An Evaluation of New Global Appearance Descriptor Techniques for Visual Localization in Mobile Robots under Changing Lighting Conditions

Vicente Román[a], Luis Payá[b], Sergio Cebollada[c], Adrián Peidró[d] and Óscar Reinoso[e]
Engineering Systems and Automation Department, Miguel Hernandez University, Elche (Alicante), Spain

Keywords: Localization, Mobile Robots, Global Appearance Descriptors, Omnidirectional Images.

Abstract: Autonomous robots should be able to carry out localization and map creation in highly heterogeneous zones. In this work, global appearance descriptors are tested to perform the localization task. It focuses on the use of an omnidirectional vision sensor as unique source of information and global appearance to describe the visual information. Global-appearance techniques consist in obtaining a unique vector that describes globally the image. The main objective of this work is to propose and test new alternatives to build and to handle global descriptors. In previous experiments the images have been processed without considering the spatial distribution of the information. In contrast, in this work, the main approach is that relevant information will be in the central rows. For this reason central rows information is given a higher weight comparing to other zones of the image. The results show that this consideration can be an interesting presumption to take into account. The experiments are carried out with real images that have been taken in two different heterogeneous environments where simultaneously humans and robots work together. For this reason, variations of the lighting conditions, people who occlude the scene and changes on the furniture may appear.

1 INTRODUCTION

In the last decade, the use of visual information has extended to improve the autonomy of mobile robots in many applications. The presence of mobile robots in different environments has increased, and in order to cope with more challenging situations their abilities in perception and interpretation have improved. To be robustly autonomous in extended, heterogeneous and changing environments, the mobile robot has a twofold task. First, a mapping task in which it has to be able to navigate around the initialy unknown environment while creating a map. Second, it has to perform localization task trying to estimate its position and orientation in the environment. Among vision sensors, omnidirectional cameras are an interesting option to carry out these tasks due to their field of view of 360^{o} around the camera axis (Sturm et al., 2011) and (Payá et al., 2017).

[a] https://orcid.org/0000-0002-3706-8725
[b] https://orcid.org/0000-0002-3045-4316
[c] https://orcid.org/0000-0003-4047-3841
[d] https://orcid.org/0000-0002-4565-496X
[e] https://orcid.org/0000-0002-1065-8944

Due to the fact that images contain a big amount of data, it is required to extract from them relevant information. Nowadays local appearance descriptors are well-known and extensively used. These methods describe specific points or local zones in the image. Among these descriptors SIFT (Lowe, 2004) and SURF (Bay et al., 2008) are the most known and used. Murillo et al. (Murillo et al., 2007) solved a mobile robots navigation problem using local descriptors, Gil et al. (Gil et al., 2011) and Valiente et al. (Valiente García et al., 2012) worked with local appearance descriptors and omnidirectional cameras. Relatively good results in navigation have been obtained using local appearance descriptors. Global appearance descriptors are an alternative method to extract characteristic information from images and use this information for mapping and localization.

Global-appearance description methods describe the image globally obtaining a unique vector per image, which is expected to be more invariant against global changes in the scene. In addition, as each image is described with a unique vector, the mapping and localization work is simplified to a pairwise comparison between vectors. Over the past few years some global-appearance descriptors have been stud-

Román, V., Payá, L., Cebollada, S., Peidró, A. and Reinoso, Ó.
An Evaluation of New Global Appearance Descriptor Techniques for Visual Localization in Mobile Robots under Changing Lighting Conditions.
DOI: 10.5220/0009595603770384
In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2020)*, pages 377-384
ISBN: 978-989-758-442-8

ied. Gist, introduced by Oliva and Torralba (Oliva and Torralba, 2001), is one of the most extended descriptors, and it has been tested in outdoor environments for example by Zhou et al. (Zhou et al., 2018) to solve the localization through matching the robot's current view with the best keyframe in the database. Other option is the Histogram of Oriented Gradients (HOG), HOG is used in (Payá et al., 2018) to solve hierarchical mapping and localization tasks. In addition, there are other important techniques to obtain global appearance descriptors based on mathematical transformation, such as the Fourier Transform (Menegatti et al., 2004) or the Radon Transform (Radon, 2005). These alternatives have been used in works as Paya et al. (Payá et al., 2009) to build maps or in Berenguer et al. (Berenguer et al., 2019) where they used Radon Transform to estimate the relative height of a mobile robot. Moreover, during the last recent years, some authors have used deep learning techniques to create new global appearance descriptors. For example, Xu et al. (Xu et al., 2019) proposed a CNN-based descriptor to obtain the most probable robot position and Cebollada et al. (Cebollada et al., 2019) perform a comparison between analytic global-appearance descriptors and CNN-based descriptors while solving a mobile robot localization work. Finally, Román et al. (Román et al., 2018) studied some of these global appearance methods in real environments to solve the localization task under illumination changes.

As shown, global appearance descriptors are defined to be invariant against rotations in the ground plane when omnidirectional images are used. Global-appearance methods have summarised the information from the panoramic images in horizontal blocks or cells traditionally. But more recently, other ways to build the descriptor have appeared, for instance (Román et al., 2019) where vertical cells are evaluated with interesting results. The current work tries to go one step beyond in the definition of global-appearance descriptors, considering that usually, the most important is condensed in the horizontal cells situated in the middle of the panoramic image, because the visual information in the upper and bottom rows often corresponds to the roof or sky and floor or terrain, which are visually less significative. For this reason a technique that increases the weight of the central rows is studied. This work compares the classic formulation with a new technique while testing them in a localization framework.

2 GLOBAL APPEARANCE DESCRIPTORS

In this section a review of the global appearance descriptors used in the presented localization task is described. The goal of these methods is to extract a unique vector that globally describes the information from an image. In this way, relevant information is keep while reducing amount of memory. Global appearance descriptors have been used to perform robot navigation tasks, for example, to solve the kidnapped robot problem in indoor environments under different conditions (Su et al., 2017) or to build hierarchical maps through clustering algorithms (Cebollada et al., 2019). To perform the localization task, HOG and Gist descriptors are modified and used in this work. In both cases the starting point is a panoramic image $i(x,y) \in \mathbb{R}^{N_x \times N_y}$ and after these methods each image is reduced to a vector $\vec{d} \in \mathbb{R}^{l \times 1}$.

The first step to build the descriptor is divide the image in a set of cells. The descriptor size depends on the number of these cells. The first option studied in this work is the classic way, used in (Román et al., 2018) where the vector is built with uniformly distributed and non-overlapped horizontal cells, figure 1 shows how cells are distributed in this classical method. Taking into account the idea of giving more importance to the central rows, a method where the descriptors are built as traditionally but they are weighted by a set of factors is evaluated. These factors are obtained from a Gaussian distribution centred in the central row of the panoramic image, in such a way that the information of the central cells is given more importance than top or bottom cells. This idea is outlined in figure 2.

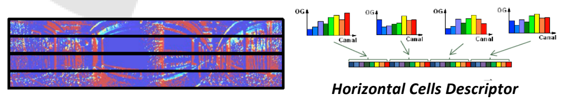

Figure 1: Classical approach to build the HOG global appearance descriptor of a panoramic image, by defining a set of horizontal cells.

Figure 2: Method to build the global-appearance descriptor multiplying descriptor by a gaussian distribution.

2.1 Histogram of Oriented Gradients, HOG

Histogram of Oriented Gradients was described by Dalal and Triggs (Dalal and Triggs, 2005) and used by Hofmeister et al. in small and controlled environments (Hofmeister et al., 2009). It basically consists in calculating the gradient of the image and obtaining magnitude and orientation of the gradient in each pixel. After that, magnitudes and orientations are evaluated and the descriptor is build by collecting together the information obtained in each cell. The methods divide the panoramic images using uniformly distributed and non-overlapped horizontal cells. As they collect information from horizontals rows and panoramic images are used, descriptors are invariant to rotations of the robot in the ground plane and they can be use in the localization tasks independently on the orientation that the robot has at a specific time instant.

The descriptor size depends on diverse parameters. This way, the vector length depends on the number of bins of the orientations histogram b and the number of cells in which the image is divided k_1. HOG descriptor reduces a panoramic image into a vector whose size is $\vec{d} \in \mathbb{R}^{b \cdot k_1 \times 1}$.

2.2 Gist

This descriptor was initially proposed by Oliva et al. (Oliva and Torralba, 2006) and it was developed by Siagian et. al. (Siagian and Itti, 2009) testing its performance in three different outdoor environments. This method exposes the image to a specific number of Gabor filters with different orientations in several resolution levels. After that, the images are reduced evaluating their mean intensity in different horizontal cells.

In this case, the descriptor size will depend on the number of orientations of Gabor filters m, the number of cells in which the images are split k_2 and the number of different resolution models used r. During the experiments this latter parameter r will be constant, r=2. For that reasons the descriptor is a vector whose size is $\vec{d} \in \mathbb{R}^{2 \cdot m \cdot k_2 \times 1}$.

2.3 Main Parameters of the Descriptors

The aim of this work is to check the efficiency of the descriptors in a localization task and to optimize the main parameters in this new description technique. Different parameters can be tuned and, as a consequence, the vector that describes the image modifies its values and size. These parameters can be seen in table 1. Different values have been tested but, as shown, these parameters define the descriptor size. The larger the descriptor is, the more information it contains but the slower the process will be.

Table 1: Parameters that impact on the location process.

Descriptor	Parameters
HOG	$b \Rightarrow$ number of bins per histogram. $k_1 \Rightarrow$ number of horizontal cells.
Gist	$m \Rightarrow$ number of Gabor filters. $k_2 \Rightarrow$ number of horizontal blocks. $r \Rightarrow$ different resolution models.

During these experiments r is constant, $r = 2$, the other parameters take values between [8-32]

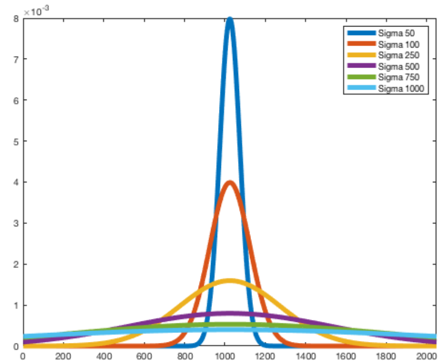

Figure 3: Gaussian distributions to obtain the coefficients to weight each cell, considering different values of sigma.

The mainly purpose of this work is to evaluate the new technique to build descriptors. As pointed out before, the most important information is situated on the central rows in a panoramic image and the top and bottom rows are visually less relevant because they contain the ceiling and the floor. For this reason, the descriptors are built using the traditional decomposition of the image in horizontal cells, but each cell is weighted with an importance coefficient. The set of coefficients is obtained from a Gaussian distribution that gives more importance to the central cells. Therefore, the information in these cells contributes to the final descriptor to a greater extent. Gaussian distribution can be built with different deviation values (σ). Figure 3 shows different Gaussian distributions used during experiments depending on σ, $\sigma = \{50, 100, 250, 500, 750, 1000\}$. Values in a gaussian distributions sum 1 and they are distributed depending on σ.

Figure 4: Real-life changes in an heterogeneous environment.

Table 2: Number of images and distance between consecutive images in each route depending on the environment and lighting conditions.

Trajectory Database	Number of images	Distance between Images
Test Route 1 (Cloudy)	2778	0.0370 ± 0.0149 m
Test Route 1 (Night)	2896	0.0357 ± 0.0192 m
Test Route 1 (Sunny)	2231	0.0462 ± 0.0213 m
Training Route 1 (Cloudy)	556	0.1835 ± 0.0594 m
Test Route 2 (Cloudy)	1450	0.1212 ± 0.0410 m
Training Route 2 (Cloudy)	750	0.2397 ± 0.0629 m

3 DATABASE

The experiments have been carried out through the use of the COLD dataset (Pronobis and Caputo, 2009) and INNOVA dataset (Amorós et al., 2018). COLD database offers three different indoor trajectories taken in three buidings (Freiburg, Saarbrücken and Ljubljana) and INNOVA database offers also another indoor laboratory trajectory the Miguel Hernandez University, Spain. While the COLD database trajectories were taken in different times of the day so they offer the same environment in three different illumination conditions, INNOVA trajectory was only captured under cloudy conditions (during the light hours but the sunlight does not considerably affect the shots). The selected databases offer the heterogeneous and dynamic environment needed to test the proposed global-appearance methods. The robot travels along some laboratories where people is normally working, so it has to deal with changes in the environment such as people walking or position of furniture and objects and also problems like lighting changes and occlusions. In the image 4 it is possible to see some shots taken from the database where it is possible to observe these real operation conditions.

Among the different trajectories offered by the datasets, three routes have been chosen as test datasets to carry out the experiments. **Route 1**: Freiburg Part A, Path 2, size 3 (Pronobis and Caputo, 2009), and **Route 2**: INNOVA (Amorós et al., 2018). In addition each dataset has its own training database. **Training database** covers the same parts of the environments with an average distance between images around 0.02 m. Other trajectory specifications can be seen in table 2, where it is possible to see the number of images and average distance between images. Within a selected route all the different specifications (cloudy, night and sunny) cover approximately the same trajectory, but they where taken in different moments. Training routes were taken in a cloudy environment. At the end, the route 1 covers approximately 103 m and route 2 176 m.

4 EXPERIMENTS

4.1 Model of the Environment

As explained in the previous section, each database offers a training trajectory which follows approximately the same route than the test ones. The distance between images in training dataset is around 0.20m and they were taken during the light hours, but the sunlight does not considerably affect the images (cloudy conditions).

Once the model is built with the descriptors of the training images, the method to solve the localization task consists in comparing the descriptor of each test image with the descriptors in the model. The program compares descriptors and calculates the nearest neighbor by means of the *correlation* distance $(d(\vec{a},\vec{b}) = 1 - \frac{\vec{a}_d^T \cdot \vec{b}_d}{|a_d| \cdot |b_d|})$. In this expression:
$\vec{a} \in \mathbb{R}^{lx1}$ and $\vec{b} \in \mathbb{R}^{lx1}$ where: $a_i, b_i, i = 1, ..., l$;
$\vec{a}_d = [a_1 - \overline{a}, ..., a_l - \overline{a}]; \overline{a} = \frac{1}{l} \cdot \sum_j \cdot a_j$ and
$\vec{b}_d = [b_1 - \overline{b}, ..., b_l - \overline{b}]; \overline{b} = \frac{1}{l} \cdot \sum_j \cdot b_j$

When the nearest neighbor is calculated the geometric distance between the capture point of the test image (ground truth) and the capture point of the nearest neighbor in the model is obtained, and the result is the error. This geometrical distance can be calculated because COLD and INNOVA databases offers the coordinates where each image had been taken but the coordinates have been only used as ground truth to check the error. The localization task is carried out with pure visual information.

4.2 Position Estimation

Initially, the training routes are used to create the reference model. Afterwards, to study the robustness of the global-appearance descriptors, the test images are used to solve the localization problem. The localization process evaluates which image in the training model is the most similar to each test image. This process has been carried out with the different lighting (cloudy, night and sunny). The error is calculated as the geometric distance between the capture points of both images. After repeating the process using different descriptors sizes, figures 5, 6 and 7 show the average error (m) using the classical method, with no weighting of the rows. Each figure shows the average localization error obtained after considering all the test images of a trajectory with specific lighting conditions.

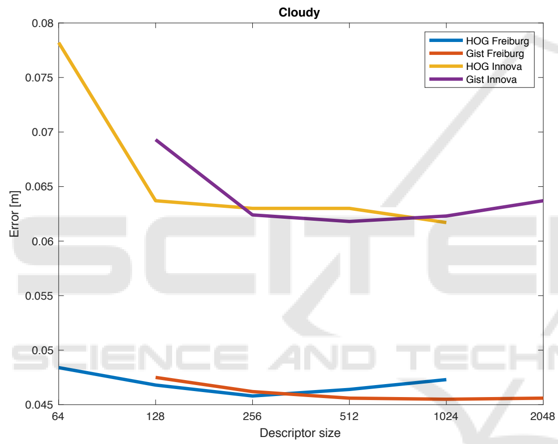

Figure 5: Localization error (m) using **cloudy** dataset versus descriptor size.

Both, HOG and Gist descriptor provide their best results when they work with middle size descriptors. Gist performs better under cloudy and night conditions, with errors between 0.0455 m and 0.2266 m respectively while HOG outputs an error of 0.0458 m with cloudy conditions and 0.2337 m with night environment. The sunny conditions have the most negative effect on the localization process, with an error of 1.903 m with Gist and 1.8675 m using HOG. On the other hand with INNOVA database only cloudy test images are available. Results are really similar and the lowest errors are 0.063 m using HOG descriptor and 0.0618 m using Gist.

The results of the proposed method, which includes the weighting coefficients, are shown in the next lines. The localization process is the same as shown before but now the information in the descriptors is weighted by a set of coefficients obtained from

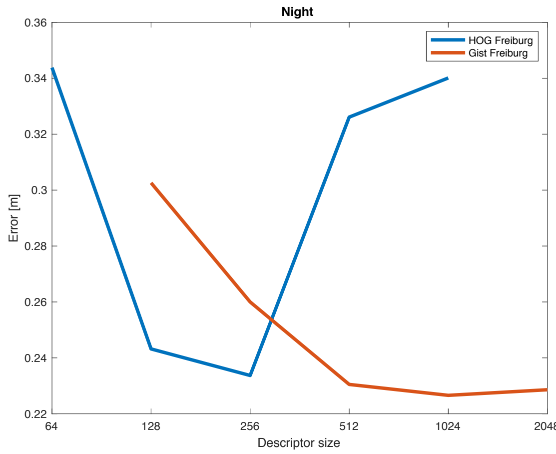

Figure 6: Localization error (m) using **night** dataset versus descriptor size.

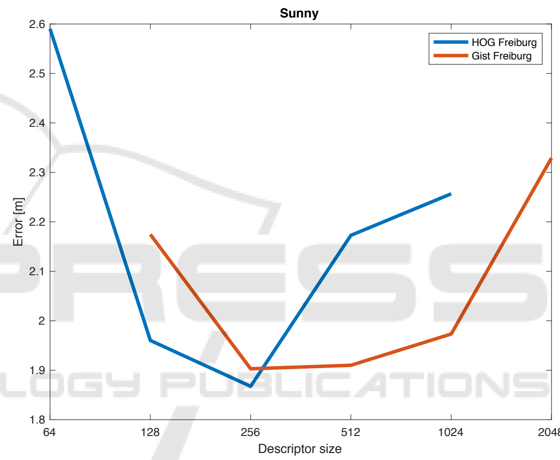

Figure 7: Localization error (m) using **sunny** dataset versus descriptor size.

a Gaussian distribution. Sigma (σ) is varied in order to evaluate the influence of this parameter. Figures 8, 9 and 10 show the average localization error (m) using the new proposed method. As before each figure corresponds with a specific lighting conditions of the test images condition (cloudy, night and sunny).

Better results are obtained using middle-high σ values. On Freiburg route and using HOG descriptor and this technique an error of 0.04598 m is obtained when σ=500 and cloudy environment, 0.1915 m when σ=500 and night environment and 0.7505 m when σ=250 and sunny environment. With Gist descriptor the errors obtained respectively were 0.04538 m 0.2040 m 1.5687 m all of them obtained with σ=250. On the other hand the lowest error in the INNOVA route are obtained using σ=750. The minimum error is 0.0619 m using HOG descriptor and 0.0637 m using Gist descriptor. These experiments clearly show

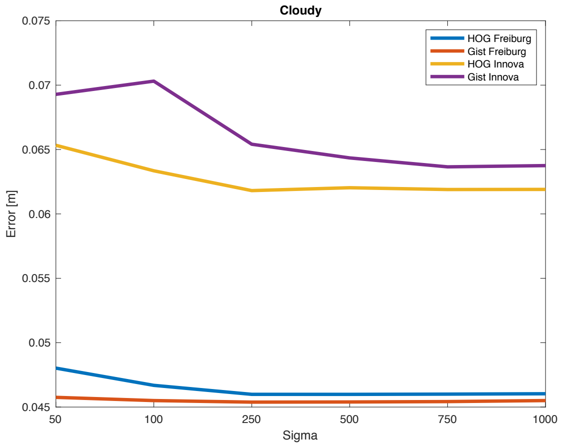

Figure 8: Localization error (m) using **cloudy** dataset versus sigma in the Gaussian distribution.

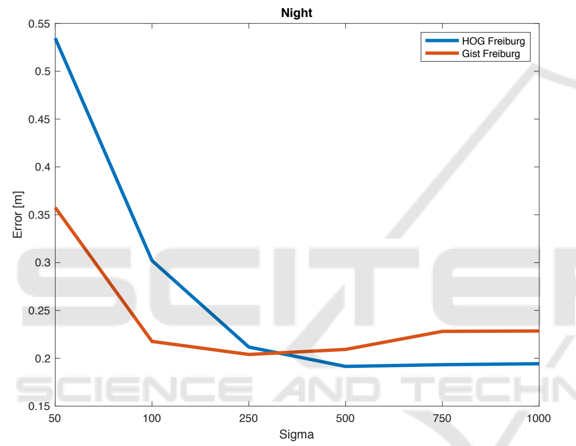

Figure 9: Localization error (m) using **night** dataset versus sigma in the Gaussian distribution.

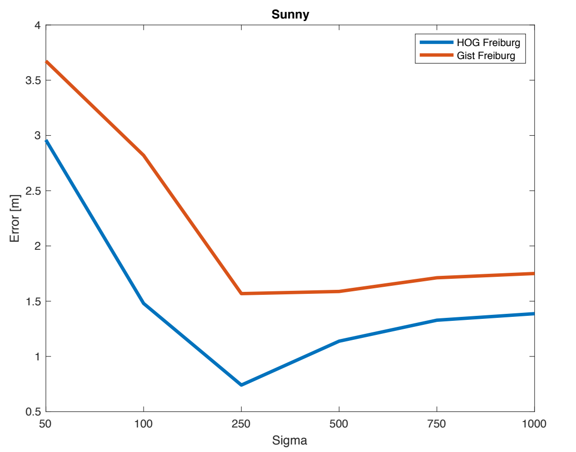

Figure 10: Localization error (m) using **sunny** dataset versus sigma in the Gaussian distribution.

that, lower errors are obtained using this new method where the descriptors are multiplied by a gaussian distribution. Best results are obtained with descriptor size 256 or 512 and taking into account the gaussian distribution, it is better to multiplied the image with a distribution σ=250 or σ=500. It can be observed on figures 11 and 12. They show the best result in each configuration and it is possible to observe that using the proposed method the error is lower. The improvement is specially important when the test images the images are taken from the sunny environment database. As seen, localisation error decreases using the proposed method, especially with HOG descriptor. The methods improve the task but to use it properly we have to supposed that the image is well distributed; there is the same quantity of cell and floor, they are not important on the scene and the camera is moved parallel to the ground floor. If these conditions are suit the proposed method should improved the localization task results.

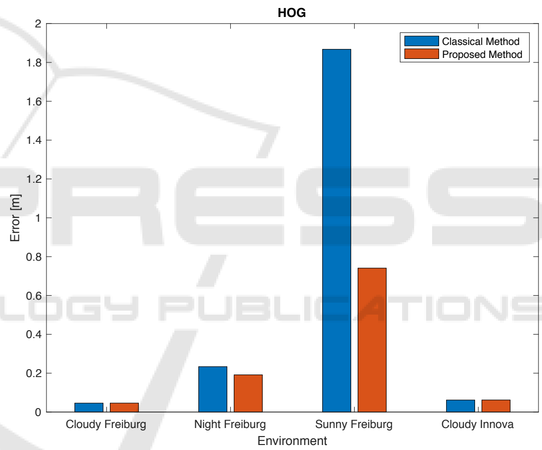

Figure 11: Comparison between classical and proposed methods vs environment while using **HOG** descriptor.

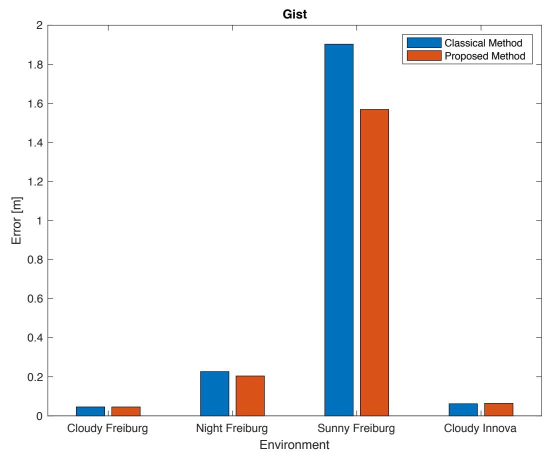

Figure 12: Comparison between classical and proposed methods vs environment while using **Gist** descriptor.

Table 3: Calculation time (s) of the localization process in the **classical method**.

Descriptor	Size	Time to build Descriptor (s)	Time Position (s)
HOG	64	0.0057	0.0836
	128	0.0070	0.0844
	256	0.0081	0.0858
	512	0.0106	0.0895
	1024	0.0132	0.1255
Gist	128	0.0295	0.0839
	256	0.0458	0.0854
	512	0.0643	0.0896
	1024	0.1041	0.1252
	2048	0.1478	0.1396

Table 4: Calculation time (s) of the localization process in the **new proposed method**.

Descriptor	Size	Time to build Descriptor (s)	Time Position (s)
HOG	64	0.0060	0.0836
	128	0.0070	0.0844
	256	0.0082	0.0858
	512	0.0106	0.0895
	1024	0.0134	0.1255
Gist	128	0.0297	0.0839
	256	0.0476	0.0854
	512	0.0692	0.0896
	1024	0.1054	0.1252
	2048	0.1478	0.1396

4.3 Computational Cost

A low error is an important characteristic to take into account when choosing a descriptor. But the computational cost is also an important issue to consider. For that reason, the necessary time to run the algorithms has also been studied. Table 3 using the classical method and table 4 using the proposed one show the time used to build the descriptor and to estimate the position. The data are given in seconds. As explained in section 2, the results depend on descriptors' size and this one depends on the parameters. Taking that into account, HOG descriptor size is $\vec{d} \in \mathbb{R}^{b \cdot k_1 \times 1}$, in the same way Gist descriptor size is $\vec{d} \in \mathbb{R}^{r \cdot m \cdot k_2 \times 1}$.

The experiments have been carried out with a CPU 8-Core Intel Xeon E5® at 3GHz and using the mathematical tool Matlab® . These time results are not absolute, they depend of the computer which runs the process. But they are comparable because all the calculations have been done with the same machine.

The lower the parameters are, the shorter the descriptor is and for that reason the runtime also is lower. If table 3 and table 4 are compared, it is possible to observe that the time used to build the descriptor is almost the same.

Finally, it is possible to observe that the process is quicker using HOG descriptor that when Gist is used, especially in the time used to build the descriptor.

5 CONCLUSIONS

The present work studies a new way to use global-appearance descriptors. The new method is compared with the traditional one in a localization task. The study has been made in real scenarios which are specially challenging due to lighting conditions and human activity. Using only visual information, global-appearance descriptors and the new proposed method described throughout this paper are studied. Once the images are described the performance of these descriptors in a localization framework is compared. Both, geometric localization error and the computational cost of the process have been studied and the parameters have been optimised.

First, about the traditional method, while Gist and HOG offer relatively good results in cloudy and night environments, the sunny conditions result more challenging, and HOG presents comparatively better results than Gist in this case. Second, about the proposed method, the experiments show that it presents substantial improvements, especially in sunny environment. About calculation times, the new method runs as quick as the traditional one. Observing the results the proposed method improves the localization task decreasing the error. Results obtained with HOG descriptor are better than the results using Gist, even though both minimise the localization error. As the run time is practically the same using both methods, using the proposed method where the image is multiplied by a gaussian vector may be a proper way to obtain better results in localization tasks where the base map has been built in an heterogeneous environment where changing lighting conditions and human activity can take part.

This work can be the first step to build more suitable description solutions in navigation tasks, specially when omnidirectional or panoramic images include a lot of information from the ceiling and/or the floor. The studied alternative may complement other classical description methods in order to achieve a robust localization. Multiply the descriptor by a gaussian distribution should not be considered as a unique solution but they can contribute towards obtaining results. Future works can include a study of a more robust solution, combining for example the new studied techniques with other measurements or techniques.

ACKNOWLEDGEMENTS

This work has been supported by the Generalitat Valenciana and the FSE through the grant ACIF/2018/224, by the Spanish government through the project DPI 2016-78361-R (AEI/FEDER, UE): “Creación de mapas mediante métodos de apariencia visual para la navegación de robots.” and by Generalitat Valenciana through the project AICO/2019/031: “Creación de modelos jerárquicos y localización robusta de robots móviles en entornos sociales”

REFERENCES

Amorós, F., Payá, L., Marín, J. M., and Reinoso, O. (2018). Trajectory estimation and optimization through loop closure detection, using omnidirectional imaging and global-appearance descriptors. *Expert Systems with Applications*, 102:273–290.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.

Berenguer, Y., Payá, L., Valiente, D., Peidró, A., and Reinoso, O. (2019). Relative altitude estimation using omnidirectional imaging and holistic descriptors. *Remote Sensing*, 11(3):323.

Cebollada, S., Payá, L., Valiente, D., Jiang, X., and Reinoso, O. (2019). An evaluation between global appearance descriptors based on analytic methods and deep learning techniques for localization in autonomous mobile robots.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1.

Gil, A., Valiente, D., Reinoso, Ó., Fernández, L., and Marín, J. M. (2011). Building visual maps with a single omnidirectional camera. In *ICINCO (2)*, pages 145–154.

Hofmeister, M., Liebsch, M., and Zell, A. (2009). Visual self-localization for small mobile robots with weighted gradient orientation histograms. In *40th International Symposium on Robotics (ISR)*, pages 87–91. Barcelona.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Menegatti, E., Maeda, T., and Ishiguro, H. (2004). Image-based memory for robot navigation using properties of omnidirectional images. *Robotics and Autonomous Systems*, 47(4):251 – 267.

Murillo, A. C., Guerrero, J. J., and Sagues, C. (2007). Surf features for efficient robot localization with omnidirectional images. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3901–3907. IEEE.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175.

Oliva, A. and Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, 155:23–36.

Payá, L., Fernández, L., Reinoso, Ó., Gil, A., and Úbeda, D. (2009). Appearance-based dense maps creation-comparison of compression techniques with panoramic images. In *ICINCO-RA*, pages 250–255.

Payá, L., Gil, A., and Reinoso, O. (2017). A state-of-the-art review on mapping and localization of mobile robots using omnidirectional vision sensors. *Journal of Sensors*, 2017.

Payá, L., Peidró, A., Amorós, F., Valiente, D., and Reinoso, O. (2018). Modeling environments hierarchically with omnidirectional imaging and global-appearance descriptors. *Remote Sensing*, 10(4):522.

Pronobis, A. and Caputo, B. (2009). COLD: COsy Localization Database. *The International Journal of Robotics Research (IJRR)*, 28(5):588–594.

Radon, J. (2005). 1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Classic papers in modern diagnostic radiology*, 5:21.

Román, V., Payá, L., Flores, M., Cebollada, S., and Reinoso, Ó. (2019). Performance of new global appearance description methods in localization of mobile robots. In *Iberian Robotics conference*, pages 351–363. Springer.

Román, V., Payá, L., and Reinoso, Ó. (2018). Evaluating the robustness of global appearance descriptors in a visual localization task, under changing lighting conditions. In *ICINCO-RA*, pages 258–265.

Siagian, C. and Itti, L. (2009). Biologically inspired mobile robot vision localization. *IEEE Transactions on Robotics*, 25(4):861–873.

Sturm, P., Ramalingam, S., Tardif, J.-P., Gasparini, S., Barreto, J., et al. (2011). Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends® in Computer Graphics and Vision*, 6(1–2):1–183.

Su, Z., Zhou, X., Cheng, T., Zhang, H., Xu, B., and Chen, W. (2017). Global localization of a mobile robot using lidar and visual features. In *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 2377–2383. IEEE.

Valiente García, D., Fernández Rojo, L., Gil Aparicio, A., Payá Castelló, L., and Reinoso García, O. (2012). Visual odometry through appearance-and feature-based method with omnidirectional images. *Journal of Robotics*, 2012.

Xu, S., Chou, W., and Dong, H. (2019). A robust indoor localization system integrating visual localization aided by cnn-based image retrieval with monte carlo localization. *Sensors*, 19(2):249.

Zhou, X., Su, Z., Huang, D., Zhang, H., Cheng, T., and Wu, J. (2018). Robust global localization by using global visual features and range finders data. In *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 218–223. IEEE.