

# ICINCO 2019

16th International Conference on  
Informatics in Control, Automation and Robotics

## PROCEEDINGS

Volume 2

Prague, Czech Republic

29-31 July, 2019

### EDITORS

Oleg Gusikhin  
Kurosh Madani  
Janan Zaytoon

<http://www.icinco.org/>

SPONSORED BY



PAPERS AVAILABLE AT



# ICINCO 2019

Proceedings of the  
16th International Conference on  
Informatics in Control, Automation and Robotics

Volume 2

Prague - Czech Republic

July 29 - 31, 2019

Sponsored by

**INSTICC - Institute for Systems and Technologies of Information, Control and Communication**

Technical Co-sponsorship by

**IFAC - International Federation of Automatic Control**  
**IEEE Robotics and Automation Society**

In Cooperation with

**AAAI - Association for the Advancement of Artificial Intelligence**

**RSJ - Robotics Society of Japan**

**APCA - Associação Portuguesa de Controlo Automático**

**INNS - International Neural Network Society**

**euRobotics AISBL**

**SPR - Portuguese Robotics Society**

Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda.  
All rights reserved

Edited by Oleg Gusikhin, Kurosh Madani and Janan Zaytoon

Printed in Portugal

ISSN: 2184-2809

ISBN: 978-989-758-380-3

Depósito Legal: 457918/19

<http://www.icinco.org>

[icinco.secretariat@insticc.org](mailto:icinco.secretariat@insticc.org)

# CONTENTS

---

## INVITED SPEAKERS

### KEYNOTE SPEAKERS

- Evolving Systems and Their Automotive Applications 5  
*Dimitar Filev*
- On Some Open-ended Challenges in Model-based Fault Management for Aerospace Systems: A Look Backwards and Forwards 7  
*Ali Zolghadri*
- The Science of Autonomy: A Holistic View at the Intersection of Learning, Control and Physics 18  
*Evangelos Theodorou*

## ROBOTICS AND AUTOMATION

### FULL PAPERS

- Camera and Lidar Cooperation for 3D Feature Extraction 23  
*Burtin Gabriel, Bonnin Patrick and Malartre Florent*
- Balancing Control of a Self-driving Bicycle 34  
*T. J. Yeh, Hao-Tien Lu and Po-Hsuan Tseng*
- Singularity Analysis for Redundant Manipulators of Arbitrary Kinematic Structure 42  
*Ahmad A. Almarkhi and Anthony A. Maciejewski*
- Automated Draping of Wide Textiles on Double Curved Surfaces 50  
*Patrick Kaumfann, Georg Braun, Andreas Buchheim and Marcin Malecha*
- Hybrid Force/Position Control of a Very Flexible Parallel Robot Manipulator in Contact with an Environment 59  
*Fatemeh Ansarieshlaghi and Peter Eberhard*
- Optimum Design and FEA of a Hybrid Parallel-deployable Structure-based 3-DOF Multi-gripper Translational Robot for Field Pot Seedlings Transplanting 68  
*Samy F. M. Assal and Isaac Ndawula*
- A Fuzzy Inference Approach to Control Robot Speed in Human-robot Shared Workspaces 78  
*Angelo Campomaggiore, Marco Costanzo, Gaetano Lettera and Ciro Natale*
- Environment-aware Sensor Fusion using Deep Learning 88  
*Caio Fischer Silva, Paulo V. K. Borges and José E. C. Castanho*
- Aerial Monitoring of Rice Crop Variables using an UAV Robotic System 97  
*C. Devia, J. Rojas, E. Petro, C. Martinez, I. Mondragon, D. Patino, C. Rebolledo and J. Colorado*
- A Generic Control Framework for Mobile Robots Edge Following 104  
*Mathieu Deremetz, Adrian Couvent, Roland Lenain, Benoît Thuilot and Christophe Cariou*
- Kinematics Modelling, Optimization and Control of Hybrid Robots 114  
*Mahmoud Tarokh and Federico Llenar*

|  |     |
|--|-----|
| Optimal Waypoint Navigation for Underactuated Cruising AUVS<br><i>Kangsoo Kim</i>  | 124 |
| Combining Ontologies and Behavior-based Control for Aware Navigation in Challenging Off-road Environments<br><i>Patrick Wolf, Thorsten Ropertz, Philipp Feldmann and Karsten Berns</i>   | 135 |
| Clustering-based Model for Predicting Multi-spatial Relations in Images<br><i>Brandon Birmingham and Adrian Muscat</i>   | 147 |
| <b>SHORT PAPERS</b>  |     |
| Path Planning of a Mobile Robot in Grid Space using Boundary Node Method<br><i>R. A. Saeed and Diego Reforgiato Recupero</i>   | 159 |
| Collision Detection for a Mobile Robot using Logistic Regression<br><i>Felix Becker and Marc Ebner</i>   | 167 |
| Integration of Open Source Arduino with LabVIEW-based SCADA through OPC for Application in Industry 4.0 and Smart Grid Scenarios<br><i>Isaías González Pérez, A. José Calderón Godoy and Manuel Calderón Godoy</i>   | 174 |
| Smart Wheelchairs: Using Robotics to Bridge the Gap between Prototypes and Cost-effective Set-ups<br><i>Matthew Aquilina, Marvin K. Bugeja and Simon G. Fabri</i>  | 181 |
| An Approach to Marker Detection in IR- and RGB-images for an Augmented Reality Marker<br><i>Aaronkumar Ehambram, Patrick Hemme and Bernardo Wagner</i>   | 190 |
| Adaptive Controller for Uncertain Multi-agent System Under Disturbances<br><i>Sergey Vlasov, Alexey Margun, Aleksandra Kirsanova and Polina Vakhvianova</i>  | 198 |
| A Comparison Study on Coupling Effects in Balance Control Methods of Humanoid Robots through an Extended Task Space Formulation<br><i>Seungjae Yoo, Joonhee Jo and Yonghwan Oh</i>   | 206 |
| Survey about the Utilization of Open Source Arduino for Control and Measurement Systems in Advanced Scenarios. Application to Smart Micro-Grid and Its Digital Replica<br><i>Isaías González Pérez, A. José Calderón Godoy, Manuel Calderón Godoy and J. Félix González González</i> | 214 |
| Robust Finite-time Position and Attitude Tracking of a Quadrotor UAV using Super-Twisting Control Algorithm with Linear Correction Terms<br><i>Yassine Kali, Jorge Rodas, Maarouf Saad, Raul Gregor, Walid Alqaisi and Khalid Benjelloun</i>   | 221 |
| Smoothing and Time Parametrization of Motion Trajectories for Industrial Machining and Motion Control<br><i>Květoslav Belda</i>  | 229 |
| Reliability Analysis of the Kalman Filter for INS/GPS Integrated Navigation System Applied to Train<br><i>Seong Yun Cho, Chang Ho Kang and Kyung Ho Shin</i>   | 237 |
| Towards an Advanced ROS Package Generator<br><i>Anthony Remazeilles and Jon Azpiazu</i>  | 243 |
| Disturbance Observer for Path-following Control of Autonomous Agricultural Vehicles<br><i>T. Hiramatsu, M. Pencelli, S. Morita, M. Niccolini, M. Ragaglia and A. Argiolas</i>  | 251 |

|   |     |
|---|-----|
| A Novel Approach for a Leg-based Stair-climbing Wheelchair based on Electrical Linear Actuators<br><i>Emiliano Pereira, Hilario Gómez-Moreno, Cristina Alén-Cordero, Pedro Gil-Jiménez and Saturnino Maldonado-Bascón</i>                     | 259 |
| Stereo Vision-based Autonomous Target Detection and Tracking on an Omnidirectional Mobile Robot<br><i>Wei Luo, Zhefei Xiao, Henrik Ebel and Peter Eberhard</i>  | 268 |
| Theoretical and Experimental Modal Analysis of a 6 PUS PKM<br><i>Francesco La Mura, Hermes Giberti, Linda Pirovano and Marco Tarabini</i>   | 276 |
| An Evaluation between Global Appearance Descriptors based on Analytic Methods and Deep Learning Techniques for Localization in Autonomous Mobile Robots<br><i>Sergio Cebollada, Luis Payá, David Valiente, Xiaoyi Jiang and Oscar Reinoso</i> | 284 |
| Time Synchronisation of Low-cost Camera Images with IMU Data based on Similar Motion<br><i>Peter Aerts and Eric Demeester</i>   | 292 |
| Quasi-serial Manipulator for Advanced Manufacturing Systems<br><i>Bryan Kelly, J. Padayachee and G. Bright</i>  | 300 |
| Integration of an Autonomous System with Human-in-the-Loop for Grasping an Unreachable Object in the Domestic Environment<br><i>Jaeseok Kim, Raffaele Limosani and Filippo Cavallo</i>  | 306 |
| Small Radius Spheres in Output Space of Nonholonomic Systems<br><i>Arkadiusz Mielczarek and Ignacy Duleba</i>   | 316 |
| Progression of Human Hand Trajectory Variabilities during a Pick-and-Place Task<br><i>Kolja Kühnlenz, Sergej Hermann, Kevin Kalb, Lucas Marscholke and Barbara Kühnlenz</i>   | 323 |
| Geometric Adaptive Robust Sliding-mode Control on SO(3)<br><i>Yulin Wang, Xiao Wang, Shengjing Tang and Jie Guo</i>   | 328 |
| Control of an Industrial Dual-arm Robot in a Narrow Space where Human Workers are Familiar with<br><i>Taeyong Choi, Hyunmin Do, Donil Park and Jinho Kyungk</i>   | 339 |
| Technology of Developing the Software for Robots Vision Systems<br><i>S. M. Sokolov, A. A. Boguslavsky and N. D. Beklemichev</i>  | 345 |
| Nonlinear Output Feedback for Autonomous U-turn Maneuvers of a Robot in Orchard Headlands<br><i>E. Le Flécher, A. Durand-Petiteville, F. Gouaisbaut, V. Cadenat, T. Sentenac and S. Vougioukas</i>  | 355 |
| Towards Skills-based Easy Programming of Dual-arm Robot Applications<br><i>Fan Dai</i>  | 363 |
| Human-aware Robot Navigation in Logistics Warehouses<br><i>Mourad A. Kenk, M. Hassaballah and Jean-François Brethé</i>  | 371 |
| Development and Implementation of Grasp Algorithm for Humanoid Robot AR-601M<br><i>Kamil Khusnutdinov, Artur Sagitov, Ayrat Yakupov, Roman Meshcheryakov, Kuo-Hsien Hsia, Edgar A. Martinez-Garcia and Evgeni Magid</i>                       | 379 |
| State Observers for Mechatronics Systems with Rigid and Flexible Drive Dynamics<br><i>Alexandra-Iulia Szedlak-Stinean, Radu-Emil Precup and Radu-Codrut David</i>   | 387 |

|   |     |
|---|-----|
| A Generalized Odometry for Implementation of Simultaneous Localization and Mapping for Mobile Robots<br><i>Kethavath Raj Kumar</i>  | 395 |
| An Improved APFM for Autonomous Navigation and Obstacle Avoidance of USVs<br><i>Xiaohui Zhu, Yong Yue, Hao Ding, Shunda Wu, MingSheng Li and Yawei Hu</i>   | 401 |
| Towards Total Coverage in Autonomous Exploration for UGV in 2.5D Dense Clutter Environment<br><i>Evgeni Denisov, Artur Sagitov, Konstantin Yakovlev, Kuo-Lan Su, Mikhail Svinin and Evgeni Magid</i>  | 409 |
| A Modular Underactuated Gripper with Force Control System<br><i>A. Margun, D. Bazylev, K. Zimenko and A. Kremlev</i>  | 417 |
| Design, Estimation of Model Parameters, and Dynamical Study of a Hybrid Aerial-underwater Robot: <i>Acutus</i><br><i>Ridhi Puppala, Nikhil Sivadasan, Abhijeet Vyas, Akshay Molawade, Thiyagarajan Ranganathan and Asokan Thondiyath</i>  | 423 |
| Multiple DOF Platform with Multiple Air Jets<br><i>Shinya Kotani, Nobukado Abe, Satoshi Iwaki, Tetsushi Ikeda and Takeshi Takaki</i>  | 431 |
| Development of an Experimental Strawberry Harvesting Robotic System<br><i>Dimitrios S. Klaoudatos, Vassilis C. Moulianitis and N. A. Aspragathos</i>  | 437 |
| Point-cloud Mapping using Lidar Mounted on Two-wheeled Vehicle based on NDT Scan Matching<br><i>Kohei Tokorodani, Masafumi Hashimoto, Yusuke Aihara and Kazuhiko Takahashi</i>  | 446 |
| Moving-object Tracking with Lidar Mounted on Two-wheeled Vehicle<br><i>Shotaro Muro, Yohei Matsui, Masafumi Hashimoto and Kazuhiko Takahashi</i>  | 453 |
| Development of Flow Rate Feedback Control in Tilting-ladle-type Pouring Robot with Direct Manipulation of Pouring Flow Rate<br><i>Yuta Sueki and Yoshiyuki Noda</i>   | 460 |
| A Testing-environment for a Mobile Collaborative Stereo Configuration with a Dynamic Baseline<br><i>Andreas Sutorma, Matthias Domnik and Jörg Thiem</i>   | 468 |
| A Novel Aerial Manipulation Design, Modelling and Control for Geometric CoM Compensation<br><i>Kamel Bouzgou, Laredj Benchikh, Lydie Nouveliere, Yasmina Bestaoui and Zoubir Ahmed-Foitih</i>   | 475 |
| Miniature Autonomy as One Important Testing Means in the Development of Machine Learning Methods for Autonomous Driving: How ML-based Autonomous Driving could be Realized on a 1:87 Scale<br><i>Tim Tiedemann, Jonas Fuhrmann, Sebastian Paulsen, Thorben Schnirpel, Nils Schönherr, Bettina Buth and Stephan Pareigis</i> | 483 |
| Sitting Assistance that Considers User Posture Tolerance<br><i>Daisuke Chugo, Masayu Koyama, Masahiro Yokota, Shohei Kawazoe, Satoshi Muramatsu, Sho Yokota, Hiroshi Hashimoto, Takahiro Katayama, Yasuhide Mizuta and Atsushi Koujina</i>  | 489 |
| A Supervised Autonomous Approach for Robot Intervention with Children with Autism Spectrum Disorder<br><i>Vinicius Silva, Sandra Queirós, Filomena Soares, João Sena Esteves and Demétrio Matos</i>   | 497 |
| Active Lower Limb Orthosis with One Degree of Freedom for Paraplegia<br><i>Takuhiro Sunada, Goro Obinata and Yanling Pei</i>  | 504 |

|   |     |
|---|-----|
| Computed Torque Control of an Aerial Manipulation System with a Quadrotor and a 2-DOF Robotic Arm<br><i>Nebi Bulut, Ali Emre Turgut and Kutluk Bilge Arıkan</i>                               | 510 |
| Investigation of Non-circular Scanning Trajectories in Robot-based Industrial X-ray Computed Tomography of Multi-material Objects<br><i>Peter Landstorfer, Gabriel Herl and Jochen Hiller</i> | 518 |
| Lean Human-Robot Interaction Design for the Material Supply Process<br><i>Marco Bonini, Augusto Urru and Wolfgang Echelmeyer</i>  | 523 |
| The Efficient Distribution Method of Limited Wireless Communication Frequency Resources for the Multi-robot Teaming<br><i>Heeseo Chae, Jae Hyuk Ju and Jae Hyun Park</i>                      | 530 |
| FPGA-based Embedded System Designed for the Deployment in the Compliant Robotic Leg CARL<br><i>Steffen Schütz, Atabak Nejadfard, Max Reichardt and Karsten Berns</i>                          | 537 |
| Vision-based Localization of a Wheeled Mobile Robot with a Stereo Camera on a Pan-tilt Unit<br><i>A. Zdešar, G. Klančar and I. Škrjanc</i>  | 544 |
| Autonomous Gripping and Carrying of Polyhedral Shaped Object based on Plane Detection by a Quadruped Tracked Mobile Robot<br><i>Toyomi Fujita and Nobuatsu Aimi</i>                           | 552 |
| A Terminal Sliding Mode Control using EMG Signal: Application to an Exoskeleton-Upper Limb System<br><i>Sana Bembli, Nahla Khraief Haddad and Safya Belghith</i>                              | 559 |
| MSER-based Framework for Classification of Objects in Thermal Images<br><i>Alia Aljasmı and Andrzej Śluzek</i>  | 566 |
| Locomotion Mode Selection Plus (LMS+) Algorithm for Resource Efficient Outdoor Navigation<br><i>Amir Sharif and Hubert Roth</i>   | 573 |
| <b>INDUSTRIAL INFORMATICS</b>   |     |
| <b>FULL PAPERS</b>  |     |
| An IoT Framework for Assembly Tracking and Scheduling in Manufacturing SME<br><i>Meysam Minoufekar, Anass Driate and Peter Plapper</i>  | 585 |
| Modular and Domain-guided Multi-robot Planning for Assembly Processes<br><i>Ludwig Nägele, Andreas Schierl, Alwin Hoffmann and Wolfgang Reif</i>  | 595 |

## SHORT PAPERS

|  |     |
|--|-----|
| The Effect of Baffles on Heat Transfer<br><i>Raheleh Jafari, Sina Razvarz, Cristóbal Vargas-Jarillo and Alexander Gegov</i>  | 607 |
| An Innovative Automated Robotic System based on Deep Learning Approach for Recycling Objects<br><i>Jaeseok Kim, Olivia Nocentini, Marco Scafuro, Raffaele Limosani, Alessandro Manzi, Paolo Dario and Filippo Cavallo</i>          | 613 |
| Interdisciplinary Approach to Cyber-physical Systems Training<br><i>Radda A. Iureva, Artem S. Kremlev, Alexey A. Margun, Sergey M. Vlasov, Sergey D. Vasilkov, Alexandr V. Penskoi, Dmitry E. Konovalov and Pavel Y. Korepanov</i> | 623 |
| Modelling of CNC Machine Tools for Augmented Reality Assistance Applications using Microsoft HoloLens<br><i>Meysam Minoufekar, Pascal Schug, Pascal Zenker and Peter Plapper</i>   | 627 |
| Formalizing the Safety Functions to Assure the Software Quality of NPP Safety Important Systems<br><i>Elena Ph. Jharko</i>   | 637 |
| I-AM: Interface for Additive Manufacturing<br><i>Marco Rodrigues, João Paulo Pereira and Pedro Miguel Moreira</i>  | 645 |
| Deep Neural Networks for New Product Form Design<br><i>Chun-Chun Wei, Chung-Hsing Yeh, Ian Wang, Bernie Walsh and Yang-Cheng Lin</i>   | 653 |
| Inventory Routing Problem with Non-stationary Stochastic Demands<br><i>Ehsan Yadollahi, El-Houssaine Aghezzaf, Joris Walraevens and Birger Raa</i>   | 658 |
| Software 2.0 for Scrap Metal Classification<br><i>Manuel Robalinho and Pedro Fernandes</i>   | 666 |
| AUTHOR INDEX   | 675 |

# An Evaluation between Global Appearance Descriptors based on Analytic Methods and Deep Learning Techniques for Localization in Autonomous Mobile Robots

Sergio Cebollada<sup>1</sup>, Luis Payá<sup>1</sup>, David Valiente<sup>1</sup>, Xiaoyi Jiang<sup>2</sup> and Oscar Reinoso<sup>1</sup>

<sup>1</sup>*Department of Systems Engineering and Automation, Miguel Hernández University, Elche, 03202, Spain*

<sup>2</sup>*Department of Computer Science, University of Münster, Münster, 48149, Germany*  
{sergio.cebollada, lpaya, dvaliente, o.reinoso}@umh.es, xjiang@uni-muenster.de

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

**Abstract:** In this work, different global appearance descriptors are evaluated to carry out the localization task, which is a crucial skill for autonomous mobile robots. The unique information source used to solve this issue is an omnidirectional camera. Afterwards, the images captured are processed to obtain global appearance descriptors. The position of the robots is estimated by comparing the descriptors contained in the visual model and the descriptor calculated for the test image. The descriptors evaluated are based on (1) analytic methods (HOG and *gist*) and (2) deep learning techniques (auto-encoders and Convolutional Neural Networks). The localization is tested with a panoramic dataset which provides indoor environments under real operating conditions. The results show that deep learning based descriptors can be also an interesting solution to carry out visual localization tasks.

## 1 INTRODUCTION

Nowadays, the use of visual information to solve mobile autonomous robotic tasks is widely expanded. In these cases, the robot must be able to build a map within the environment and estimate its position within that environment. These tasks are known as mapping and localization. Among the different sensors used, the omnidirectional cameras introduce an interesting solution since they are able to provide information that covers a field of view of 360 deg.

Global appearance descriptors have been proposed by several authors to extract characteristic information from images and use this information for mapping and localization. For instance Zhou *et al.* (Zhou et al., 2018) propose the use of the descriptor *gist* to solve the localization through matching the best keyframe in the dataset based on the given robot's current view. Korrapati and Mezouar (Korrapati and Mezouar, 2017) introduced the use of omnidirectional images through global appearance descriptors to create a topological mapping approach and a loop closure detection method. More recently, Román *et al.* (Román et al., 2018) evaluate the use of global appearance descriptors for localization under illumination changes. In this work, several distance measurements were also evaluated with the aim to obtain a

similitude distance between images which represents the geometrical distance between the positions where those images were captured.

The computation of these descriptors is based on analytic methods, nevertheless, during the last years, some authors have proposed the use of deep learning techniques to create global appearance descriptors. For example, on the one hand, Xu *et al.* (Xu et al., 2016) proposed the use of auto-encoders to detect histopathological images of breast cancer. On the other hand, Xu *et al.* (Xu et al., 2019) used a CNN-based descriptor to obtain the most probable position within an indoor map through Monte Carlo Localization and also to solve the kidnapping problem; Payá *et al.* (Payá et al., 2018) proposed also the use of the CNN-based descriptors but in this case for hierarchical mapping. Both works are based on the net *places* (Zhou et al., 2014). The descriptors extracted from this network correspond to the ones calculated in some of the fully convolutional layers within the network.

Through this work, we carry out a comparison between global appearance descriptors based on analytic methods and global appearance descriptors based on deep learning techniques to solve the visual localization task. The goodness of these methods are measured according to the accuracy (error of localiza-

tion) and computing time (to calculate the descriptor and to estimate the position of the robot).

The remainder of the paper is structured as follows: Section 2 explains the algorithm used to estimate the position of the test images within the environment. After that, section 3 outlines the global appearance descriptors which will be evaluated. Section 4 explains and presents the dataset used as well as the experimental results and the discussion about them. Finally, section 5 outlines the conclusions and future research lines.

## 2 LOCALIZATION METHOD

The localization task consists in an image retrieval problem. This is, obtaining the image which presents higher similitude in relation to the new captured image. For this purpose, the robot has previously obtained visual information from the environment, i.e., global appearance descriptors that are calculated from the  $N_{Train}$  images captured from different positions of the environment. This task is known as mapping and this step must be carried out before starting the localization. Therefore, the localization task is solved through the following steps:

- The robot captures a new omnidirectional image from an unknown position.
- That image is transformed to panoramic ( $im_{test}$ ) and after that, the corresponding descriptor is calculated ( $\vec{d}_{test}$ ).
- Once the descriptor is available, the robot calculates the cosine distance (selected in (Cebollada et al., 2019) as the best distance method for global appearance descriptors) between the test descriptor ( $\vec{d}_{test}$ ) and each descriptor from the visual model ( $\vec{d}_j$ , where  $j = 1, \dots, N_{Train}$ ).
- A vector of distances is obtained as  $\vec{h}_t = \{h_{t1}, \dots, h_{tN_{Train}}\}$  where  $h_{tj} = dist\{\vec{d}_{test}, \vec{d}_j\}$ .
- The node which presents the minimum distance ( $d_t^m | t = argmin_j h_{tj}$ ) corresponds to the estimated position of the robot.

## 3 THE GLOBAL APPEARANCE DESCRIPTORS

Visual localization has been commonly solved either using local features along a set of scenes or using a unique descriptor per image which contains information on its global appearance. These second methods

are known as global appearance description and have been used to solve the localization task since they allow straightforward localization algorithms. For instance, Naseer *et al.* (Naseer et al., 2018) propose a localization method from global appearance (by using histogram of oriented gradients descriptors and features from deep convolutional neural networks) to solve the localization problem and to keep in parallel several possible trajectories hypotheses.

Basically, the steps to calculate a global appearance descriptor are the following: (1) The starting point is a panoramic image expressed as a bidirectional matrix ( $im_j \in \mathbb{R}^{M_x \times M_y}$ ). (2) Then the specific mathematical calculations are applied and a vector which characterizes the original image will be obtained ( $\vec{d}_j \in \mathbb{R}^{l \times 1}$  and corresponds to the image  $im_j$ ).

The first global appearance descriptors used in computer vision were descriptors based on analytic methods. Nevertheless, during the last years, the emergence of the deep learning techniques have empowered the use of descriptors based on these new methods.

### 3.1 Methods based on Analytic Methods

These methods are basically based on calculations of gradients and orientation of the different pixels which compose the image. Their use has been quite often to solve mobile robotics issues. For instance, Su *et al.* (Su et al., 2017) used a global descriptor to reduce pose search space with the aim to solve the kidnapped robot problem in indoor environments under different conditions. An interesting study was carried out by Román *et al.* (Román et al., 2018) which evaluates the use of global appearance descriptors for localization under illumination changes. More recently, Cebollada *et al.* (Cebollada et al., 2019) evaluate the use of global appearance descriptors to build hierarchical maps through clustering algorithms and then to solve the localization in those maps.

Among the different methods, this work proposes the use of HOG and gist, which have been used in previous works (Cebollada et al., 2019) and have proved to present interesting results for localization tasks.

Regarding the **HOG** descriptor, it was introduced by Dalal and Triggs (Dalal and Triggs, 2005) to solve the detection of pedestrians. In this work, the procedure is the one proposed by Leonardis and Bischof (Leonardis and Bischof, 2000): the panoramic image is divided into  $k_1$  horizontal cells and a histogram of gradient orientation (with  $b$  bins per histogram) is compiled per each cell. Finally, the set of histograms are arranged in a unique row to compose the final descriptor  $\vec{d} \in \mathbb{R}^{b \cdot k_1 \times 1}$

As for the *gist* descriptor, it was introduced by Oliva *et al.* (Oliva and Torralba, 2006). In this work, the version used consists on: (1) obtaining  $m_2$  different resolution images, (2) applying Gabor filters over the  $m_2$  images with  $m_1$  different orientations, (3) grouping the pixels of each image into  $k_2$  horizontal blocks and (4) arranging the obtained orientation information into one row to create a vector  $\vec{d} \in \mathbb{R}^{m_1 \cdot m_2 \cdot k_2 \times 1}$ .

### 3.2 Methods based on Deep Learning

During the last years, the use of deep learning methods to solve computer vision issues has extensively grown. Regarding the localization task through the use of visual information, this work studies the use of Convolutional Neural Networks (CNN) and the use of auto-encoders. The idea is to obtain vectors which characterize the images through some deep learning technique. On the one hand, these methods can result very interesting since their use can be focused on specific kind of images (such as indoor environments in our case) and, hence, providing more efficient descriptors. On the other hand, these methods lead to previous training which normally implies huge processing data and noteworthy time.

Regarding the use of CNNs, these networks have been commonly designed for classification. In this sense, (1) a set of images correctly labeled are collected and introduced into the network to tackle the learning process and after that, (2) the network is properly available to face the classification (test image as input and the CNN outputs the most likely label option). The CNNs are composed by several hidden layers whose parameters and weights are tuned through the training iterations. In this work, some hidden layers outputs are used to obtain global appearance descriptors. This idea have already been proposed by some authors such as Mancini *et al.* (Mancini *et al.*, 2017), who use them to carry out place categorization with the Naïve Bayes classifier or Payá *et al.* (Payá *et al.*, 2018), who proposed CNN-based descriptors to create hierarchical visual models for mobile robot localization. The CNN architecture that has been used in this work is *places* (Zhou *et al.*, 2014), which was trained with around 2.5 million images to categorize 205 possible kinds of scenes (no re-training is carried out in this work). Fig. 1 shows the architecture of the *places* CNN, which is based on the caffe CNN. The net basically consists in (1) an input layer, (2) several intermediate hidden layers and (3) an output layer. Within the intermediate layers, the first phase consists in (2.1) layers for featurizing learning (whose layers incorporate several filters and the output gen-

erated are used as input for the next layer) and (2.2) layers for classification (whose layers are fully connected and they generate vectors which provide information for classification).

In this work, we have evaluated the output information from 5 layers. Three fully convolutional layers ('*fc6*', '*fc7*' and '*fc8*') whose output size are  $4096 \times 1$ ,  $4096 \times 1$  and  $205 \times 1$  respectively. Moreover, we have obtained two descriptors from the output of 2D convolution layers ('*conv4*' and '*conv5*'). These layers apply several sliding convolutional filters to the input images with the aim to activate certain characteristics of the image. Hence, the output of these layers is a set of images which are the input image after being filtered. Finally, a descriptor is basically obtained from these layers through selecting an image from the output dataset and arranging the data (matrix) in a single row (vector). Since the size of the output images is  $13 \times 13$ , the size of the descriptor is  $169 \times 1$ .

As for the use of **auto-encoders**, the aim of these neural networks is to reconstruct the output through compressing the input into a latent-space representation (Hubens, 2018). The fig. 2 shows the architecture design of the auto-encoders. These networks firstly compress the input (encoding) and secondly reconstruct the input departing from the latent space representation (decoding). The idea consists in building a latent representation to obtain useful features with small dimension, i.e., training the auto-encoder to extract the most salient features. For example, Gao and Zhang (Gao and Zhang, 2017) used auto-encoders to detect loops for visual Simultaneous Localization And Mapping (SLAM).

For this experiment, two types of auto-encoder are proposed. Both have been trained using the same parameters (Coefficient for the  $L_2$  weight regularizer, 0,004; Coefficient that controls the impact of the sparsity regularizer, 4; Desired proportion of training examples a neuron reacts to, 0.15; Encoder Transfer Function, "Logistic sigmoid function"; and Maximum number of training epochs, 1000) and also both have been trained using a GPU (NVIDIA GeForce GTX 1080 Ti), but whereas the first option (*auto-enc-Frib*) is trained with the images obtained from the dataset used to evaluate the localization (explained in sec. 4), the second alternative (*auto-enc-SUN*) is trained with images obtained from a dataset (SUN 360 DB (Xiao *et al.*, 2012)) which contains generic panoramic images. The aim of this second option is to create a generic auto-encoder based on indoor panoramic images which provides a good-enough solution to obtain descriptors for panoramic images independently the environment. This solution would solve the handicap that introduces the descrip-

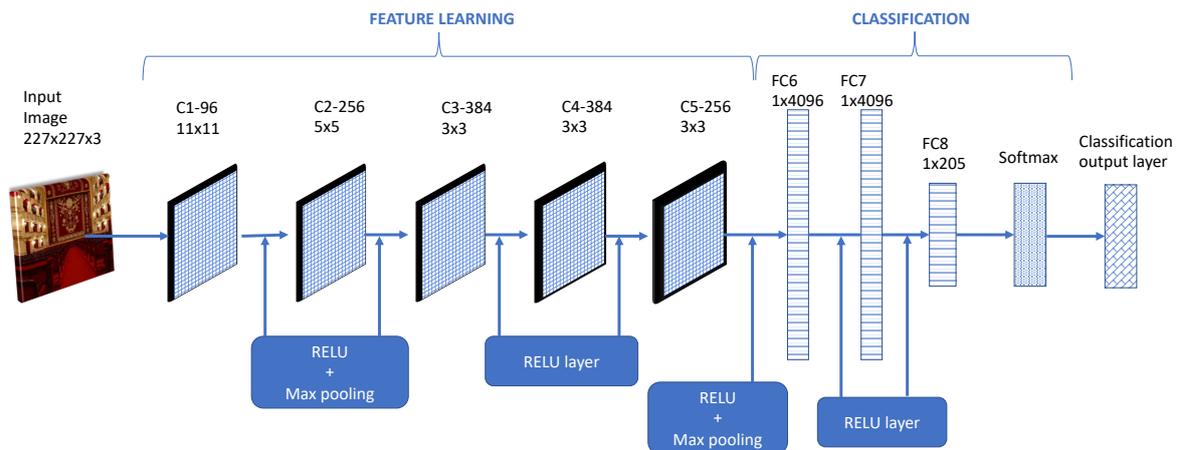


Figure 1: CNN architecture design of the pre-trained 'caffe' model.

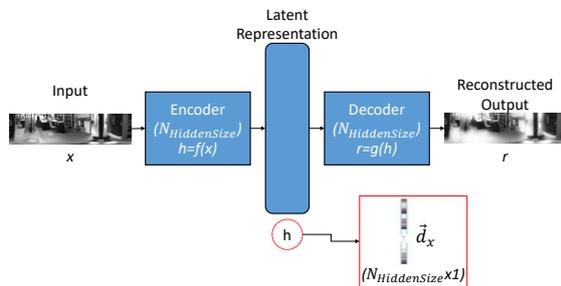


Figure 2: Auto-encoder architecture design and extraction of features departing from the latent representation.

tor based on auto-encoders regarding the need to carry out a previous training before calculating the descriptors. For *auto-enc-Frib*, the training dataset consists in 519 panoramic images whose size is  $512 \times 128$ ; for *auto-enc-SUN*, the training dataset consists in 541 panoramic images whose size is also  $512 \times 128$ . Furthermore, the auto-encoders are trained varying the size of hidden representation of the auto-encoder (this number is the number of neurons in the hidden layer) and the resultant descriptor size obtained depends directly on that number ( $N_{HiddenSize} \times 1$ ). Regarding the computing time to train these auto-encoders, the computer needs between 6 min and 2,94 hours, being directly proportional to the number of neurons (the more number of neurons there are, the more computing time is required).

## 4 EXPERIMENTS

### 4.1 Dataset

The experiments were carried out through the use of the COLD dataset (Pronobis and Caputo, 2009),

which contains visual information along a trajectory. It contains three indoor laboratory environments in three cities (Freiburg, Saarbrücken and Ljubljana) and three different illumination conditions. Nevertheless, for the experiments purposes, only the images related to the Freiburg environment were used and no illumination changes have been considered, i.e., the images used were only captured under cloudy conditions (during the light hours but the sunlight does not considerably affect the shots). This lack of illumination changes is due to the fact that this work is focused on studying the goodness of the descriptors for localization task, however, in future works, an extension to study the illumination changes effects will be considered. This dataset includes changes in the environment such as people walking or position of furniture and objects. An example of these dynamic conditions can be seen in fig. 3.



Figure 3: Panoramic image from COLD database.

Among the different paths, the red one was selected for this experiment because it is the longest. Afterwards, the images are split into two datasets: training and test datasets. Training dataset is composed by 519 images which present an average distance around 20 cm between an image and the following one. The test dataset is composed by 2595 images and the average distance between images is 4,10 cm. The table 1 shows the information about the datasets in detail.

Table 1: Number of images in each room of the training and test datasets created from the Freiburg environment.

| Name                | Number of images in Training | Number of images in Test |
|---------------------|------------------------------|--------------------------|
| Printer area        | 44                           | 223                      |
| Corridor            | 212                          | 1044                     |
| Kitchen             | 51                           | 255                      |
| Large Office        | 34                           | 175                      |
| 2-persons office 1  | 46                           | 232                      |
| 2-persons office 2  | 26                           | 131                      |
| 1-person office     | 31                           | 154                      |
| Bathroom            | 49                           | 247                      |
| Stairs area         | 26                           | 134                      |
| <b>Total number</b> | 519                          | 2595                     |

## 4.2 Evaluation of the Localization

To evaluate the goodness of each descriptor method for localization, two parameters are considered: On the one hand, (1) the average localization error, which measures the Euclidean distance between the position estimated and the real position where the test image was captured. To obtain this value, the ground truth provided by the dataset is used. Nevertheless, the ground truth is only used for this purpose (it is not to solve the localization task). On the other hand, (2) the average computing time, which is analyzed through two values, (2.a) the computing time to calculate the descriptor and (2.b) the computing time to estimate the position of the test image.

The results obtained through the use of analytic descriptors (HOG and *gist*) and the descriptors based on deep learning (auto-encoders and CNNs) are shown in the tables 2, 3 and 4. These tables show the size of the descriptor, the average localization error (cm), the average computing time to calculate the descriptor (ms) and the average computing time to estimate the position of the test images (ms).

Regarding the results obtained through the use of descriptors based on analytic methods (see table 2), for the HOG case, the localization error does not significantly decrease as the size increases; the computing time to calculate the descriptor is also barely constant but the time to estimate the pose increases as the size of the descriptor does. Hence, the descriptor whose size is 64 is considered the best option, because this configuration presents good accuracy and the minimum computing time. Regarding the *gist* descriptor, the localization error decreases millimetres as the size of the descriptor increases, however the time to calculate the descriptors as well as the time to estimate the pose increases significantly as the size

does. Therefore, in this case, the minimum size is selected as the best option.

As for the descriptors obtained through the use of auto-encoders (see table 3), for both cases (auto-enc-Frib and auto-enc-SUN), the outputs obtained by using the auto-encoders whose size of hidden representation (number of neurons) is 10 show the worst localization error results. In the case of auto-enc-Frib, the descriptors obtained from auto-encoders with  $N_{HiddenSize} = 50 - 500$  behaves well (localization error between 7,04 and 7,45 cm), but for the auto-enc-SUN, only the case  $N_{HiddenSize} = 500$  outputs similar values. Regarding the computing times (to compute the descriptor and to estimate the pose), the longer the size of the descriptor is, the more time the method needs. Furthermore, the computing time values increase severally as the size does. For instance, in the case of  $N_{HiddenSize} = 500$ , with auto-enc-Frib and auto-enco-SUN, the average time are 1166 ms and 1125 ms respectively. Therefore, for auto-enc-Frib, the best configuration is reached through the auto-encoder whose number of neurons is 100, because the localization error is the minimum and the computing time is the third lowest. For auto-enc-SUN, despite the configuration with  $N_{HiddenSize} = 500$  presents the worst times, it is selected as the best one because the rest of options do not provide solutions that can be used to solve the localization task.

Finally, for the CNN-based descriptors case (see table 4), in general, all the layers evaluated present good results. The first layers achieve an accuracy of around 5 cm. This behaviour is reasonable since the aim of the first layers in a CNN is to obtain global characteristic information from the images and the further CNN layers are focused on optimizing the classification task. Special consideration for the layers '*conv4*' and '*conv5*', whose use to obtain global appearance descriptors is scarce until today and they present very optimal solutions. Regarding the computation time to calculate the descriptor, none of the layers need high values and, as it was expected, the further the corresponding layer is, the higher the time is. Moreover, the computing time to estimate the pose is directly proportional to the size of the descriptor, but the layer '*conv5*' needs less time than '*conv4*'. Hence, '*conv5*' is selected as the best layer to calculate descriptors.

## 5 CONCLUSIONS

In this work, a study is tackled regarding the use of global appearance descriptors for localization. This task is solved as an image retrieval problem. A dy-

Table 2: Results obtained through the use of global appearance descriptors based on analytic methods (HOG and *gist*) to solve visual localization.

| Descriptor  | Size       | Error loc. (cm) | Time comp. descriptor (ms) | Time pose est. (ms) |
|-------------|------------|-----------------|----------------------------|---------------------|
| HOG         | <b>64</b>  | 16,34 ± 0,78    | 44,64                      | 0,38                |
|             | 128        | 16,23 ± 0,73    | 45,27                      | 0,51                |
|             | 256        | 16,22 ± 0,69    | 45,33                      | 2,48                |
|             | 512        | 16,17 ± 0,69    | 46,52                      | 4,75                |
| <i>gist</i> | <b>128</b> | 5,19 ± 0,18     | 10,30                      | 0,45                |
|             | 256        | 5,11 ± 0,17     | 11,98                      | 2,19                |
|             | 512        | 5,09 ± 0,16     | 21,21                      | 4,17                |
|             | 1024       | 5,08 ± 0,16     | 40,07                      | 10,72               |

Table 3: Results obtained through the use of global appearance descriptors based on auto-encoders (auto-enc-Frib and auto-enc-SUN) to solve visual localization.

| Descriptor    | Size       | Error loc. (cm) | Time comp. descriptor (ms) | Time pose est. (ms) |
|---------------|------------|-----------------|----------------------------|---------------------|
| auto-enc-Frib | 10         | 599,83 ± 3,83   | 49,79                      | 0,25                |
|               | 50         | 8,61 ± 2,29     | 138,64                     | 0,44                |
|               | <b>100</b> | 7,04 ± 0,85     | 249,55                     | 0,59                |
|               | 200        | 7,45 ± 0,23     | 473,59                     | 0,93                |
|               | 500        | 7,22 ± 0,19     | 1166,49                    | 4,54                |
| auto-enc-SUN  | 10         | 362,73 ± 22,77  | 54,99                      | 0,28                |
|               | 50         | 520,85 ± 29,66  | 138,61                     | 0,43                |
|               | 100        | 916,16 ± 31,58  | 252,39                     | 0,59                |
|               | 200        | 327,25 ± 21,39  | 477,48                     | 0,90                |
|               | <b>500</b> | 5,31 ± 0,34     | 1125,06                    | 4,66                |

Table 4: Results obtained through the use of global appearance descriptors based on *places* CNN (layers '*conv4*', '*conv5*', '*fc6*', '*fc7*' and '*fc8*') to solve visual localization.

| Layer        | Size       | Error loc. (cm) | Time comp. descriptor (ms) | Time pose est. (ms) |
|--------------|------------|-----------------|----------------------------|---------------------|
| conv4        | 169        | 5,03 ± 0,02     | 6,64                       | 1,62                |
| <b>conv5</b> | <b>169</b> | 5,09 ± 0,17     | 6,66                       | 0,63                |
| fc6          | 4096       | 5,14 ± 0,18     | 7,42                       | 34,38               |
| fc7          | 4096       | 16,71 ± 0,84    | 8,58                       | 33,22               |
| fc8          | 205        | 24,22 ± 6,44    | 8,88                       | 0,72                |

dynamic dataset with panoramic images has been used to evaluate the experiments. Five global appearance descriptors have been evaluated: two based on analytic methods (HOG and *gist*), two based on auto-encoders and one based on CNN layers. The size of each descriptor is varied through either tuning some parameters (such as the number of bins in HOG or the size of hidden representation of the auto-encoders) or selecting a different layer in the CNN case. The localization error, the computing time to calculate the descriptor and the computing time to estimate the position of the robot have been used as parameters to measure the efficiency of these descriptors. The fig. 4 shows the results obtained for the best configuration of each descriptor evaluated. From that figure, we can conclude that the minimum localization error is ob-

tained through the CNN-based descriptor option, but the *gist* descriptor and the auto-enc-SUN descriptor show results quite similar. The CNN-based descriptor introduces also the best option regarding the computing time to calculate the descriptor. Nevertheless, regarding the time to estimate the pose of the robot, HOG is the fastest.

Regarding the use of auto-encoders, using an auto-encoder which has been trained with images that belong to the environment outputs good-enough accuracy results. The general auto-encoder proposed through training a generic panoramic dataset works acceptably in the case of high size of hidden representation, hence this leads to high computing times. Nevertheless, its use as tool to obtain global appearance descriptors for panoramic images would be valid and

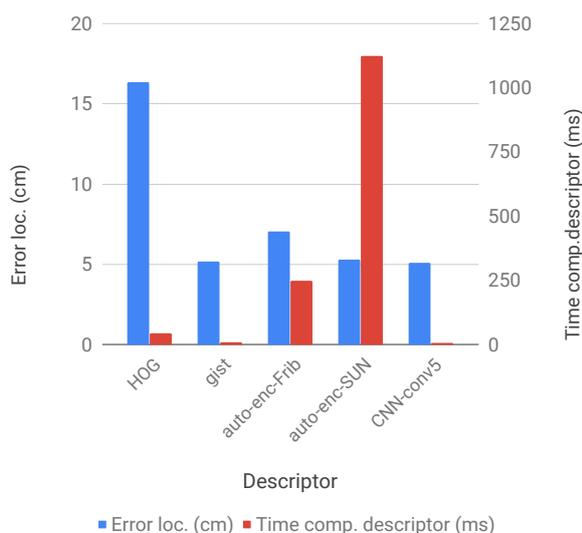


Figure 4: Summary of the best configuration for each descriptor studied.

the advantage of this method is that the auto-encoder is trained just once, then the tool is suitable independently the environment.

As for the use of CNN-based descriptors, we have proved that the first layers can output very interesting descriptors despite these are not fully convolutional layers (typically proposed to obtain descriptors). Moreover, the descriptors related to the 'conv4' and 'conv5' layers have produced the optimal localization solutions among all the methods evaluated: size of descriptor relatively small (which leads to fast times to estimate the position), low computing time to calculate the descriptor and very accurate localization (average error around 5 cm for a test dataset and a training dataset whose average distance between images is around 4 cm and 20 cm respectively).

## ACKNOWLEDGEMENTS

This work has been supported by the Generalitat Valenciana through grant ACIF/2017/146 and 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."

## REFERENCES

Cebollada, S., Payá, L., Mayol, W., and Reinoso, O. (2019). Evaluation of clustering methods in compression of topological models and visual place recognition us-

ing global appearance descriptors. *Applied Sciences*, 9(3):377.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA. Vol. II*, pp. 886-893.

Gao, X. and Zhang, T. (2017). Unsupervised learning to detect loops using deep neural networks for visual slam system. *Autonomous robots*, 41(1):1-18.

Hubens, N. (2018). *Deep inside: Autoencoders*. <https://towardsdatascience.com/deep-inside-autoencoders-7e41f319999f>. Accessed February 11, 2019.

Korrapati, H. and Mezouar, Y. (2017). Multi-resolution map building and loop closure with omnidirectional images. *Autonomous Robots*, 41(4):967-987.

Leonardis, A. and Bischof, H. (2000). Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99-118.

Mancini, M., Bulò, S. R., Ricci, E., and Caputo, B. (2017). Learning deep nbnn representations for robust place categorization. *IEEE Robotics and Automation Letters*, 2(3):1794-1801.

Naseer, T., Burgard, W., and Stachniss, C. (2018). Robust visual localization across seasons. *IEEE Transactions on Robotics*, 34(2):289-302.

Oliva, A. and Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. In *Progress in Brain Research: Special Issue on Visual Perception. Vol. 155*.

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.

Román, V., Payá, L., and Reinoso, O. (2018). Evaluating the robustness of global appearance descriptors in a visual localization task, under changing lighting conditions. In *ICINCO 2018. 15th International Conference on Informatics in Control, Automation and Robotics, Porto (Portugal), 29-31 July 2018*, pages 258-265.

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 (RO-BIO)*, pages 2377-2383.

Xiao, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2012). Recognizing scene viewpoint using panoramic place representation. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 2695-2702.

Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., and Madabhushi, A. (2016). Stacked sparse auto-encoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE Transactions on Medical Imaging*, 35(1):119-130.

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, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495.

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.