

POLITECNICO DI MILANO
DIPARTIMENTO DI ELETTRONICA, INFORMAZIONE E BIOINGEGNERIA
CORSO DI LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA



Cognitive SLAM: knowledge-based self-localization and mapping

AI & R Lab
Laboratorio di Intelligenza Artificiale
e Robotica del Politecnico di Milano

Relatore:
Prof. Andrea Bonarini

Tesi di Laurea di:
Davide Tateo Matricola n. 799311

ANNO ACCADEMICO 2013-2014

A qualcuno...

Sommario

Lo SLAM è uno dei principali problemi nello sviluppo di robot autonomi. Gli approcci correnti sono afflitti da un pesante costo computazionale e si comportano male in ambienti dinamici e disordinati; le performance di questi sistemi diventano ancora peggiori se si utilizzano sensori a basso costo, necessari per le applicazioni commerciali. Questa tesi affronta il problema da un punto di vista nuovo e originale, usando feature ad alto livello come punti chiave e sfruttando la conoscenza di un esperto e un linguaggio fuzzy per riconoscerli, per tenere traccia di punti chiave forti e stabili e permettere mappe più intelligenti e una localizzazione robusta in ambienti complessi. L'idea fondamentale è quella di mantenere il tasso di errore della localizzazione limitato e ridurre il costo necessario a far navigare con successo un robot autonomo in un ambiente interno, usando solo i dati provenienti da una webcam e una unità di misura inerziale a basso costo. Il principale problema è il riconoscimento di feature ad alto livello, come porte e scaffali, affrontato tramite un classificatore fuzzy ad albero, definito da un esperto, in modo da evitare fasi di allenamento e migliorare la generalità del riconoscimento.

Abstract

SLAM is one of the key issues in autonomous robots development. Current approaches are affected by heavy computational load and misbehave in cluttered and dynamic environment; their performance get even worse with low cost sensors, needed for market applications. This thesis faces the problem in a new and original way, working with high level features as key points and using expert knowledge and a fuzzy language to detect them, in order to track strong and stable key points and allow smarter maps and robust localization in complex environments. The key idea is to keep the error rate of the localization process limited, and to reduce the cost of an autonomous robot to successfully navigate into an indoor environment using only the data from a webcam and a low cost Inertial measurement unit. The main issue is the high level feature recognition, like doors and shelves, done by an expert-defined fuzzy tree classifier, in order to avoid training and improve generalization of recognition.

Indice

Sommario	I
Abstract	III
Ringraziamenti	VII
1 Introduzione	1
2 Stato dell'arte	3
2.1 SLAM	3
2.2 Rappresentazione del Mondo	5
2.3 Riconoscimento di Oggetti	6
2.4 Reasoning	7
3 Concetti Fondamentali	9
4 Architettura del Sistema	11
4.1 Introduzione	11
4.2 Diagramma del sistema	12
4.3 Comunicazione tra i nodi	12
4.4 Parametri degli algoritmi	14
5 Reasoning	15
6 Riconoscimento degli Oggetti	17
7 Tracking	19
8 Mapping	21
9 Risultati sperimentali	23
10 Conclusioni e Sviluppi Futuri	25

Bibliografia	27
A Il manuale utente	31
B Esempio di impiego	33

Ringraziamenti

Ringrazio

Capitolo 1

Introduzione

“Qualunque cosa che accade, accade”

“Qualunque cosa che, accadendo, ne fa accadere un'altra, ne fa accadere un'altra.”

“Qualunque cosa che, accadendo, induce se stessa a riaccadere, riaccade.”

“Però non è detto che lo faccia in ordine cronologico.”

Douglas Adams, Mostly Harmless

Uno dei problemi chiave della robotica è la localizzazione di un robot in un ambiente sconosciuto. Questo problema è noto come “SLAM”, Simultaneous localization and Mapping, ed è attualmente una delle aree più importanti della ricerca riguardante i robot autonomi. In particolare, questo problema risulta ancora più difficile se applicato a robot dotati di sensori a basso costo, quali webcam e unità di misura inerziale economiche, restrizione fondamentale per la diffusione di applicazioni di robotica autonoma nel mondo reale.

Attualmente questo problema è affrontato attraverso tecniche probabilistiche, che sono ottime per analizzare ambienti statici e per ottenere una localizzazione precisa. Le mappe create sono utilizzabili principalmente per la navigazione. I sensori più usati sono gli scanner laser, che tuttavia sono molto costosi, i sensori RGB-D, sensori economici che nascono per applicazioni videoludiche commerciali e hanno alcune limitazioni se applicati alla robotica, le telecamere stereo.

Lo scopo della tesi è di sviluppare un framework per risolvere il problema della localizzazione di robot autonomi dotati di sensori a basso costo, quali ad esempio i quadricotteri. L'idea è di sviluppare una metodologia che non solo permetta al robot di localizzarsi nella mappa in maniera efficiente, ma anche di interagire con l'ambiente in maniera avanzata, di compiere ragionamenti, di adattarsi a eventuali cambiamenti ed agli eventi che possono

occorrere, quali la presenza di persone o altri agenti.

Basandosi principalmente sulle informazioni provenienti da una videocamera monoculare, è stato sviluppato un sistema che è in grado di processare informazioni provenienti da un esperto, espresse in un linguaggio formale, di estrarre feature di basso livello e aggregarle, grazie alla base di conoscenza, per riconoscere e tracciare feature di alto livello, quali porte e armadi, e di basare conseguentemente su di essi il processo di localizzazione. Per permettere all'esperto di trasmettere la sua conoscenza all'agente, è stato sviluppato un linguaggio formale, basato sulla logica fuzzy, che permetta di esprimere sia regole fuzzy linguistiche, sia di definire un classificatore fuzzy ad albero in grado di definire una gerarchia di modelli e le relazioni tra di essi.

La tesi è strutturata nel modo seguente:

Nel capitolo 2 si illustra lo stato dell'arte.

Nel capitolo 3 si illustrano le basi teoriche necessarie.

Nel capitolo 4 si descrive l'architettura generale del sistema.

Nel capitolo 5 si parla del reasoner e del linguaggio formale utilizzato.

Nel capitolo 6 si descrive l'estrazione delle feature di basso livello.

Nel capitolo 7 si mostra come gli oggetti riconosciuti sono tracciati.

Nel capitolo 8 si illustra la creazione della mappa e la localizzazione.

Nel capitolo 9 si analizzano i risultati sperimentali del sistema proposto.

Nel capitolo 10 si riassumono gli scopi, le valutazioni di questi e le prospettive future.

Capitolo 2

Stato dell'arte

Doc: Ecco perché non ha funzionato: c'è scritto "Made in Japan".

Marty: E che vuol dire Doc? Tutta la roba migliore è fatta in Giappone.

Doc: Incredibile!

Ritorno al Futuro, parte III

2.1 SLAM

Il problema della localizzazione di un robot e costruzione della mappa in contemporanea in un ambiente sconosciuto è stato affrontato fin dagli anni '90 a partire da [20], articolo nel quale per la prima volta si delineava un framework per localizzare un robot costruendo contemporaneamente la mappa dell'ambiente.

Tramite l'utilizzo di sonar, venivano estratte feature geometriche con cui veniva costruita la mappa, nella quale il robot si localizzava. Il problema principale della localizzazione è il "problema della correlazione": se la posizione della feature rispetto alla quale ci si localizza è affetta da incertezza, la conseguente stima della posizione effettuata rispetto a tale feature sarà affetta da un errore che dipende dall'errore della posizione della feature stessa. Questo problema diventa tanto più grave se si pensa che la posizione del robot in ogni istante non è nota a priori, ma deve essere stimata sulla base delle osservazioni precedenti.

È necessario risolvere questo problema per evitare che l'errore della generazione della mappa e l'errore della stima della posizione divergano nel tempo. Per risolverlo, gli autori hanno spesso utilizzato un filtro di Kalman esteso.

Il filtro di Kalman è uno stimatore Bayesiano ricorsivo, che, supposto noto il modello lineare che regola la generazione dei dati e la loro osservazione, supposto che l'errore di misura e di modello siano gaussiani, restituisce la densità di probabilità del sistema osservato. Il filtro di Kalman, se utilizzato secondo le ipotesi, è uno stimatore ottimo dello stato del sistema osservato, secondo i minimi quadrati. Tuttavia, nell'ambito della robotica, e in particolare nel problema della localizzazione, il modello di generazione e osservazione dei dati non può essere considerato lineare. E' quindi necessario utilizzare un'estensione del filtro di Kalman al caso non lineare: il filtro di Kalman esteso (EKF) è una delle possibili soluzioni al problema. L'idea alla base del filtro di Kalman esteso è quella di lavorare sul modello linearizzato, stimato ricorsivamente dal modello non lineare sulla base della stima corrente.

Per avere una buona stima della posizione è necessario utilizzare un gran numero di feature, numero che cresce molto rapidamente con l'aumentare della dimensione dell'ambiente. La complessità computazionale dell'approccio tradizionale basato sul filtro di Kalman esteso è $\mathcal{O}(N^3)$, con N numero di feature, e quindi il tempo di calcolo diventa ben presto inaccettabile per prestazioni in tempo reale. Per risolvere questo problema è stato introdotto in [23] un nuovo algoritmo detto FastSLAM, che consiste nell'utilizzo del Particle Filter, e del filtro di Kalman esteso in combinazione. L'algoritmo associa ad ogni feature considerata, un filtro di Kalman esteso; la densità di probabilità congiunta, invece, viene calcolata sfruttando il Particle filter. Il Particle Filter è un altro stimatore Bayesiano ricorsivo, che, invece di un modello e dell'assunzione di rumore gaussiano, sfrutta metodi di tipo Monte Carlo per stimare la densità di probabilità del sistema che genera i dati. Il risultato è un algoritmo che ha complessità computazionale $\mathcal{O}(N \log M)$, con M numero di feature e N il numero di particelle usate dal Particle Filter. Questo approccio rende il problema trattabile nella maggior parte dei casi, pur essendo pesante computazionalmente, dato che è necessario un elevato numero di particelle per avere una buona localizzazione.

Vista la particolarità del problema quando il sensore utilizzato è una videocamera monoculare, sono stati sviluppati algoritmi ad hoc. Uno degli algoritmi più usati è PTAM [16], [17]. L'idea alla base di questo algoritmo è dividere in due thread separati il tracking e la creazione della mappa: un thread si occupa del tracking robusto di feature a basso livello, mentre l'altro thread si occupa della creazione della mappa. Per rendere efficiente il processo di mapping, solo i keyframe, ossia i frame che contengono maggiore informazione rispetto a quella già presente, vengono considerati. Per rendere il processo di mapping robusto, vengono utilizzate tecniche batch per

costruire la mappa, come ad esempio il bundle adjustment. Il bundle adjustment consiste in un processo iterativo di raffinamento della stima dei punti 3D ricostruiti e della posa della videocamera. PTAM, tuttavia, nasce per applicazioni di realtà aumentata, e quindi necessita di una inizializzazione, per risolvere i problemi dell'acquisizione del primo keyframe e per gestire la scala della mappa.

Un approccio alternativo consiste nell'usare tutti i dati dell'immagine per eseguire la localizzazione, questo approccio è alla base, ad esempio, di DTAM, Dense Tracking and Mapping [24]. Questo algoritmo crea un modello denso dell'ambiente e usa l'allineamento della videocamera

Si sono dimostrati efficaci anche i metodi semi-diretti, come SVO [8], Semi Direct Visual Odometry, un algoritmo che riesce a ottenere altissime prestazioni limitando l'estrazione delle feature ad alto livello ai soli keyframe, operando direttamente sulle intensità dei pixel nei frame successivi, eliminando le fasi computazionalmente più onerose, che sono l'estrazione e l'abbinamento delle feature. SVO si basa sulle idee di PTAM, ma ne migliora sia le prestazioni, riuscendo a essere computazionalmente più leggero, sia la precisione della mappa e della localizzazione riducendo di molto gli outlier.

Recentemente, stanno avendo molto successo i sistemi basati su sensori RGB-D [14], [9]. Questi sensori vengono utilizzati come scanner laser a basso costo per creare una mappa dell'ambiente. Tuttavia, sono soggetti a molte limitazioni, non essendo stati progettati per questo scopo, e soffrono tra l'altro di un raggio d'azione limitato. Nonostante queste limitazioni i sistemi riescono comunque a ottenere buone prestazioni in ambienti indoor [30].

2.2 Rappresentazione del Mondo

Sono noti diversi modi per rappresentare un ambiente tridimensionale. Il più semplice possibile è quello di usare delle nuvole di punti, direttamente estratte dai sensori. Un'altra rappresentazione comune è quella di filtrare le nuvole di punti ottenute tramite una griglia di voxel, come, ad esempio, in [29]. Metodi più avanzati permettono una rappresentazione geometrica dell'ambiente con un minor uso di memoria, come, ad esempio, le mappe di quota, in cui una mappa a due dimensioni è estesa con il calcolo del valore medio di altezza di ciascun punto 2D [10], le mappe di quota estese [27], che tengono conto di possibili aperture attraversabili dai robot, oppure le mappe di quota multi livello [31], che riescono a descrivere complesse geometrie multi livello.

Tra le mappe più promettenti esistono le mappe basate sugli octree, che permettono un'efficiente rappresentazione in memoria sia dello spazio occupato sia dello spazio libero, pur potendo descrivere geometrie complesse. Inoltre, questo tipo di rappresentazione permette di scalare facilmente la risoluzione della mappa, permettendo di utilizzare la stessa mappa per compiti che richiedono una precisione differente. Un'efficiente implementazione di questo tipo di mappa può essere trovata in [13].

Recentemente, si sta sviluppando l'idea di rappresentare il mondo ad alto livello, incorporando informazioni semantiche e geometriche che possano essere utilizzate non solo dagli algoritmi di navigazione, ma anche per svolgere compiti ad alto livello e ragionamenti. Una possibile soluzione al problema è l'approccio basato su Scene Graph [2]. Sono stati sviluppati anche linguaggi specifici di dominio (DSL) per poter interagire ad alto livello, ad esempio generando istanze di oggetti a partire da un modello generico, come ad esempio in [3].

2.3 Riconoscimento di Oggetti

La quasi totalità degli algoritmi di riconoscimento di oggetti nell'immagine è basata su tecniche di machine learning. Una delle classi di algoritmi più usati sono quelli basati sulle Haar-like feature [26]. Queste feature sono calcolate in aree rettangolari, all'interno delle quali vengono sommati i valori dell'intensità dei pixel. Le somme sono in seguito usate per calcolare differenze tra aree di interesse, per poter riconoscere feature geometriche quali angoli, linee o bordi. Un efficiente algoritmo per il riconoscimento di oggetti è descritto in [32] ed è stato ampliato in [21], e consiste nell'utilizzare in cascata una serie di classificatori via via più restrittivi; ognuno dei classificatori di ogni stadio è costruito attraverso una tecnica di boosting, ossia sono formati da un insieme di classificatori deboli che creano un classificatore complessivo più restrittivo. I classificatori di base sono solitamente alberi di decisione. Il classificatore complessivo dà una risposta binaria; per riconoscere effettivamente l'oggetto deve essere applicato tramite una finestra mobile su tutta l'immagine.

Un algoritmo simile è descritto in [1], dove, però, invece che feature Haar-like, vengono utilizzati direttamente un insieme di pixel selezionati nell'immagine.

Un'altra importante classe di algoritmi è basata sulle feature HOG (Histogram of Oriented Gradients) [5].

Queste feature sono ricavate contando le occorrenze dell'orientamento dei gradienti in sotto-blocchi dell'immagine. Su queste feature si basa l'al-

goritmo descritto in [7]. Questo algoritmo è in grado di definire un oggetto a partire dalle sue parti, e quindi è robusto anche a occlusioni parziali dell'oggetto. Tuttavia è computazionalmente molto più pesante dell'approccio basato su feature Haar-like.

Più recenti sono i metodi basati sul deep learning. Tra questi, i più promettenti nel riconoscimento di oggetti sono basati sulle reti neurali convoluzionali. La particolarità di queste reti è di essere basata su kernel di nodi collegati in maniera fissa, e ripetuti per coprire tutta l'immagine. La struttura rigida sfrutta la località dell'informazione nell'immagine, e permette un training più efficiente. Un esempio di applicazione si può trovare in [18].

Metodi recenti, sfruttano sensori RGB-D, come ad esempio Kinect, per riconoscere oggetti nella scena, come ad esempio in [19]. La maggior parte delle tecniche usate sono un'estensione delle tecniche già descritte, spesso estendendo i descrittori delle feature grazie alle informazioni sulla profondità.

2.4 Reasoning

I sistemi di ragionamento sono stati alla base dell'Intelligenza Artificiale fin dalla nascita ed hanno trovato sbocco in diverse applicazioni tra cui i sistemi esperti, per lungo tempo una delle branche dell'intelligenza artificiale più attive [4] [6] [15].

Con la maturazione della tecnologia per lo sviluppo dei sistemi esperti, e l'avvento di Internet, l'attenzione della ricerca collegata ai sistemi di ragionamento formale si è spostata dai sistemi esperti al web semantico. Il linguaggio più diffuso per i sistemi di inferenza è il linguaggio OWL 2 [25], [11], fortemente basato sulle logiche descrittive che permettono un buon compromesso tra espressività, e complessità computazionale e mantengono la logica utilizzata decidibile.

Fin dai primi sistemi esperti si è riconosciuta l'importanza del trattamento dell'incertezza in sistemi di ragionamento che avessero a che fare con il mondo reale [6], ed in particolare si è avuta una larga diffusione in moltissimi settori della logica fuzzy [33], adottata anche nel nostro lavoro.

Un esempio di reasoning applicato al riconoscimento delle immagini può essere trovato in [22], in cui feature a basso livello vengono estratte da un meccanismo di machine learning, e utilizzate da un sistema di inferenza che sfrutta una ontologia per l'analisi ad alto livello dell'immagine. Un altro esempio di utilizzo di sistemi di inferenza nel riconoscimento di oggetti si può trovare in [12], dove viene utilizzato il linguaggio OWL per specificare

una base di conoscenza e riconoscere tramite la conoscenza di dominio gli oggetti nell'ambiente.

Capitolo 3

Concetti Fondamentali

Capitolo 4

Architettura del Sistema

4.1 Introduzione

In questo capitolo descriveremo l'architettura del sistema implementato. Il sistema è stato progettato in maniera modulare per una serie di motivi: questo tipo di architettura permette di riutilizzare i moduli sviluppati, di estendere semplicemente il sistema con altri moduli, di sostituire eventualmente un intero modulo con un altro che compia le stesse funzioni, di comunicare in maniera semplice con altri sistemi.

Per raggiungere questo scopo, si è deciso quindi di utilizzare il middleware ROS (Robot Operating System) [28]. ROS è un middleware che, oltre ad offrire molti tool utili per lo sviluppo di complesse applicazioni robotiche, implementa due pattern architetturali importanti, entrambi usati nel nostro lavoro: il pattern publish-subscribe e il pattern client-server. Questi due pattern sono implementati rispettivamente tramite messaggi e servizi: i messaggi definiscono un formato comune per la pubblicazione di dati in topic, i servizi invece specificano un'interfaccia per le chiamate a procedure remote, dichiarando gli input e gli output tra client e server. Ogni processo che viene eseguito in ROS è chiamato nodo. Per eseguire qualsiasi sistema basato su ros, devono essere presenti tre ulteriori nodi: il nodo Master, che si occupa di garantire la comunicazione tra gli altri nodi, tramite i due pattern supportati, il server dei parametri, che implementa un dizionario condiviso accessibile tramite la rete, in cui i nodi possono memorizzare e recuperare

parametri usati dai loro algoritmi a runtime, il nodo `rosout`, che si occupa di mantenere i log prodotti dalle applicazioni.

Il sistema sviluppato è pensato per interagire con qualsiasi tipo di robot che sia provvisto di una videocamera monoculare e una unità di misura inerziale (IMU) tramite le due interfacce standard di ROS. Esse consistono in tre topic: “`imu`”, per i dati provenienti dall’unità di misura inerziale, “`image_raw`”, per l’immagine proveniente dalla videocamera, “`camera_info`”, che contiene i parametri intrinseci della videocamera, tra cui la matrice di calibrazione della videocamera utilizzata, necessaria per il funzionamento del sistema. I dati estratti dall’unità di misura inerziale devono essere riferiti al sistema di coordinate della videocamera, infatti l’algoritmo di visione implementato utilizza i dati provenienti dalla IMU per stimare approssimativamente la posa della telecamera.

4.2 Diagramma del sistema

L’architettura del sistema implementato è descritta in Figura 4.1.

Il diagramma rappresenta i nodi del sistema, segue una breve descrizione di ogni elemento per specificare la loro funzione:

`image_proc` si occupa di eliminare la distorsione radiale della videocamera causata dalla curvatura della lente.

`c_vision_detector` si occupa di estrarre possibili feature dall’immagine.

`c_tracking` si occupa di seguire le feature a basso livello estratte dal nodo “`c_vision_detector`” nell’immagine, mantenendo un modello in modo da riuscire a riconoscere la feature anche dopo essere stata persa.

`c_vision_slam` si occupa dell’analisi approfondita delle feature, in modo da determinare il tipo di oggetto e la sua posizione nello spazio.

`c_fuzzy_reasoner` implementa un reasoner fuzzy; data una base di conoscenza e un classificatore, analizza le feature in ingresso e le classifica.

4.3 Comunicazione tra i nodi

Questo sistema sfrutta sia il paradigma client-server, per l’interazione con il reasoner, che il paradigma publish-subscribe, in tutti gli altri casi. Il reasoner offre 4 servizi:

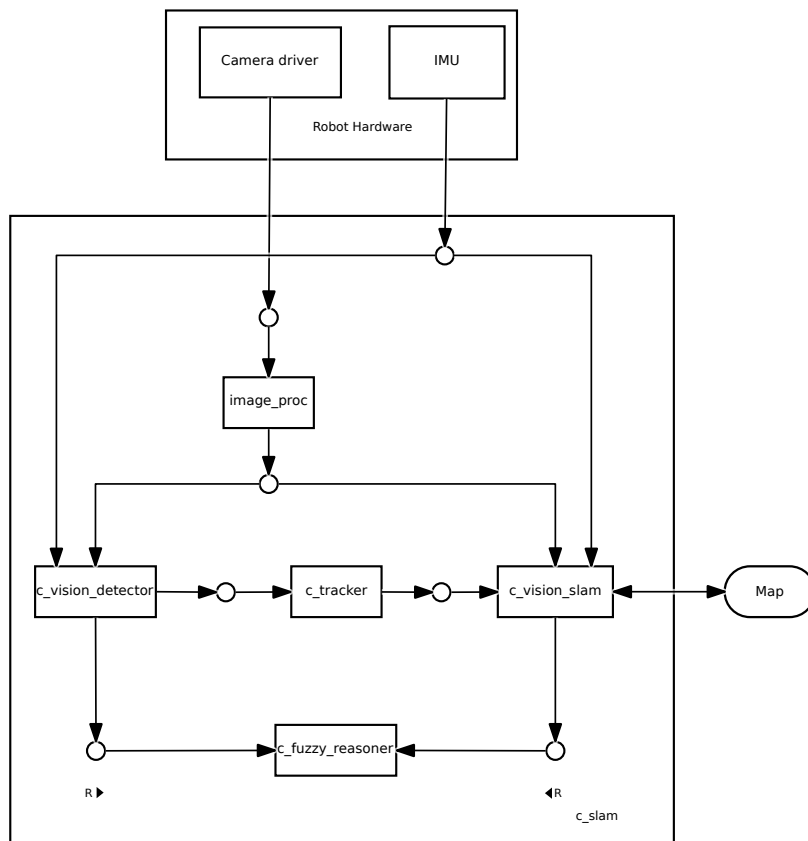


Figura 4.1: Architettura del sistema implementato

reasoning questo servizio è un normale servizio di reasoning basato su una knowledgebase fuzzy.

classification questo servizio si occupa di classificare istanze di features riconosciute.

getDependencyGraph questo servizio restituisce il grafo delle dipendenze del classificatore.

getReasoningGraph questo servizio restituisce il grafo di reasoning utilizzato dal classificatore.

Il reasoner e i suoi servizi verranno discussi approfonditamente nel Capitolo 5.

le informazioni riguardanti gli oggetti riconosciuti e seguiti dal sistema sono scambiate tramite i seguenti topic:

to_track in questo topic vengono pubblicate le possibili feature riconosciute dall'intera immagine. Viene fornito il contorno della feature e la sua possibile classificazione.

tracks in questo topic vengono pubblicati i risultati dell'algoritmo di tracking: contiene il bounding box e il contorno della feature, quest'ultimo viene maggiorato del 20%, per assicurarsi di mantenere all'interno di esso il reale contorno della feature.

Il riconoscimento di features viene descritto nel Capitolo 6, il tracking delle feature riconosciute nel Capitolo 7, e infine l'analisi ad alto livello viene effettuata nel Capitolo 8.

4.4 Parametri degli algoritmi

Gli algoritmi utilizzati nel sistema, in particolare gli algoritmi di visione, hanno alcuni parametri che è possibile tarare per adattare il sistema a qualsiasi tipo di robot utilizzato. Per gestire i parametri abbiamo utilizzato il server dei parametri di ROS. Il server dei parametri è un dizionario, ossia ogni parametro, identificato da una stringa, è salvato nel server dei parametri, e può essere recuperato tramite il nome. Il server dei parametri supporta dizionari gerarchici, in modo da poter rappresentare tipi di dati strutturati. Inoltre è possibile definire parametri privati per qualsiasi nodo. I parametri privati restano accessibili da tutto il resto del sistema, ma nel dizionario vengono salvati sotto il nome del nodo a cui appartengono; questo comportamento è studiato per evitare collisioni tra parametri con lo stesso nome in nodi differenti.

Visto che gli algoritmi di estrazione e analisi delle features si basano fortemente sugli stessi strumenti, questo sistema usa in maniera estesa i parametri privati, in modo da avere lo stesso parametro, che rappresenta concettualmente la stessa quantità, differente in base al nodo che lo utilizza.

Il sistema permette anche di cambiare i parametri a runtime, essi vengono aggiornati nei nodi che ne fanno uso con una frequenza fissa. Si possono conoscere i parametri usati dal sistema e il loro significato nell'Appendice A.

Capitolo 5

Reasoning

Marty: Aspetta un momento Doc. Se vado dritto verso lo schermo, andrò a sbattere contro quegli indiani!

Doc: Marty, non stai pensando quadridimensionalmente!

Ritorno al Futuro, parte III

Capitolo 6

Riconoscimento degli Oggetti

Capitolo 7

Tracking

Capitolo 8

Mapping

Capitolo 9

Risultati sperimentali

Capitolo 10

Conclusioni e Sviluppi Futuri

Bibliografia

- [1] Yotam Abramson, Bruno Steux, and Hicham Ghorayeb. Yet even faster (yef) real-time object detection. *IJISTA*, 2(2/3):102–112, 2007.
- [2] S. Blumenthal, H. Bruyninckx, W. Nowak, and E. Prassler. A scene graph based shared 3d world model for robotic applications. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 453–460, May 2013.
- [3] Sebastian Blumenthal and Herman Bruyninckx. Towards a domain specific language for a scene graph based robotic world model. *arXiv preprint arXiv:1408.0200*, 2014.
- [4] Bruce G Buchanan, Edward Hance Shortliffe, et al. *Rule-based expert systems*, volume 3. Addison-Wesley Reading, MA, 1984.
- [5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, June 2005.
- [6] Randall Davis, Bruce Buchanan, and Edward Shortliffe. Production rules as a representation for a knowledge-based consultation program. *Artificial intelligence*, 8(1):15–45, 1977.
- [7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.
- [8] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. SVO: Fast semi-direct monocular visual odometry. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

- [9] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research*, 31(5):647–663, 2012.
- [10] M Herbert, C Caillas, Eric Krotkov, In So Kweon, and Takeo Kanade. Terrain mapping for a roving planetary explorer. In *Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on*, pages 997–1002. IEEE, 1989.
- [11] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph, editors. *OWL 2 Web Ontology Language: Primer*. W3C Recommendation, 27 October 2009. Available at <http://www.w3.org/TR/owl2-primer/>.
- [12] Helmi Ben Hmida, Christophe Cruz, Frank Boochs, and Christophe Nicolle. Knowledge base approach for 3d objects detection in point clouds using 3d processing and specialists knowledge. *CoRR*, abs/1301.4991, 2013.
- [13] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 2013. Software available at <http://octomap.github.com>.
- [14] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011.
- [15] Gary D. Riley Joseph C. Giarratano. *Expert Systems Principles and Programming*. 1994.
- [16] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’07)*, Nara, Japan, November 2007.
- [17] Georg Klein and David Murray. Improving the agility of keyframe-based SLAM. In *Proc. 10th European Conference on Computer Vision (ECCV’08)*, pages 802–815, Marseille, October 2008.

- [18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [19] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Detection-based object labeling in 3d scenes. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1330–1337. IEEE, 2012.
- [20] J.J. Leonard and H.F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*, pages 1442–1447 vol.3, Nov 1991.
- [21] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, pages 900–903, 2002.
- [22] Nicolas Maillot and Monique Thonnat. Ontology based complex object recognition. *Image Vision Comput.*, 26(1):102–113, 2008.
- [23] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002. AAAI.
- [24] Richard A. Newcombe, Steven Lovegrove, and Andrew J. Davison. Dtam: Dense tracking and mapping in real-time. In Dimitris N. Metaxas, Long Quan, Alberto Sanfeliu, and Luc J. Van Gool, editors, *ICCV*, pages 2320–2327. IEEE, 2011.
- [25] W3C OWL Working Group. *OWL 2 Web Ontology Language: Document Overview*. W3C Recommendation, 27 October 2009. Available at <http://www.w3.org/TR/owl2-overview/>.
- [26] C.P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Computer Vision, 1998. Sixth International Conference on*, pages 555–562, Jan 1998.
- [27] Patrick Pfaff, Rudolph Triebel, and Wolfram Burgard. An efficient extension to elevation maps for outdoor terrain mapping and loop closing. *The International Journal of Robotics Research*, 26(2):217–230, 2007.

- [28] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5, 2009.
- [29] Y. Roth-Tabak and R. Jain. Building an environment model using depth information. *Computer*, 22(6):85–90, June 1989.
- [30] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 573–580. IEEE, 2012.
- [31] R. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop closing. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 2276–2282, Oct 2006.
- [32] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. pages 511–518, 2001.
- [33] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

Appendice A

Il manuale utente

Manuale utente per l'utilizzo del sistema

Appendice B

Esempio di impiego

Un esempio di impiego del sistema realizzato.