the humanoid robot. It also contains good quality and blurry 640×480 images, but it does not have images with drastic light changes, as in the previous dataset. We used 94 images as a visual memory and 94 images for testing. Both datasets *CIMAT-NAO-A* and *CIMAT-NAO-B* are available in http://personal.cimat.mx:8181/~hmbecerra/CimatDatasets.zip.

The *Bicocca 2009-02-25b* dataset is available online [13] and was acquired by a wheeled robot inside a university. The 320×240 images have no rotation around the optical axis nor blur. We used 120 images as a visual memory and 120 images for testing. Unlike the three previous datasets that were obtained indoors, the *New College* dataset was acquired outside the Oxford University by a wheeled robot [14], with important light changes. The 384×512 images are of good quality with no rotation nor blur. For this dataset, 122 images were chosen as a visual memory and 117 images for testing.

4.2 Evaluation Metrics

Since the goal of this work is to evaluate different descriptors in VBoW approaches, it is critical to define corresponding metrics to assess the quality of the result from our application. We propose two metrics; the first one is:

$$\mu_1(\mathcal{I}) = \text{rank}(\bar{k}(\mathcal{I}))$$

where $\bar{k}(\mathcal{I})$ is defined as the ground truth index of the key image associated to $\mathcal{I}$. In the best case, the rank of the closest image to ours should be one, so $\mu_1(\mathcal{I}) = 1$ means that the retrieval is perfect, whereas higher values correspond to worse evaluations. The second metric is:

$$\mu_2(\mathcal{I}) = \sum_l \frac{z_l(\bar{k}(\mathcal{I}))}{\sum_{l'} z_{l'}(\bar{k}(\mathcal{I}))} \text{rank}(l)$$

where the $z_l(k)$ is the similarity score between the key images k and l inside the visual memory. This metric is proposed to handle similar key images within the dataset; hence, with this metric, the final score integrates weights (normalized by $\sum_{l'} z_{l'}(\bar{k}(\mathcal{I}))$ to sum to one) from the key images l similar to the closest ground truth image $\bar{k}(\mathcal{I})$; this ensures that all the closest images are well ranked.

4.3 Parameters Selection

There are three free parameters: the number of clusters k, the tree depth L and the measure of similarity. Tests were performed by varying k from $k = 8$ to $k = 10$ and varying L from $L = 4$ to $L = 8$. Also, different similarity measures between histograms were tested: L1-Norm, L2-Norm, χ^2, Bhattacharyya and dot product. To do the tests, we generated ground truth data, by defining manually the most similar key image to the input image. The parameters were selected so that the confidence levels μ_2 were close to 1. We obtained the best results with $k = 8$, $L = 8$ and the metric χ^2. The dataset used for the parameters selection was the CIMAT-NAO-A, since it is the most challenging dataset for the type of images it contains.

4.4 Analysis of the Results Obtained on the Evaluation Datasets

We present the results in the following tables for the seven vocabularies created. On the one hand, the efficiency of the vocabularies is observed using the confidence μ_1. In this case, the threshold chosen for the test to be classified as correct was 1. On the other hand, for the level of confidence μ_2 we choose a threshold of 2.5, all tests below 2.5 were considered correct. This level of confidence takes into account the possible similarity between the images on the visual memory.

In Table 2, we present the results obtained for the *CIMAT-NAO-A* dataset. In this case, the BRIEFROT vocabulary obtained the best behavior for both levels of confidence. For the case of μ_1, it has an efficiency of 60.85 % and for μ_2 of 75.19 %. Also, the ORB vocabulary offered good performance for μ_2 and was the second best for μ_1. The Color-Half vocabulary obtained the worst results. On the other hand, for the *CIMAT-NAO-B* dataset, the SURF vocabulary behaved better than BRIEFROT, but with higher computation times. The times reported were measured from the stage of features extraction to the stage of comparison. ORB, again, behaved well and was the second best vocabulary. The Color-Random vocabulary obtained the worst performance for μ_1, but for μ_2 it was

Table 1. Percentages of correct results for the dataset CIMAT-NAO-A.

Descriptor	Number of tests	Correct tests μ_1	Effectiveness μ_1 (%)	Correct tests μ_2	Effectiveness μ_2 (%)	Average time comparison (ms)
BRIEF	258	132	51.16	185	71.71	122.6
BRIEFROT	258	157	60.85	194	75.19	132.4
Color-Random	258	110	42.64	117	45.35	129.4
Color-Half	258	104	40.31	160	62.01	93.1
Color-Whole	258	110	42.64	162	62.79	101.8
ORB	258	144	55.81	194	75.19	107.5
SURF	258	135	52.32	187	72.48	296.5

Table 2. Percentages of correct results for the dataset CIMAT-NAO-B.

Descriptor	Number of tests	Correct tests μ_1	Effectiveness μ_1 (%)	Correct tests μ_2	Effectiveness μ_2 (%)	Average time comparison (ms)
BRIEF	94	63	67.02	82	87.23	87.23
BRIEFROT	94	65	69.14	81	86.17	112.0
Color-Random	94	62	65.96	86	91.49	109.9
Color-Half	94	64	68.09	78	82.98	63.2
Color-Whole	94	68	72.34	83	88.3	73.1
ORB	94	69	73.40	83	88.3	77.7
SURF	94	70	74.46	86	91.49	267.9

Table 3. Percentages of correct results for the dataset Bicocca25b.

Descriptor	Number of tests	Correct tests μ_1	Effectiveness μ_1 (%)	Correct tests μ_2	Effectiveness μ_2 (%)	Average time comparison (ms)
BRIEF	120	111	92.5	116	96.67	73.4
BRIEFROT	120	111	92.5	116	96.67	98.5
Color-Random	120	60	50	69	57.50	72.7
Color-Half	120	57	47.5	67	55.83	36.0
Color-Whole	120	59	49.17	65	54.17	79.4
ORB	120	110	91.67	114	95.00	60.2
SURF	120	111	92.5	114	95.00	120.0

Table 4. Percentages of correct results for the dataset New College.

Descriptor	Number of tests	Correct tests μ_1	Effectiveness μ_1 (%)	Correct tests μ_2	Effectiveness μ_2 (%)	Average time comparison (ms)
BRIEF	117	70	59.83	85	72.65	105.9
BRIEFROT	117	70	59.83	87	74.36	134.6
Color-Random	117	48	41.02	69	58.97	110.9
Color-Half	117	34	29.06	64	54.70	60.4
Color-Whole	117	44	37.61	74	63.25	110.5
ORB	117	69	58.97	85	72.65	107.0
SURF	117	74	63.25	89	76.07	302.3

one of the best vocabulary; this means that it tends to put the correct key image in the second rank. Color-Half had the worst results for μ_2.

In the *Bicocca 2009-02-25b* dataset, three vocabularies obtained the best results for μ_1: BRIEFROT, Color-Random and SURF. The difference between these three vocabularies is in the computation time: SURF consumes much more time. For μ_2, BRIEFROT was the best. In the *New College* dataset, the SURF vocabulary obtained the best behavior for both levels of confidence. In both cases the BRIEFROT vocabulary obtained a good behavior, close to SURF, but BRIEFROT consumes less than half the time required by SURF (Tables 1, 3 and 4).

5 Conclusions

This paper addresses the problem of vision-based localization of humanoid robots, i.e., determining the most similar image among a set of previously acquired images (visual memory) to the current robot view. To this end, we use a hierarchical visual bag of words (VBoW) approach. A comparative evaluation of local descriptors to use to feed the VBoW is reported: Real-valued, binary and color descriptors were compared on real datasets captured by a small-size humanoid robot. We presented a novel use of the BRIEF descriptor suited to the VBoW approach for humanoid robots: BRIEFROT. According to our evaluation, the BRIEFROT vocabulary is very effective in this context, as reliable as SURF to solve the localization problem, but in much less time. We also show that keypoints-based vocabularies performed better than color-based vocabularies.

As future work, we will explore the combination of visual vocabularies to robustify the localization results. We will implement the method onboard the NAO robot using a larger visual memory. We also wish to use the localization algorithm in the construction of the visual memory to identify revisited places.

References

1. Courbon, J., Mezouar, Y., Martinet, P.: Autonomous navigation of vehicles from a visual memory using a generic camera model. IEEE Trans. Intell. Transp. Syst. **10**(3), 392–402 (2009)
2. Ido, J., Shimizu, Y., Matsumoto, Y., Ogasawara, T.: Indoor navigation for a humanoid robot using a view sequence. Int. J. Robot. Res. **28**(2), 315–325 (2009)
3. Delfin, J., Becerra, H.M., Arechavaleta, G.: Visual path following using a sequence of target images and smooth robot velocities for humanoid navigation. In: IEEE International Conference on Humanoid Robots, pp. 354–359 (2014)
4. Ulrich, I., Nourbakhsh, I.: Appearance-based place recognition for topological localization. In: IEEE International Conference on Robotics and Automation, pp. 1023–1029 (2000)
5. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: IEEE International Conference on Computer Vision, pp. 1–8 (2003)
6. Botterill, T., Mills, S., Green, R.: Bag-of-words-driven, single-camera simultaneous localization and mapping. J. Field Robot. **28**(2), 204–226 (2011)
7. Galvez-Lopez, D., Tardos, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Trans. Robot. **28**(5), 1188–1197 (2012)
8. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
9. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
10. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision, pp. 2564–2571 (2011)
11. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
12. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2161–2168 (2006)
13. Bonarini, A., Burgard, W., Fontana, G., Matteucci, M., Sorrenti, D.G., Tardos, J.D.: Rawseeds: robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In: International Conference on Intel, Robots and Systems (2006)
14. Smith, M., Baldwin, I., Churchill, W., Paul, R., Newman, P.: The new college vision and laser data set. Int. J. Robot. Res. **28**(5), 595–599 (2009)

Springer Link

Chapter

Pattern Recognition

Volume 9116 of the series *Lecture Notes in Computer Science* pp 179-189

Date: 04 June 2015

Evaluation of Local Descriptors for Vision-Based Localization of Humanoid Robots

- Noé G. Aldana-Murillo
- , Jean-Bernard Hayet
- , Héctor M. Becerra

Abstract

In this paper, we address the problem of appearance-based localization of humanoid robots in the context of robot navigation using a visual memory. This problem consists in determining the most similar image belonging to a previously acquired set of key images (visual memory) to the current view of the monocular camera carried by the robot. The robot is initially kidnapped and the current image has to be compared with the visual memory. We tackle the problem by using a hierarchical visual bag of words approach. The main contribution of the paper is a comparative evaluation of local descriptors to represent the images. Real-valued, binary and color descriptors are compared using real datasets captured by a small-size humanoid robot. A specific visual vocabulary is proposed to deal with issues generated by the humanoid locomotion: blurring and rotation around the optical axis.

Keywords

Vision-based localization Humanoid robots Local descriptors comparison Visual bag of words

References

1. Courbon, J., Mezouar, Y., Martinet, P.: Autonomous navigation of vehicles from a visual memory using a generic camera model. IEEE Trans. Intell. Transp. Syst. **10**(3), 392–402 (2009)
CrossRef *(http://dx.doi.org/10.1109/TITS.2008.2012375)*

2. Ido, J., Shimizu, Y., Matsumoto, Y., Ogasawara, T.: Indoor navigation for a humanoid robot using a view sequence. Int. J. Robot. Res. **28**(2), 315–325 (2009)
CrossRef *(http://dx.doi.org/10.1177/0278364908095841)*

3. Delfin, J., Becerra, H.M., Arechavaleta, G.: Visual path following using a sequence of target images and smooth robot velocities for humanoid navigation. In: IEEE International Conference on Humanoid Robots, pp. 354–359 (2014)

4. Ulrich, I., Nourbakhsh, I.: Appearance-based place recognition for topological localization. In: IEEE International Conference on Robotics and Automation, pp. 1023–1029 (2000)

5. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: IEEE International Conference on Computer Vision, pp. 1–8 (2003)

6. Botterill, T., Mills, S., Green, R.: Bag-of-words-driven, single-camera simultaneous localization and mapping. J. Field Robot. **28**(2), 204–226 (2011)
CrossRef *(http://dx.doi.org/10.1002/rob.20368)*

7. Galvez-Lopez, D., Tardos, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Trans. Robot. **28**(5), 1188–1197 (2012)
CrossRef *(http://dx.doi.org/10.1109/TRO.2012.2197158)*

8. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
CrossRef *(http://dx.doi.org/10.1016/j.cviu.2007.09.014)*

9. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
CrossRef *(http://dx.doi.org/10.1007/978-3-642-15561-1_56)*

10. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision, pp. 2564–2571 (2011)

11. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
CrossRef *(http://dx.doi.org/10.1007/11744023_34)*

12. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2161–2168 (2006)

13. Bonarini, A., Burgard, W., Fontana, G., Matteucci, M., Sorrenti, D.G., Tardos, J.D.: Rawseeds: robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In: International Conference on Intel, Robots and Systems (2006)

14. Smith, M., Baldwin, I., Churchill, W., Paul, R., Newman, P.: The new college vision and laser data set. Int. J. Robot. Res. **28**(5), 595–599 (2009)
CrossRef *(http://dx.doi.org/10.1177/0278364909103911)*

About this Chapter

Title

Evaluation of Local Descriptors for Vision-Based Localization of Humanoid Robots

Book Title

Pattern Recognition

Book Subtitle

7th Mexican Conference, MCPR 2015, Mexico City, Mexico, June 24-27, 2015, Proceedings

Pages

pp 179-189

DOI

10.1007/978-3-319-19264-2_18

Print ISBN

978-3-319-19263-5

Online ISBN

978-3-319-19264-2

Series Title

Lecture Notes in Computer Science

Series Volume

9116

Series ISSN

0302-9743

Publisher

Springer International Publishing

Additional Links

- *About this Book*

Topics

- *Pattern Recognition*
- *Artificial Intelligence (incl. Robotics)*
- *Image Processing and Computer Vision*
- *Information Systems Applications (incl. Internet)*
- *Data Mining and Knowledge Discovery*

Keywords

- Vision-based localization
- Humanoid robots
- Local descriptors comparison
- Visual bag of words

Industry Sectors

- *Electronics*
- *Telecommunications*
- *IT & Software*

eBook Packages

- *eBook Package english Computer Science*
- *eBook Package english full Collection*

Editors

- *Jesús Ariel Carrasco-Ochoa* ✉ (13)
- *José Francisco Martínez-Trinidad* ✉ (14)
- *Juan Humberto Sossa-Azuela* ✉ (15)
- *José Arturo Olvera López* ✉ (16)
- *Fazel Famili* ✉ (17)

Editor Affiliations

- 13. National Institute of Astrophysics, Optics, and Electronics
- 14. Nataional Institute of Astrophysics, Optics, and Electronics
- 15. National Polytechnic Institute of Mexico
- 16. Autonomous University of Puebla
- 17. University of Ottawa

Authors

- *Noé G. Aldana-Murillo* (18)
- *Jean-Bernard Hayet* (18)
- *Héctor M. Becerra* (18)

Author Affiliations

- 18. Centro de Investigación en Matemáticas (CIMAT), C.P. 36240, Guanajuato, GTO, Mexico