DR 5.2.3: Hierarchical Structure of Learned Skills, Scan-paths, Saliency Map of Activities and Communication Interfaces

Mario Gianni, Panagiotis Papadakis, Fiora Pirri, Matia Pizzoli and Arnab Sinha

Dipartimento di Ingegneria Informatica, Automatica e Gestionale “Antonio Ruberti”-Sapienza Università di Roma, via Ariosto 25, 00185 Rome, Italy
⟨pirri@dis.uniroma1.it⟩

This document describes the progress status of the research on Hierarchical Structure of Learned Skills, Scan-paths, Saliency Map of Activities and Communication Interfaces performed by the NIFTi Consortium. As per the description of work, the research reported in this document concerns the WP5 for the Year 2 of the NIFTi project.

Planned work is introduced and the actual work is discussed, highlighting the relevant achievements, how these contribute to the current state of the art and to the aims of the project.
1 Tasks, objectives, results

1.1 Planned work ................................................. 5
1.2 Addressing reviewers' comments ..................................... 5
1.3 Actual work performed ........................................... 6
  1.3.1 Task T5.2: Learning skills for functioning processes and task execution 6
  1.3.2 Task T5.3: Task-driven attention for coordination and communication 7
  1.3.3 Task T5.4: Adaptive behaviours in flexible temporal planning ........ 10
  1.3.4 Task T5.5: Situated exploration history ...................... 13
1.4 Relation to user-centric design .................................. 16
1.5 Relation to the state-of-the-art .................................. 17

2 Annexes

2.1 Gianni et al. “A unified framework for planning and execution-monitoring
  of mobile robots ”(aaaiPAMR2011) .................................. 22
2.2 Gianni et al. “Awareness in Mixed-Initiative Planning ”(aaaiRHTDAE2011) .......... 23
2.3 Ntouskos et al. “A Comprehensive Analysis of Human Motion Capture
  Data for Action Recognition ”(visapp2012) .......................... 24
2.4 Sinha et al. “Gap Detection and Traversability Analysis using LIDAR for
  Safe Robot Navigation”(IVS2012) ..................................... 25
2.5 Pirri et al. “3D Saliency Maps”(WBCV2011) .......................... 26
2.6 Mancas et al. “Human-Motion Saliency in Multi-Motion Scenes and in
  Close Interaction ”(GW2011) .......................................... 27
2.7 Pirri et al. “Human Motion Saliency in Complex Scenes”(LNAI2012) .......... 28
2.8 Mancas et al. “From Saliency to Eye Gaze: Embodied Visual Selection for
  a Pan-Tilt-Based Robotic Head”(ISVC2011) .......................... 29
2.9 Pirri et al. “A general method for the Point of Regard estimation in 3D
  space.”(CVPR 2011) .................................................. 30
2.10 F. Pirri and M. Pizzoli “Inference about Actions: Levesques view on action
  ability and Dirichlet processes.”(KRA 2011) .......................... 31
2.11 H. Khambhaita et al. “Help Me to Help You: how to Learn Intentions,
  Actions and Plans”(AAAI-SSS11) ........................................ 32
2.12 Papadakis et al. “Constraint-free Topological Mapping and Path-Planning
  by Maxima Detection of the Spatial Clearance Density. ”(ICPRAM 2012) .......... 33
2.13 P. Papadakis and F. Pirri “Consistent Pose Normalization of Non-Rigid
  Shapes using One-Class Support Vector Machines. ”(NORDIA 2011) .......... 34
Executive Summary

This document describes progress of the research on Hierarchical Structure of Learned Skills, Scan-paths, Saliency Map of Activities and Communication Interfaces performed in NIFTi in WP5, Year 2.

Planned work is introduced and the actual work is discussed, highlighting the relevant achievements, how these contribute to the current state of the art and to the aims of the project.

Most of the work of the second year is related to the planning requirements for both the exploration of an unknown and dynamic environment, and for mixed initiative. These concepts, now, had to suitably adapted to a completely new robot platform. The new NIFTi robot has a complex morphology, requiring control to be flexible, in any sense, so as to accommodate its behaviors to the variability of the context.

Under this perspective, WP5 concentrates on those planning capabilities that are related to switching among different control behaviors, namely modalities, and cognitive robot control. To this end a partial model of the 3D structure of the environment, according to saliency and the need to change the focus of attention, possibly induced by a close interaction with the operator, has been done. Desirable skills at the current state of the platform functionalities development are morphological adaptation, switching tasks capabilities, and mixed initiative, that is asking for operator intervention or accepting operator intervention, when required.

In particular:

1. Task 5.2 (T5.2) has concentrated on modeling the cognitive skills of the NIFTi UGV related to task switching.

2. Task 5.3 (T5.3) addresses the research on task driven attention, and more specifically the problem of defining saliency measures suitable for robot communication and for 3D navigation.

3. Task 5.4 (T5.4) addresses the morphological adaptation problem. It provides methods and paradigms for learning the required skills to control the reconfigurable components of the UGV using a dynamic Bayesian model.

4. Task 5.5 (T5.5) addresses the advancements in the situated exploration history.

In synthesis, this document reports the research carried on by WP5, more specifically the results of tasks T5.2, T5.3, T5.4 and T5.5.
Role of Flexible planning, learning and execution for joint exploration in NIFTi

During the exploration of an unknown area, the NIFTi human-robot team holds a continuous interaction so as to yield a joint plan. Still, one of the team operators, including the robot, can identify sudden events which require a change of plan. Hence the robot has to adapt to the new requirements issued by the dynamic context. This implies that execution should address sudden, asynchronous requests.

Under these principles, WP5 addresses the following items:

- Definition of the skills required to select and coordinate multiple tasks in operation.
- Specification of the paradigms and methods to learn the required skills from continuous interactions with operators.
- Design of the cognitive execution and monitoring architecture supporting flexible planning extended with task switching.
- Design of the models providing robot adaptation to address the different scenarios and user requirements derived from joint exploration.

Contribution to the NIFTi scenarios and prototypes

WP5 addresses the development of methods and paradigms indispensable to learn skills. These, in turn, are needed to both making good plans and manage the tasks that the team has to accomplish to cope with the goal requests in the USAR scenario. A substantial contribution to the integration of all the planning facilities both in ROS, lower levels, and in CAST (in cooperation with WP3), higher level and memory, has been done, for the development of the NIFTi cognitive architecture.

Likewise for the NIFTi scenario and the user model, the saliency data collected in Year 1 have been widely used. This is a training set of observations on demonstrations, performed by humans, made available by experiments with the Gaze Machine framework. Where humans look, how they move, what actions they perform and how they report what they see and do through speech constituted the training data for skill learning. Indeed, human visual attention as a selection strategy for demonstration data represents a novel contribution definition of a model for flexible planning. In the context of human action recognition to support joint planning and execution, we have performed extensive experiments on databases of human Motion Capture sequences that have prescribed the degree of dimensionality of human actions that is necessary to be captured.
Furthermore, methods that endow the NIFTi ugv with the skills for adapting its morphology can be evaluated within the 3D simulation environment in ROS wherein we have modelled the physics of the ugv within a given environment. Various categories of common negative or positive obstacles encountered in NIFTi scenarios such as gaps, inclined planes and staircases have been determined and perception methods together with traversability analysis methods are developed on this basis (in cooperation with WP1).

1 Tasks, objectives, results

1.1 Planned work

Planned work for WP5 in Year 2 amounted to improve the flexible planning framework with learned skills to ensure mapping internal states to execution. Also it was necessary to ensure that the robot knows the resources that have to be deployed in the ongoing mission. With the new NIFTi platform, a crucial issue arises for the ability to generate a 3D scan-path according to task driven attention and, possibly, selecting those regions in which some human activity is taking place.

1.2 Addressing reviewers’ comments

According to the reviewers comments, concerning the timing of a fully operational UGV, we have concentrated the work of this second year on the problem of morphological adaptation and integration of planning with the new components asked for by the new robot. Thereupon we have worked on the design of switching tasks so as to allow for a suitable control of the new components with new requisites. Due to the great novelty of morphological adaptation, our work on a full embedding of the UGV motion into the planner is still in progress. Despite this, we have integrated the flexible planning functionalities and we have specified the structure of a planner that can suddenly change task execution according to incoming stimuli.

A great effort has been pursued to improve the software architecture for planning and execution, which is now able to support sophisticated requirements as it holds all the proprio- and extero-perceptual functionalities of the robot, including them into appropriate representation at the memory level in CAST and therefore at the suitable representation level in the logical framework.

On the other hand despite the suggestion of a reviewer to make available online the functionalities of the Gaze Machine, so as to obtain a huge amount of data for skill learning, we have considered to delay this goal. Indeed, we gathered that it would have had a better effect to conclude the development of the methods for the construction of a 3D saliency map which is a completely novel contribution, before making our software available. The
methodology for the construction of the saliency map is succinctly addressed in §1.3.2.

1.3 Actual work performed

WP5 contributes to NIFTi project via the realization of a flexible time planner with time interval compatibilities, resources and components management related to different robot processes. It also contributes to the execution and monitoring of the planned actions, and to all those activities that make planning adaptable to the user needs, to the task requirements, to joint-proviso of team-work and to the processes and resources to be allocated for a designated mission. These activities include skill learning, accommodation to end-user planning procedures, modeling of end-users behaviors and most significantly of end-users instructed attentive behaviors while performing risky operations.

1.3.1 Task T5.2: Learning skills for functioning processes and task execution

Task 5.2 initiates at month 0 and ends at month 24. It commits 13 PM. T5.2 addresses the problem of how a robot can control task execution in dynamic scenarios.

1. We have studied the degrees of complexity in terms of dimensionality of human action representation, by extensive experimentation in databases of human Motion Capture (MOCAP) sequences. Based on the obtained results, we can largely constrain the number of dimensions needed to capture the evolution of human actions and thereafter enable the robot to learn the necessary skills for functioning and executing its own process §2.3.

2. We have formulated a task-switching paradigm which combines the declarative temporal model of the NIFTi UGV robot with a probabilistic stimuli-response model to account for the adaptation and the flexibility of the behaviours of the rescue platform in response to changing environmental demands. This paradigm allows us to represent the task set which is composed of the task-relevant stimuli, the task-relevant responses and the corresponding stimulus-response mapping. The hybrid model-based control allows the NIFTi UGV robot to not only respond reactively but also to behave in a more proactive way to achieve goals and perform tasks.

In the stimuli-response model the stimuli are represented as differences between events generated by the processes running on the system whenever output data are produced (e.g. the process for detecting objects in
the acquired video stream generates an event whenever an object has been detected. We call this events *yields* of the running processes.

The yields are a representation of the data which can be produced by the running processes. The model provides a way to measure the events associated to the incoming raw data in so measuring the change of the events over the time and determining the occurrence of the stimulus. The measure of an event is based on a probability density function defined on the *features* of the yields. Features represent the types of the incoming data (e.g. *car*, *victim* and *landmark* are the types of the detected object).

The instances of the yield are represented in the stimuli-response model by a K-dimensional binary vectors and modelled using a multinomial distribution conditioned on the parameters of the features which have been determined in a Bayesian treatment by choosing the Dirichlet distribution as conjugate prior. Given the probability distribution on the instances of the yields we compute the discrete derivative of the probability of the feature vectors at each time step and we verify the occurrence of the stimuli by suitably analysing the values of these derivatives which contain information about the rarity of the *yields*. The occurrences of the stimuli select all those instances of the yields which have determined them. This subset of instances is used to choose the responses of the system to that stimuli. The mapping stimulus-response is modelled by a multivariate Pólya distribution defined on the possible behaviours of the system.

### 1.3.2 Task T5.3: Task-driven attention for coordination and communication

Task 5.3 initiates at month 0 and ends at month 24. It commits 13 PM.

In Year 2 the research on task-driven attention for coordination and communication has focused on the modeling of the attentive selection strategies to support the coordination of the robot intervention and the communication with the human operators. In particular, three main directions can be seen in the T5.3-related research:

1. the definition of a novel concept of saliency, taking into account 3D cues, which can help reducing the complexity of the operations involving 3D perception;

2. the development of motion selection strategies tuned on human gestures, supporting action recognition and non-verbal communication;

3. the investigation on how to actually use the saliency of a scene to control an active vision system.

During Year 2, the NIFTi robotic platform has become available to the consortium. The requirements to dynamically adapt the robot morphology
according to the current task and terrain have posed challenging problems related to the robot 3D perception.

Under the above conditions a strong collaboration with other WPs such as WP1, WP2 and WP6 has been pursued during the meetings so as to model the data necessary to develop the formalization indicated in the above mentioned items 1 and 3. In fact, part of the work on the point cloud 3D cues had required a common work with task T1.2 and T1.3, likewise for item 3 a collaboration has been pursued mainly on the management of the IMU together with T2.4 and T2.5.

On the other hand, at the conceptual level a collaboration is implicit with WP3 and WP4. Clearly, while in our methodology we work mainly with raw data and their transformations under suitable mathematical mapping (from projective transformation to 3D Gabor transformation), we collaborate with WP3 and WP4 when the conceptualization of the modalities is required. So we have collaborated with Task T3.2 on adaptive multi-modal HRI for joint exploration, especially on the problem of reducing the span of the 3D point-cloud. It can be shown that an appropriate sparsity of the cloud to cope with interesting-salient regions simplifies several interface problems, concerning situation awareness, and we are still working on this issue. With respect to WP4 we have collaborated on the model of attention to motion, that can be used in the end-user experiments. Indeed, selectional attention (Task4.1, and Task 4.2) requires a tracking model of human states contextual dynamics. This, under the perspectives of item 2 above, includes the association of the robot ego-motion with the subject potential ego-motion. Potential since the operator is in fact not moving, therefore this differential might be conceptualized under task allocation for selectional attention.

The study on three-dimensional saliency, which is based on the data collected in GM experiments §2.5,§2.11 has been applied to the NIFTi robot sensors, namely to the image streams from the omni-camera and the point-clouds from the laser scanner. Attentive features are computed from the images and the correspondent location in the 3D world is recovered by matching in the point-cloud making use of the fixed, known transformation between the camera image planes and the laser scanner. This attentive search for 3D features is limited to a perceptual box, a volume around the robot whose dimensions vary according to the task.

**Problem with point cloud registration**  Point cloud registration and robot pose estimation from point cloud are synonymous problems. If we know the current and past robot pose estimations, then we can estimate the transformation between these two poses, hence the point clouds from these two different poses can be registered into a world coordinate frame. There are mainly three problems in our specific case:

1. Non-uniform point cloud: Due to the rotating laser, the point cloud
distribution depends not only on the surface of the object, but also the distance and orientation with respect to the robot.

2. Dynamic registration: Since the laser rotating speed is comparative to the robot speed, therefore, two consecutive scans will not come from the same coordinate system. The dynamic registration problem can be minimized by considering other inputs, e.g., IMU, track odometry, etc.

3. Computation complexity or time complexity: This is a well known problem in point cloud registration. Since, in our specific case it has to be real-time, it becomes a major constraint.

All the three problems described above are very crucial in our particular case. If the number of points are minimized in such a way, that, it satisfies a minimax criterion, then the registration algorithm can be also real-time required for the 3D saliency and other modules. We term it as a minimax criterion, since the criterion is minimum of maximum with respect to a measure. Maximum in the sense that, if we include more points it will only increase the redundancy in the optimization process. And, minimum in the sense that, if we reduce it more than the present size then it might incur huge error in the robot pose estimation (i.e., the point cloud registration). Moreover, this initial processing of the point cloud to minimize the number of points, has to be also within some milli-seconds, otherwise it will loose its purpose.

Concerning task driven attention for communication, WP5 research has focused on developing a new and original method for human motion analysis and evaluation, developed on the basis of the role played by attention in the perception of human motion. The paper §2.6 introduces a combination of two different attentional filtering mechanisms with the aim of clustering salient human gestures:

- Attentional selectivity towards human motion, which has the purpose of segmenting human motion in scenes depicting several different kinds of motion.

- Clustering of salient gestures, represented as novel, non-repetitive gestures while the repetitive and less informative movements are discarded.

In order to move from a measure of scene saliency to active vision and attentive exploration, during the Year 2 of the NIFTi project the use of saliency maps to control the acquisition of visual information has been investigated. Indeed, saliency models provide a biologically consistent prediction of eye movements, in the sense of the frequency with which human beings attend regions in the observed scene. In contrast, control models of eye movements have
been designed to compute the optimal sequence to move the visual sensors in order to achieve a given task, usually modeled as the problem of maximizing a certain measure of information or reward. The work presented in §2.8 bridges the two approaches, by providing a description of an active vision system that is controlled according to a dynamic model on the basis of a saliency map. The goal of this research is to provide the consortium with a strategy to computationally generate plausible scan-paths on the basis of a saliency map. The gaze predicted accordingly depends on the selection of:

- Where to look next? The saliency maps are usually characterized by several maxima. In addition, the maxima are not stable and a strategy is needed in order to select the next location to attend.

- What gaze behavior should be implemented? The system have been provided with the capability to fixate or pursue a current target represented by a saliency map local maximum or to change the focus of attention, by performing a saccade to the local maximum to be attended next.

1.3.3 Task T5.4: Adaptive behaviours in flexible temporal planning

Task 5.4 initiates at month 12 and ends at month 48. The objective of T5.4 is to develop methods for adaptive flexible plan execution on the basis of the learned skills.

1. We have modelled the kinematics and physics model of the NIFTi UGV robot within the ROS simulation environment (Gazebo) which provides the off-line training setup wherein the morphological adaptation of the robot can be learned and evaluated off-line so that it can later be deployed on-line §2.2.

2. We have formulated and tested a preliminary approach to detect and analyse the traversability of gaps in the vicinity of the UGV. The acquired 3D point cloud is processed, possible gaps are detected and their contours are extracted together with features that describe if and how the UGV should adapt its behaviour in terms of motion planning, in order to safely execute its motion plan (submission §2.4).

In order to optimally adapt the pose and morphology of the robot with respect to the terrain, we formulate an optimization problem. A measure of optimal robot pose can be defined combining together various criteria used to formulate the optimization problem. Candidate optimality-cost criteria reflecting robot properties are the following:

- Ground Clearance; the minimum distance from the robot coordinate frame to the terrain below it.
• **Robot orientation;** roll/pitch.

• **Distance Stability Margin;** the minimum distance between a support robot point and its projected center of mass, required to tip over the robot.

• **Gradient Stability Margin;** the minimum angle between the gravitational vector and a support robot point (both emanating from the robot center of mass), required to tip over the robot.

• **Wheel traction;** depends on the ratio of tangential to normal contact forces over all wheels.

Next to the above properties, inherent terrain properties contributing to the assessment of traversability are the following:

• **Elevation statistics;** Roughness, slope.

• **Proximity to obstacles.**

Using a sufficiently large variety of terrain maps (some examples are shown in Figure 1) that are characterized from the different ranges of the parameters involved in the optimization, we solve the optimization problem off-line within the simulation environment. This off-line stage constitutes the learning phase where the necessary skills for the morphological adaptation are learned and later employed in the real scenario wherein the robot would be asked to traverse uneven terrain.

**Configuration space:** The configuration space for the robot is defined by the position of the robot, its orientation (roll, pitch, yaw), the rotation angle of the flippers and the differential lock of the tracks.

![Figure 1: Terrain Classes; (a) hole, (b) step and (c) inclined plane.](image-url)
State space: Based on the constraints imposed by the \textit{arc-based} path planning process and the contact of the robot with the terrain (the elevation map), part of the configuration space is largely constrained. The remaining sub-state space that is under the control of the morphological adaptation regards mainly the angles of the flippers and the differential switch control. Each state is assigned a cost attribute that represents the traversability of the robot, in this particular state and stores the full state (including the rotations of the flippers of the robot in the optimal setting. Optimality and traversability are equivalent concepts in the sense that the state with minimum cost is the optimal-best state. The cost is computed as a feature aggregating the criteria listed earlier. Different weightings-aggregating cost functions can be mapped to different planning strategies that can be called by the high-level planner, depending on the requirements.

Action space-Control space: The action space is determined based on the motion-path planning algorithm. Constraints with respect to flippers concern avoiding self-collision and over-excessive weight to be put on any of the flippers. The action space with respect to the vehicle’s velocities and accelerations are mainly under the responsibility and control of the path planner.

With respect to gap perception and analysis, we have developed an algorithm that is based on: (i) application of image morphological and contour detection algorithms for gap perception and (ii) application of Principal Component Analysis in the orientation domain of the gap contours and extraction of optimal traversable direction. Afterwards we reason about
the traversability of detected gaps in order to determine whether a robot should completely avoid approaching the gap, or alternatively, it could be able to traverse over it under certain criteria. A representative example of the applications of this method is shown in Figure 2. For more details please refer to the paper (submitted) §2.4. With respect to the reasoning and detection of the gaps, a collaborative work has been pursued with respect to work-packages WP1 and WP6. Moreover, the representation of the gaps inside the user interface is done with work-package WP3. The methods for adaptive flexible plan execution which have been developed for the UGV, have been also extended in coordination with WP7 in order to be usable for the UAV.

1.3.4 Task T5.5: Situated exploration history

Task 5.5 initiates at month 0 and ends at month 48.

1. We have designed a model-based control of the main components and processes of the NIFTi UGV robot §2.1. The model is specified in the Temporal Flexible Situation Calculus (TFSC) which is a framework to suitably represent temporally-flexible behaviours in the Situation Calculus [18, 19]. This hybrid framework combines temporal constraints reasoning and reasoning about actions.

2. We have integrated the model-based control with incoming perceptual information from vision, SLAM, topological map segmentation and dialogue.

3. We have deployed an hybrid CAST subarchitecture which embeds the ECLIPSE Prolog implementation of the model-based control as well as a set of ROS nodes which are responsible of the communication with the ROS layers of the architecture of the NIFTi UGV robot. The CAST subarchitecture for the planning allows the human operator to switch between several operational modalities lying between autonomous and teleoperated modes during the execution of a task.

4. The planning subarchitecture provides communication interfaces with the human in both the directions: from the human to the robot and vice versa. The research related to representing tasks and responsibilities and authorizations in a shared display has been contributed by TNO as described in the work [16]. The research related to the situated dialogue processing and its integration into task-driven collaborative context as well as how the communication has been grounded in the human-robot collaboration has been contributed by DFKI as described in the work [31].
Figure 3: The network states that in order to reach the destination node \textit{Node} of the graph \( G \), the system has to scan the environment, rotating the 3D laser with the angular velocity \( \omega \), segment the free space in order to build the graph \( G \). During the execution of the navigation goal a suitable set of parameters \( \alpha \) (e.g. flipper angles, velocity, acceleration) have to be available for the reconfiguration of the pose of the robot.

The model-based control of NIFTi UGV robot is composed of a TFSC model of the controllable activities and a planning engine. The TFSC model explicitly represents the main components and processes of the NIFTi UGV robot, the cause-effect relationships as well as the temporal constraints among the processes. These processes are represented in the TFSC through fluents and instantaneous starting and ending actions which are defined in terms of preconditions and effects. Hard time constraints among the processes are managed by the TFSC model using Allen-like temporal relations [1].

The interactions between the set of the processes of the NIFTi UGV robot can be defined by a Temporal Constraint Network (TCN) [11], which represents the temporal relations among a set of parametric processes. Figure 3 shows a Temporal Constraint Network which defines a set of the temporal constraints on a subset of activities of the NIFTi UGV robot.

Temporal Flexible Situation Calculus intermediates between Situation Calculus formulae and Temporal Constraint Networks. Each temporal constraint between processes can be defined by a \( \mathcal{L}_{TFSC} \) formula mentioning sets of timelines. The consistency of the temporal constraint network allows us to represent the set of temporal constraints on a subset of processes of our system with a consistent formula of the language \( \mathcal{L}_{TFSC} \). The time variables mentioned in the formulae are assigned according to a specified set of behaviours, in so being very well suited for representing the flexible behaviours of the NIFTi UGV robot.
The planning engine is composed of two main logical modules: the plan generator and the execution monitoring. The plan generator relies on a library of Prolog scripts designating the set of tasks which the mobile robot can perform, according to the specified processes, their temporal constraints and preconditions. The tasks are based on the perceived information of the environment which is represented in the knowledge of the planning engine in the form of a graph augmented with information about the type and position of detected objects, within the map, such as cars or containers. The nodes of the graph correspond to topologically segmented regions or approachable areas around a detected object and edges determine the traversability between the regions.

The execution-monitoring ensures that the set of action sequences, generated by the plan generator, according to the TFSC model and the current state of the domain knowledge, are consistently executed. Furthermore the execution-monitoring manages the interventions of the human operator during the interaction with the control system.

Both the TFSC model and the planning engine are implemented in ECLIPSE Prolog [2] which optimally combines the power of a constraint solver with inference in order to generate the set of action sequences, and also enable the continuous update due to incoming new knowledge.

An hybrid CAST subarchitecture has been deployed in our mobile system in order to fully embed the TFSC model and the planning engine as well as a set of ROS nodes which are responsible of the main communication tasks.
with the ROS layers. Figure 4 illustrates the role of each component of the planning subarchitecture which we have designed for the robot.

The core of the planning subarchitecture is implemented by the Execution-Monitoring component which plays a crucial role in orchestrating the other components. This component manages the communication with the human interfaces and embeds into the subarchitecture the logical part of the control system. It enables the human operator to interact with the control system during the computational cycle. Further, this component allows the human operator to modify the control sequence produced by the planner by skipping some activities, adding new actions. It also allows the operator to take the control of some functional activities while the rescue robot is executing a task.

On the other hand the Execution-Monitoring component manages the external and internal perceptual components of the planning subarchitecture which communicate with the CAST subarchitectures/ROS modules implementing the perception capabilities of the NIFTi UGV robot. The information acquired from the perceptual components is written in the working memory WM and compiled in order to build the domain knowledge of the planning engine. The design of messages and data exchange coming in and out from the visual detector and associated with auxiliary ROS nodes has also involved CTU.

The component is also responsible for sending task activation signals to the actuator components of the subarchitecture in order to perform the sequences of actions generated by the planning engine.

The actuator components rely on the skill components to select the suitable set of parameters of the actions according to the internal state of the system and the features extracted from the 3D point cloud of the environment.

In other words the morphological adaptation generates sequences of parameters to configure the pose of the robot in order to reduce instabilities that could cause the tip-over. These sequences are for example flipper angles, velocity and acceleration. In order to generate the above defined parameters, the morphological adaptation component takes as inputs the 3D point cloud registered by the lidar component and the route computed by the navigation component to reach the destination point in the map.

The research related to temporal organizations over grounded planning and communication content started at the beginning of Year 1 and it will be concluded at the end of Year 4.

1.4 Relation to user-centric design
We have approached all main tasks of WP5 from a user-centric point of view.

Data collected in the end-user evaluation in Montelibretti have shown
that the operators need to feel more confident even in tele-operated driving. The operator, indeed, expects to feel in the field when observing the robot acting autonomously and when tele-operating it. To partially cope with these problems we have addressed attentive exploration and mixed initiative.

Endowing machines with the capability to attentively select relevant information in a natural scene requires modeling the phenomenon that is responsible of the impressive performance exhibited by humans. The first step is to obtain measures of the studied phenomenon and, thus, an instrument is needed to measure the deployment of attention. A novel framework, the Gaze Machine, has been developed to collect and analyze experimental data.

The planning subarchitecture in CAST has been developed so as to allow the human operator to switch between several operational modalities lying between autonomous and teleoperated modes during the execution of a task. Our user-centric endeavor is to support the cooperation between the user and the NIFTi UGV robot in accomplishing the task. The subarchitecture effectively intermediates between a model of the human interventions and the declarative control model of the NIFTi UGV robot. The incoming user commands are suitable parsed in order to be integrated in the control plan and in turn the behaviours of the NIFTi UGV robot are interpreted to be understandable for the operator. The mixed-initiative planning approach has enhanced both the operator’s situation awareness and human-robot interaction for the execution and control of diverse activities within USAR.

1.5 Relation to the state-of-the-art

Various directions have been explored in addressing the gap detection problem, still this considered an open problem. In the following we try to sum-
marize the seminal and most representative papers in the field. Dima et al. [12] used feature and classifier fusion for obstacle detection and terrain traversability where the basis features that are computed for various perceptual modalities, correspond to the mean and variance of pixel values along a set of image patches that span the whole image. Combining features that incorporate domain knowledge [32], different classifier fusion strategies are evaluated that show improved classification scores for road, human and negative obstacle detection, in comparison to single feature-based classifiers. In the work of Larson et al. [34], terrain traversability is partly determined by the presence of negative obstacles, wherein patches of missing range data that exceed some size are considered as potential negative obstacles and a consecutive filtering process determines whether they could be the result of shadowing from positive obstacles. In [33], the authors explore a two-stage (long and short-range) negative obstacle detection framework. Initially, potential negative obstacles are detected at a distance using the NODR classification approach and then further refined and filtered using support vector machines (SVM) when the robot has sufficiently approached the surrounding area. NORD comprises a multi-pass detection process that first looks for steps and next for gaps whose characteristics can either be directly measured from the available range data, or inferred by using contextual cues, such as sudden elevation drops. Eventually, using an SVM model trained on ground truth data, true and false positives of negative obstacles are distinguished once the robot has sufficiently approached. An alternative approach for perception of negative obstacles during night is proposed by Matthies et al. in [38], wherein range data are combined with thermal terrain features that highlight cavities as potential negative obstacles. Their approach was based on the observation that negative obstacles retain more heat during night than planar surfaces. In [15], obstacles are detected in the vertical image direction using dense 3D terrain data reconstructed from stereo disparities. First, a disparity validity measure is used together with an image pyramid to produce reliable disparity estimates. In the following, traversability is computed for each pixel of the disparity image by estimating the maximum vertical slope and using hysteresis thresholding driven by morphological opening and region filling. Heckman et al. [23] perform detection of potential negative obstacles by first performing ray-tracing for occlusion labelling and finally context-based labelling. Given a 3D voxel grid where cells are classified into linear, surface and scatter, ray-tracing is used to propagate the class of occupied voxels to the corresponding occluded voxels while context-based labelling is used to differentiate between four cases that could be the cause of data absence and hence reason about the presence of negative obstacles. Kelly et al. [28] describe the design and operation of a human-robot team for off-road navigation, wherein terrain classification is based on geometry-based features combined with multi-spectral image-based features. The robot support surface is extracted by ray-tracing of
the laser-beams and training a neural network to derive the load-bearing surface when traversing over vegetated areas, while negative obstacles are found by the absence of laser hits in the direction perpendicular to the support surface. Finally we are trying to address the problem of traversability analysis, path planning, terrain analysis and adaptive morphology from an optimization perspective according to the work [24].

As far as task driven attention is concerned, the state of the art mainly addresses computational attention models, which have been successfully applied to improve the efficiency of computer algorithms and to guide the process of understanding what is observed in a scene. Recent years saw the development of computational models of visual attention [46, 47, 25] that allow a system to focus on crucial aspects for the task at hand.

The notion of saliency map [30] was introduced to predict eye movements in static scenes, the first models prevalently accounting for bottom up features and subsequently extended to model top down influences. Recently, motion cues have been taken into consideration [36, 3, 37], in so introducing time in the computation of saliency and moving the analysis of visual attention to dynamic sequences. Spatio-temporal analysis is performed to qualitatively segment a sequence according to primitive motion components [49], discriminating interesting motion patterns (e.g. coherent, static) from those portions that account for flickering, depict incoherent motion or are just too unstructured to support subsequent analysis. In the NIFTi scenario, the capability to select human motion in a scene is crucial in order to support subsequent analysis and interpretation. The work described in §2.6 advances the state of the art by contributing a computational model of visual attention that filters human motion in a scene depicting different kinds of motion and is also capable of selecting novel interesting gestures in a scenario of close interaction. This contributes to the definition of a saliency map of human activities.

Based on the eye physiology, sophisticated computational models of eye movements have been introduced for smooth pursuit [35] and saccades [51]. Still, the problem of defining a scan-path in the 3D world represents an open challenge. To support active vision, control models of eye movements have been designed to compute the optimal sequence to move the visual sensors in order to achieve a given task [6], usually modeled as the problem of maximizing a certain measure of information or reward. These models need to be explicit about the measure of information to maximize. In contrast, saliency models [30, 36] provide a biologically consistent prediction of eye movements, in the sense of the frequency with human beings attend regions in the observed scene. The work presented in §2.8 bridges the gap between control and saliency models, as it combines a motion-based saliency map with a model of eye movements, based on a non-linear Bayesian state-space process, in order to predict the focus of the attention accordingly and generate plausible scan paths.
State of the art approaches have been designed and validated on a very specific class of data: still images and videos. Eye tracking experiments have been performed extensively to study visual attention, on the other hand it is becoming clear that those stimuli often used in laboratory experiments, such as still images or videos that have been suitably edited, may not be representative of natural viewing behavior [13], hence the need to rely on approaches for data collection in the wild. Our research has focused on deriving a computational model of visual attention that can address the natural scenarios of a system perceiving and acting in the real 3D world. The goal is to learn how to compute the saliency of a natural scene, in which the observer freely moves and interact with an environment inducing dynamic stimuli. This resides in understanding what kind of features should be taken into consideration in similar settings. Therefore, the WP5 research on saliency maps of activities and scan path generation has been funded on data collected making use of the GM framework §2.9 §2.5, which has been employed to design and validate the computational models.

Learning 3D point cloud (registered with image) cues for learning a novel attentive model is the NIFTi vision. There are some recent works in this domain, such as [27]. In this work, the authors have tried to segment the point cloud according to the saliency from the stereo images. The salient point cloud is then given to a Gaussian mixture model to learn the color characteristics of the point cloud. Thereafter, an optimization module is performed on the graphical model (which was created with respect to the point cloud with colored points according to the images) for the final multi-class segmentation purpose.

The problem of task switching is well known in neurosciences, since the work of Jersild [26] further rediscovered by Spector and Biederman (the geons inventor) in [44]. Recently the research on task switching in psychology and neuroscience is becoming very active, a quick review is done in Monsell [39], and one of the most cited works is [43], whereas a recent work of reference is that of [10]. On the other hand the concept has been also used in operating systems to deal with interruptions and multitasking. In both cases the concept of switching is based on forms of inhibition and costs when concerning humans and on preemption when concerning machines. In machines the switch cost is seen as kernel preemption, requiring to set priorities among threads in order to ensure context switch. Clearly when it comes to CPU there is no need to model a stimulus. As opposed to CPUs, inhibition in live systems explains how a subject (an artificial system) in the presence of several stimuli responds selectively and is able to resist inappropriate urges (see [45]). In particular, Harnischfeger [21] defines cognitive inhibition as a form of forgetting previously activated cognitive processes and harnessing interference from processes or contents not relevant to the main current task. Inhibition, as a general function, explains flexibly switching between tasks, when reconfiguration of memory and perception is required, by disengaging
from previous goals or task sets.

When it comes to robots behaviours very few papers have addressed the problem, despite Rubinstein et al., in [43] refer to the cognitive architectures of Norman and Shallice [42] and Newell [41], mainly because in these works the concepts of cognitive control were highlighted.

Indeed the few papers that have so far addressed the concept such as [7] and [48] do this with a meaning closer to the use made in operating systems, without considering the stimulus activation and inhibition, which is otherwise crucial to model task switching for cognitive control. For example in [7] switching is defined between the tasks of reaching the sound source and getting close to as many light sources as possible.

The declarative temporal model of the NIFTi UGV robot has been specified in the Temporal Flexible Situation Calculus [18, 19]. The integration of the dynamic model of the main components and activities of the NIFTi UGV robot into the CAST subarchitecture follows the approaches §2.1, [22, 9, 8, 8]. The proposed model-based approach to control for the execution and control of the NIFTi UGV robot is related to [17, 4, 50]. The interfaces for the interaction between the human operator and the NIFTi UGV robot for the mixed-initiative planning have been designed taking into account the works [20, 40, 5, 29, 14].
2 Annexes

2.1 Gianni et al. “A unified framework for planning and execution-monitoring of mobile robots” (aaaiPAMR2011)


Abstract We present an original integration of high level planning and execution with incoming perceptual information from vision, SLAM, topological map segmentation and dialogue. The task of the robot system, implementing the integrated model, is to explore unknown areas and report detected objects to an operator, by speaking loudly. The knowledge base of the planner maintains a graph-based representation of the metric map that is dynamically constructed via an unsupervised topological segmentation method, and augmented with information about the type and position of detected objects, within the map, such as cars or containers. According to this knowledge the cognitive robot can infer strategies in so generating parametric plans that are instantiated from the perceptual processes. Finally, a model-based approach for the execution and control of the robot system is proposed to monitor, concurrently, the low level status of the system and the execution of the activities, in order to achieve the goal, instructed by the operator.

Relation to work performed In this work, we have described a mobile robot system that employs high-level control in order to operate in a real-world setting where the main task is human-assisted exploration of an environment. We have integrated multi-modal perception from vision and mapping with a model-based executive control. We have also showed how the system allows the interaction between the human operator and the robot platform via the dialogue-based communication. In this framework, action planning is performed using a high-level representation of the environment that is obtained through topological segmentation of the metric map and object detection and 3D localization in the map. This representation has the form of a graph where all the information related to the spatial characteristics of the environment is stored into properties that are annotated to the nodes and the edges of the graph that is used by the planner to generate task-dependent plans. The control system monitors the execution of the action sequences and communicates the status through the dialogue.
2.2 Gianni et al. “Awareness in Mixed-Initiative Planning” (aaaiRHTDAE2011)


Abstract For tasks that need to be accomplished in unconstrained environments, as in the case of Urban Search and Rescue (USAR), human-robot collaboration is considered as an indispensable component. Collaboration is based on accurate models of robot and human perception consistent with one another, so that exchange of information critical to the accomplishment of a task is performed efficiently and in a simplified fashion to minimize the interaction overhead. In this paper, we highlight the features of a human-robot team, i.e. how robot perception may be combined with human perception based on a task-driven direction for USAR. We elaborate on the design of the components of a mixed-initiative system wherein a task assigned to the robot is planned and executed jointly with the human operator as a result of their interaction. Our description is solidified by demonstrating the application of mixed-initiative planning in a number of examples related to the morphological adaptation of the rescue robot.

Relation to work performed In this paper, we describe the main features of a control system where HRI is fully based on a mixed-initiative planning activity that integrates the operator interventions with the robot activities. This control system allows us to define a set of hybrid operational modalities lying between teleoperated mode and autonomous mode that are crucial in a collaborative planning setting. The benefits of the mixed-initiative approach have been shown by focusing on the problem the morphological adaptation of the NIFTi robot with respect to particular tasks being performed in a simulated 3D environment, or in a real scenario.
2.3 Ntouskos et al. “A Comprehensive Analysis of Human Motion Capture Data for Action Recognition” (visapp2012)

Bibliography


Abstract

In this paper, we present an analysis of human motion that can assist the recognition of human actions irrespective of the selection of particular features. We begin with an analysis on the entire set of preclassified motions in order to derive the generic characteristics of articulated human motion and complement the analysis by a more detailed inter-class analysis. The statistical analysis concerns features that describe the significance-contribution of the human joints in performing an action. Furthermore, we adopt a hierarchical analysis on the human body itself in the study of different actions, by grouping joints that share common characteristics. We present our experiments on standard databases for human motion capture data as well as a new commercial dataset with additional classes of human motion and highlight certain interesting results.

Relation to work performed

We have analysed the complexity of a high-level, articulated human action representation in a plurality of classes across several standard datasets. The purpose of the analysis is to derive a lower dimensional subspace wherein human actions can be modelled by retaining the maximum amount of information using the minimum number of dimensions. In combination with the learned model of attention, this experimental analysis is useful in the learning of the robot skills required for functioning its processes through human demonstration.
2.4 Sinha et al. “Gap Detection and Traversability Analysis using LIDAR for Safe Robot Navigation” (IVS2012)


Abstract  In this paper, we address the problem of negative obstacle detection and analysis in the form of gaps within the vicinity of a robot, by processing LIDAR data acquired from an onboard 3D laser scanner. Our approach is based on the application of image morphological operations in order to detect gaps that could pose a danger to the robot and reason about the traversability of the gaps, by extracting PCA-based features of the orientation-direction distribution of the gap contour. We have evaluated our approach within a realistic scenario of a tunnel car accident and demonstrate its performance in a number of representative examples, using a contemporary Search and Rescue robot.

Relation to work performed  We have developed an algorithm that processes the 3D point cloud acquired from the UGV in order to detect gaps in the vicinity of the robot and evaluate how the UGV would safely navigate over or completely avoid them. This functionality integrates and augments the morphological capability of the UGV so that it can flexibly operate within a challenging Search and Rescue environment.
2.5 Pirri et al. “3D Saliency Maps” (WBCV2011)


**Abstract**  Eye tracking devices have been extensively used to study human selection mechanisms and promoted the development of computational models of visual attention, whose well known outcomes are the saliency maps. Among the eye trackers, wearable ones have the advantages of allowing the estimation of the Point of Regard (POR) while performing natural tasks, instead of experimental, static lab settings. The motion of the viewer makes localization necessary to collect data in a coherent reference frame. In this work we present a framework for the estimation and mapping of the sequence of 3D PORs collected by a wearable device in unstructured, experimental settings. The result is a three-dimensional map of gazed objects, which we call 3D Saliency Map and constitutes the novel contribution of this work.

**Relation to work performed**  In this paper the complete Gaze Machine Framework for the collection and analysis of salient gaze behaviors in real-life, unconstrained scenarios is presented and the 3D map of the volumes attended during the experiment is introduced. The collection of these maps is at the core of the WP5 research on skill learning, as it provided data to T5.2 on Learning skills for functioning processes and task execution and T5.3 on Task-driven attention for coordination and communication.
2.6 Mancas et al. “Human-Motion Saliency in Multi-Motion Scenes and in Close Interaction” (GW2011)


Abstract We briefly account for an original method for human motion analysis and evaluation, taken from two perspectives: human motion in a multi-motion scene and in close interaction. The first aspect concerns the saliency of human motion with respect to any other kind of motion in the scene. The second aspect, identified by upper limbs and thus involving gestures, concerns the extent of human attention to gesture modalities both in peripheral and central vision. We used a dynamic 3D gaze tracker for modeling and verification, while experiments were achieved on movies of gestures in several scenarios and generic youtube movies.

Relation to work performed This paper reports on the results of a preliminary investigation on saliency models for attention to human motion. The plausibility of two different models, one for attentive motion clustering and the other for the detection of salient gestures, is evaluated with the help of a 3D gaze tracker. The work is related to the WP5 research on saliency maps of activities and, in general, on task driven attention.
2.7 Pirri et al. “Human Motion Saliency in Complex Scenes” (LNAI2012)


Abstract  We present a new and original method for human motion analysis and evaluation, developed on the basis of the role played by attention in the perception of human motion. Attention is particularly relevant both in a multi-motion scene and in social interactions, when it comes to select and discern why and what to focus on. The first crucial role of attention concerns the saliency of human motion within a scene where other dynamics might occur. The second role, in social-close interactions, is highlighted by the selectivity shown towards gesture modalities both in peripheral and central vision. Experiments for both modeling and testing have been based on a dynamic 3D gaze tracker.

Relation to work performed  The paper contributes an integrated model to segment motion based on a novel saliency definition. As a first step, human motion is segmented in scenes depicting several different kinds of motion; then, interesting gestures are selected among less informative, repetitive gestures. The capability to automatically detect human motion and, in particular, gestures, is extremely valuable in the NIFTi scenario, in which the focus on natural human-robot cooperation demands the capability by the robot to interpret any form of communication issued by the human operators.
2.8 Mancas et al. “From Saliency to Eye Gaze: Embodied Visual Selection for a Pan-Tilt-Based Robotic Head” (ISVC2011)


Abstract  This paper introduces a model of gaze behavior suitable for robotic active vision. Built upon a saliency map taking into account motion saliency, the presented model estimates the dynamics of different eye movements, allowing to switch from fixational movements, to saccades and to smooth pursuit. We investigate the effect of the embodiment of attentive visual selection in a pan-tilt camera system. The constrained physical system is unable to follow the important fluctuations characterizing the maxima of a saliency map and a strategy is required to dynamically select what is worth attending and the behavior, fixation or target pursuing, to adopt. The main contributions of this work are a novel approach toward real time, motion-based saliency computation in video sequences, a dynamic model for gaze prediction from the saliency map, and the embodiment of the modeled dynamics to control active visual sensing.

Relation to work performed  This paper addresses the problem of using a saliency model to control an active vision system. The research described in the paper contributes to the WP5 objectives by providing a model to computationally generate plausible scan-paths. The model is trained making use of demonstration data collected in GM experiments.
2.9 Pirri et al. “A general method for the Point of Regard estimation in 3D space.” (CVPR 2011)


Abstract A novel approach to 3D gaze estimation for wearable multi-camera devices is proposed and its effectiveness is demonstrated both theoretically and empirically. The proposed approach, firmly grounded on the geometry of the multiple views, introduces a calibration procedure that is efficient, accurate, highly innovative but also practical and easy. Thus, it can run online with little intervention from the user. The overall gaze estimation model is general, as no particular complex model of the human eye is assumed in this work. This is made possible by a novel approach, that can be sketched as follows: each eye is imaged by a camera; two conics are fitted to the imaged pupils and a calibration sequence, consisting in the subject gazing a known 3D point, while moving his/her head, provides information to 1) estimate the optical axis in 3D world; 2) compute the geometry of the multi-camera system; 3) estimate the Point of Regard in 3D world. The resultant model is being used effectively to study visual attention by means of gaze estimation experiments, involving people performing natural tasks in wide-field, unstructured scenarios.

Relation to work performed The paper describes the novel contributions introduced by the video-oculography subsystem of the Gaze Machine, the framework that is being used to collect demonstration data and thus is as at the core of the skill learning paradigms for the WP5 tasks.
2.10  F. Pirri and M. Pizzoli “Inference about Actions: Levesques view on action ability and Dirichlet processes.” (KRA 2011)


Abstract  In this paper we have considered Levesques approach to modeling the belief that a cognitive robot has about observed actions. Modeling observations of actions is a basic step for any further inference about actions, when the inference concerns what a robot can do or how it can perform a certain task. This requires the robot to be able, for example, to correctly discern two observed actions, even when the two motion sequences are not identical. To treat sequences of point actions, we have interpreted actions as non parametric densities, defined by a countable mixture of normal densities, which can thus approximate any function. We have faced the empirical Bayes problem of estimating their parameters, using Dirichlet processes. Finally, on the basis of the axiomatization provided by Levesque et al. in (Shapiro, Pagnucco, Lesperance, and Levesque 2000), we have shown how beliefs can be used to compare two actions, as defined via countable mixtures.

Relation to work performed  The paper addresses the problem of modeling human actions observed by a robot in order to perform inference. The problem is relevant in the context of the research performed by WP5 as it introduces a complete framework to model actions from perception and it thus contributes to the NIFTi vision of natural human-robot communication.


Abstract The collaboration between a human and a robot is here understood as a learning process mediated by the instructor prompt behaviours and the apprentice collecting information from them to learn a plan. The instructor wears the Gaze Machine, a wearable device gathering and conveying visual and audio input from the instructor while executing a task. The robot, on the other hand, is eager to learn both the best sequence of actions, their timing and how they interlace. The cross relation among actions is specified both in terms of time intervals for their execution, and in terms of location in space to cope with the instruction interaction with people and objects in the scene. We outline this process: how to transform the rich information delivered by the Gaze Machine into a plan. Specifically, how to obtain a map of the instructor positions and his gaze position, via visual slam and gaze fixations; further, how to obtain an action map from the running commentaries and the topological maps and, finally, how to obtain a temporal net of the relevant actions that have been extracted. The learned structure is then managed by the flexible time paradigm of flexible planning in the Situation Calculus for execution monitoring and plan generation.

Relation to work performed The paper outlines a model of human robot collaboration in which the final goal is to learn the best actions needed to achieve the required goals, in this case, reporting hazards due to a crash accident in a tunnel, identifying the status of victims and, possibly, rescuing them. The collaboration is here viewed as a learning process involving the extraction of information from the instructor behaviours, thus providing data for skill and affordances learning. The instructor communicate his actions both visually (Using the GM) and with the aid of his comments delivered while executing the actions.


Abstract Asserting the inherent topology of the environment perceived by a robot is a key prerequisite of high-level decision making. This is achieved through the construction of a concise representation of the environment that endows a robot with the ability to operate in a coarse-to-fine strategy. In this paper, we propose a novel topological segmentation method of generic metric maps operating concurrently as a path-planning algorithm. First, we apply a Gaussian Distance Transform on the map that weighs points belonging to free space according to the proximity of the surrounding free area in a noise resilient mode. We define a region as the set of all the points that locally converge to a common point of maximum space clearance and employ a weighed mean-shift gradient ascent onto the kernel space clearance density in order to detect the maxima that characterize the regions. The spatial intra-connectivity of each cluster is ensured by allowing only for linearly unobstructed mean-shifts which in parallel serves as a path-planning algorithm by concatenating the consecutive mean-shift vectors of the convergence paths. Experiments on structured and unstructured environments demonstrate the effectiveness and potential of the proposed approach.

Relation to work performed The paper contributes a novel method to address topological segmentation of robot-acquired maps. The presented approach has supported the WP5 research on higher level reasoning by providing the stratum between the metric and functional maps.
2.13 P. Papadakis and F. Pirri “Consistent Pose Normalization of Non-Rigid Shapes using One-Class Support Vector Machines.” (NORDIA 2011)


Abstract The estimation of 3D surface correspondence constitutes a fundamental problem in shape matching and analysis applications. In the presence of non-rigid shape deformations, the ambiguity of surface correspondence increases together with the complexity of registration algorithms. In this paper, we alleviate this problem by using One-Class Support Vector Machines (OCSVM) in order to normalize the pose of 3D objects. We show how OCSVM are employed in order to increase the consistency of translation and scale normalization under articulations, extrusions or the presence of outliers. To estimate the relative translation and scale of an object, we use the 3D distribution of points that is modeled by employing OCSVM to estimate the decision surface corresponding to the surface points of the object. To evaluate the performance, we use a dataset of 3D objects where we introduce various extrusions, articulations or outliers and demonstrate the increased robustness of the proposed methodology.

Relation to work performed The paper addresses the problem of pose normalization, a fundamental step that needs to be accomplished in order to approach the more general problem of 3D object recognition. In particular, pose normalization represents a hard challenge when dealing with non-rigid and articulated objects. This research is related the scenario addressed by the NIFTi project, as it contributes to the development of the foreseen natural human-robot cooperation and, in particular, it contributes to the WP5 research on the detection of salient activities and the definition of communication interfaces.
References


EU FP7 NIFTi (ICT-247870) 35


