



UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA
**Dipartimento di Informatica, Sistemistica e
Comunicazione**
Corso di Laurea in Informatica

Percezione dell'Ambiente e Costruzione della Base di Conoscenza per Robot Umanoidi

Relatore: Prof. Dimitri Ognibene

Tesi di Laurea di:
Luca Brini
Matricola 879459

Anno Accademico 2023-2024

Indice

Introduzione	3
1 Stato dell'Arte	4
2 Mappe semantiche	5
2.1 Definizione	5
2.1.1 Nodi	6
2.1.2 Archi	8
2.1.3 Grafo stanze	8
2.1.4 Grafo Oggetti	8
2.2 Esempi ed utilizzi	9
3 Grafo di Scena	10
3.1 Generazione del Grafo di Scena	10
3.1.1 Inferenza	10
3.1.2 Coordinate oggetto	10
3.2 Aggiornamento del Grafo della Mappa	10
3.2.1 Stanza corrente robot	10
3.2.2 Proiezione del Camera Frustum	10
3.2.3 Controllo della posizione degli oggetti	10
3.2.4 Aggiornamento e salvataggio a DB	10
3.3 Conclusioni	10
4 Riconoscimento di Stanze	11
4.1 Analisi e risultati	11
4.2 Conclusioni	11
5 Analisi e Risultati	12
5.1 Errore della posizione degli oggetti	12
5.2 Punti di forza e svantaggi	12
5.2.1 Inferenza efficiente	12
5.2.2 Merging efficiente	12
6 Conclusioni e Sviluppi Futuri	13
6.1 Miglioramenti	13
6.1.1 Database a grafo	13
6.1.2 Finetuning OpenPSG	13
6.1.3 Object tracking	13
6.1.4 Utilizzo di OpenPVSG e Open4PSG	13

A	Appendice	14
A.1	RoBee System	14
A.1.1	Dashboard and Console	14
A.1.2	Infrastructure architecture, microservices and MQTT . . .	14
A.1.3	Maps, navigation and LiDaRs	14
A.1.4	Joints and transformations	14
A.1.5	Cameras and point cloud	14

Introduzione

Negli ultimi anni il campo della robotica ha vissuto un significativo incremento di applicazioni e innovazioni. Lo sviluppo di nuove tecnologie e la disponibilità di nuovi strumenti hanno reso possibile la creazione di robot in grado di svolgere compiti sempre più complessi. La **pianificazione automatica delle missioni** è sempre stata una delle attività di sviluppo in questo campo più affascinanti, pur essendo una delle più tediosa. Con l'avvento di ChatGPT e modelli simili, si è iniziato a pensare di integrare i **Large Language Models**, come alternativa ai classici planner, all'interno del sistema robot, con l'obiettivo di pianificare missioni autonome sulla base della descrizione in linguaggio naturale di ciò che si vuole far eseguire al robot.

La percezione dell'ambiente circostante è dunque una delle attività più importanti per un robot, soprattutto nell'ambito del **Mission Planning**. La capacità di riconoscere gli oggetti e di calcolarne la posizione è fondamentale per poterci interagire. Inoltre, è essenziale potersi localizzare nella mappa, sia in modo geometrico che topologico, in modo da poter pianificare anche eventuali movimenti verso gli oggetti desiderati che si trovano in punti non raggiungibili al momento dal robot.

In questo documento definiremo il significato di **Mappa Semantica**, le ragioni alla base della sua esistenza, la struttura e come viene utilizzata per pianificare le missioni del robot. Successivamente entreremo nel dettaglio del **Grafo di Scena**, come viene generato e tenuto aggiornato con i cambiamenti dell'ambiente. Infine analizzeremo il **Riconoscimento delle Stanze** a partire dalla mappa SLAM generata attraverso i sensori LiDaR del Robot, essenziale per suddividere l'insieme degli oggetti nelle loro stanze e gestire le missioni che necessitano lo spostamento in altre stanze.

Capitolo 1

Stato dell'Arte

Capitolo 2

Mappe semantiche

Gli esseri umani, talvolta senza rendersene conto, riescono a integrare continuamente nuove informazioni riguardo l'ambiente che li circonda, sia esso una casa, un edificio pubblico o un parco. Questa capacità, conscia e inconscia, è essenziale per la successiva pianificazione di obiettivi o movimenti poichè basati sulle informazioni appena apprese.

Così come per gli essere umani, anche i robot hanno bisogno di informazioni per poter essere considerati "cognitivi" e pianificare rispetto alla propria base di conoscenza. In particolar modo quando l'obiettivo è pianificare missioni data la descrizione in linguaggio in naturale di ciò che il robot deve fare, come in questo caso.

Esempio Consideriamo una persona che entra per la prima volta in una biblioteca. Egli osserva scaffali pieni di libri, tavoli per lo studio e un'area dedicata ai computer. Queste informazioni vengono immagazzinate e utilizzate successivamente per trovare un libro specifico o un luogo tranquillo per studiare.

Allo stesso modo, immaginiamo un robot progettato per operare in una casa intelligente. Riceve l'istruzione: "Prendi il libro dal tavolo del soggiorno e portalo in cucina." Per svolgere questo compito, il robot deve comprendere la struttura della casa, identificare il tavolo corretto e navigare verso la cucina.

2.1 Definizione

La Mappa Semantica è di fatto un grafo orientato che rappresenta questa base di conoscenza dove:

- Un nodo può essere un:
 - Nodo stanza
 - Nodo oggetto
 - Nodo tag
- Un arco può rappresentare:
 - La relazione tra due oggetti
 - Il collegamento tra due stanze

- L'appartenenza di un oggetto o un tag ad una ed una sola stanza

Di conseguenza, per trovare gli oggetti o i nodi appartenenti ad una stanza x è sufficiente considerare il sottografo del nodo stanza x

2.1.1 Nodi

Nodi oggetto

I nodi oggetto rappresentano gli oggetti riconosciuti all'interno dell'ambiente attraverso la segmentazione panoptica del feed video proveniente dalle camere del robot. Ogni nodo oggetto ha i seguenti attributi:

- Identificativo: utilizzato per identificare un oggetto all'interno della Mappa Semantica
- Nome: label inferita dal modello di segmentazione panoptica
- Posizione: terna (x, y, z) rappresentante la posizione dell'oggetto all'interno dell'ambiente rispetto alla Reference Posizione
- Reference Posizione: origine del sistema di riferimento delle posizioni degli oggetti. Può essere l'origine del sistema Mappa o l'origine del sistema Robot
- Tipo: rappresenta la tipologia dell'oggetto che può essere scelta tra:
 - Pickable: qualora l'oggetto possa essere preso attraverso gli end effectors del robot
 - Non Pickable: qualora l'oggetto non possa essere preso attraverso gli end effectors del robot
 - Asset: in tutti gli altri casi

Il modello è stato addestrato su dataset COCO, dunque gli oggetti riconoscibili sono i soli che appartengono a questo dataset.

I nodi oggetto vengono aggiornati con le inferenze di nuovi frame video: possono dunque essere eliminati dalla mappa semantica qualora un oggetto non si presenti più all'interno della scena oppure aggiornati, per esempio a livello di posizione, qualora l'oggetto venga spostato.

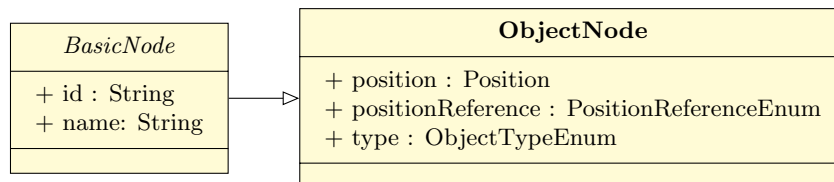


Figura 2.1: Diagramma delle classi - Nodo Oggetto

Nodi tag

I nodi tag rappresentano tutti quei particolari riferimenti che vengono utilizzati dal robot per localizzarsi o localizzare oggetti di particolare rilevanza (come la stazione di ricarica o il tavolo di lavoro). Ogni nodo tag ha i seguenti attributi:

- Identificativo: utilizzato per identificare il tag all'interno della Mappa Semantica
- Nome: assegnato dall'utente
- Posizione: terna (x, y, z) rappresentante la posizione del tag all'interno dell'ambiente rispetto all'origine della mappa
- Dimensione: dimensione del tag in millimetri
- Di Navigazione: flag che indica se il tag è utilizzato per la navigazione del robot
- Per Picking: flag che indica se il tag è utilizzato per identificare un oggetto di cui fare il picking con gli end effectors

I nodi tag sono permanenti all'interno della mappa semantica perché si assume che questi non vengano mai spostati o rimossi dall'ambiente del robot

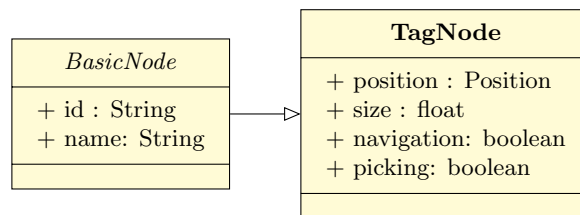


Figura 2.2: Diagramma delle classi - Nodo Tag

Nodi Stanza

I nodi stanza rappresentano le stanze all'interno della mappa slam generata attraverso i sensori LiDaR del robot. Queste stanze vengono individuate grazie a un algoritmo apposito. La presenza delle stanze nella mappa è fondamentale per diverse ragioni:

- Permette di suddividere gli oggetti in stanze, facilitando la discriminazione degli omonimi in base alla stanza di appartenenza.
- Consente l'utilizzo di algoritmi di ricerca su grafo per la pianificazione del percorso per raggiungere gli oggetti.
- Permette di creare percorsi pianificati che evitano determinate stanze.

Ogni nodo stanza ha i seguenti attributi:

- Identificativo: utilizzato per identificare la stanza all'interno della Mappa Semantica

- Nome: assegnato dall'utente
- Segmenti: Lista di segmenti che delimitano il poligono della stanza. Viene usato per verificare se un oggetto appartiene ad una stanza o no
- Oggetti: Sottografo degli oggetti appartenenti alla stanza

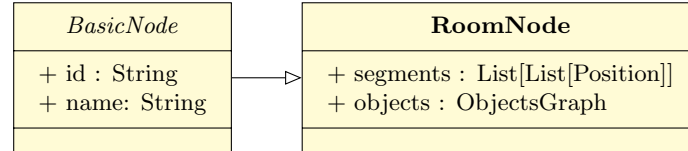


Figura 2.3: Diagramma delle classi - Nodo Stanza

2.1.2 Archi

Archi tra stanze

Come detto in precedenza, gli archi tra i nodi di tipo stanza rappresentano il collegamento diretto tra due stanze. Questi sono molto importanti poichè in futuro permetteranno di verificare quali stanze sono raggiungibili da una stanza x e pianificare di conseguenza.

2.1.3 Grafo stanze

Definiamo il grafo delle stanze come il sottografo (V_s, E_s) tale che:

- $V_s = \{v \in V_m \mid v \text{ è un nodo stanza}\}$
- $E_s = \{(v, u) \in E_m \mid v \in V_s \wedge u \in V_s\}$

dove V_m è l'insieme dei vertici del grafo della Mappa Semantica e E_m è l'insieme degli archi del grafo della Mappa Semantica.

Questo grafo, viene generato a partire dall'algoritmo di riconoscimento delle stanze, illustrato nel Capitolo 4, ed è la prima parte di Mappa Semantica creata. Solo successivamente sarà possibile la costruzione del grafo degli oggetti poichè ad ogni oggetto va associata una stanza.

2.1.4 Grafo Oggetti

Definiamo il grafo degli oggetti (o grafo di scena) come il sottografo (V_o, E_o) tale che:

- $V_o = \{v \in V_m \mid v \text{ è un nodo oggetto o tag}\}$
- $E_o = \{(v, u) \in E_m \mid v \in V_o \wedge u \in V_o\}$

dove V_m è l'insieme dei vertici del grafo della Mappa Semantica e E_m è l'insieme degli archi del grafo della Mappa Semantica.

Una versione grezza del grafo degli oggetti viene generato dal modello OpenP-SG di **yang2022psg**. Successivamente, come verrà mostrato nel Capitolo 3, questa verrà fusa con la Mappa Semantica applicando rendendo dunque la base di conoscenza coerente rispetto l'ambiente.

2.2 Esempi ed utilizzi

Esempio di Mappa Semantica

La figura di seguito mostra un esempio di mappa semantica.

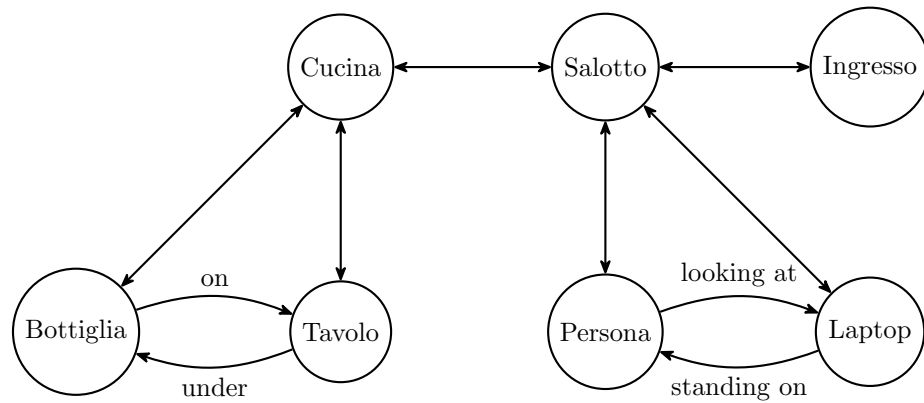


Figura 2.4: Esempio di Mappa Semantica

Capitolo 3

Grafo di Scena

3.1 Generazione del Grafo di Scena

3.1.1 Inferenza

3.1.2 Coordinate oggetto

3.2 Aggiornamento del Grafo della Mappa

3.2.1 Stanza corrente robot

3.2.2 Proiezione del Camera Frustum

3.2.3 Controllo della posizione degli oggetti

3.2.4 Aggiornamento e salvataggio a DB

3.3 Conclusioni

Capitolo 4

Riconoscimento di Stanze

4.1 Analisi e risultati

4.2 Conclusioni

Capitolo 5

Analisi e Risultati

5.1 Errore della posizione degli oggetti

5.2 Punti di forza e svantaggi

5.2.1 Inferenza efficiente

5.2.2 Merging efficiente

Capitolo 6

Conclusioni e Sviluppi Futuri

6.1 Miglioramenti

6.1.1 Database a grafo

6.1.2 Finetuning OpenPSG

6.1.3 Object tracking

6.1.4 Utilizzo di OpenPVSG e Open4PSG

Appendice A

Appendice

A.1 RoBee System

A.1.1 Dashboard and Console

A.1.2 Infrastructure architecture, microservices and MQTT

A.1.3 Maps, navigation and LiDaRs

A.1.4 Joints and transformations

A.1.5 Cameras and point cloud