

# Project log - Robotica

Augello Andrea

Castiglione Francesco Paolo

La Martina Marco

2 gennaio 2021

## Indice

<b>1 Setup</b>	<b>1</b>
<b>2 Nome</b>	<b>1</b>
<b>3 Ambiente</b>	<b>2</b>
<b>4 ROS</b>	<b>2</b>
4.1 Bug . . . . .	2
<b>5 Dipendenze</b>	<b>2</b>
<b>6 Obbiettivo</b>	<b>2</b>
<b>7 TIAGo Iron</b>	<b>2</b>
<b>8 Modello del moto e posizionamento</b>	<b>3</b>
<b>9 Object recognition</b>	<b>3</b>
9.1 Campionamento delle immagini . . . . .	3
9.2 Yolo . . . . .	3
9.3 Scarto dei duplicati . . . . .	4
<b>10 Posizione dei target</b>	<b>4</b>
10.1 Triangolazione . . . . .	4
10.2 Calcolo della distanza . . . . .	5
10.3 Modello probabilistico . . . . .	6
<b>11 Pianificazione del moto</b>	<b>9</b>
11.1 Modalità di movimento . . . . .	9
11.1.1 Modalità esplorazione . . . . .	9
11.1.2 Bug mode . . . . .	9
11.1.3 Campi di potenziale . . . . .	10
11.2 Pianificazione . . . . .	10

## 1 Setup

OS	Ubuntu 18.04 Ubuntu 20.04
ROS version	melodic noetic
Webots	R2020b revision 1
Target hardware	Raspberry Pi 4B Raspberry Pi 3B+

## 2 Nome

Il team ha scelto il nome **Change** in onore di **Chang'e 4** [1], la missione parte della seconda fase del programma cinese di esplorazione lunare, durante il quale è andato a buon fine il primo atterraggio morbido sulla faccia nascosta della luna.

## 3 Ambiente

Abbiamo considerato opportuno analizzare e studiare il package **webots\_ros** [2] al fine di raggiungere una comprensione più profonda sulle metodologie per interfacciare i nodi ROS con il controller ROS standard per Webots. Inoltre è risultato necessario approfondire la documentazione ROS [3] al fine di installare e configurare l'ambiente ROS ed inoltre per capire i concetti fondamentali relativi ai nodi e topics. Infine abbiamo impostato l'interfaccia ROS su Webots seguendo la documentazione cyberbotics rilevante [3].

## 4 ROS

### 4.1 Bug

Logical values did not allow callbacks.

## 5 Dipendenze

La seguente è una lista delle librerie utilizzate nel nostro progetto ed una breve spiegazione della loro funzione e rilevanza:

- opencv 4.x, una libreria per la computer vision, usata per operazioni di segmentazione [4];
- imutils, che include funzioni per semplici operazioni di image processing quali traslazioni, rotazioni, ridimensionamento. Utilizzato inoltre per effettuare Non Maxima Suppression(NMS) [5];
- sklearn, una libreria per il machine learning comprendente algoritmi di clustering quali DBSCAN [6];
- numpy, una libreria che fornisce supporto per array multidimensionali, matrici ed operazioni matematiche per lavorare su detti array [7];
- matplotlib, una libreria per creare visualizzazioni di dati (statiche, dinamiche, interattive) [8];
- math, una libreria che fornisce funzioni matematiche definite dallo standard C [9].

## 6 Obiettivo

L'obiettivo del robot è di **evitare assembramenti in ambienti indoor**.

Nella dimostrazione presentata il nostro robot rileva le persone nella stanza e individua i possibili assembramenti. In seguito alla fase di rilevazione si sposterà verso l'assembramento evitando gli ostacoli e, arrivato, esorterà le persone al rispetto del distanziamento sociale.

## 7 TIAGo Iron

Il robot scelto per l'obiettivo proposto è il **TIAGo Iron**.

Il **PAL Robotics TIAGo Iron** [10] è un robot umanoide a due ruote con torso e testa ma senza braccia articolate. Il modello è una piattaforma modulare mobile che permette l'interazione fra esseri umani e robot.

Il datasheet del **TIAGo** indica la presenza di speaker e display, tuttavia questi non sono presenti nel modello Webots. Abbiamo dunque ritenuto necessario per il nostro scopo aggiungere uno **speaker** e un **display** con corrispondente solido di supporto al modello. La camera del **TIAGo**, come indicato dal datasheet, è RGB-D. Il modello Webots ne è sprovvisto, di conseguenza è stata utilizzata una camera monoscopica RGB. Inoltre è stato necessario contattare gli sviluppatori del **TIAGo** per chiedere informazioni circa le dimensioni esatte delle **ruote** in quanto tale informazione è omessa dal datasheet. Ci è stato comunicato che le ruote del **TIAGo** hanno raggio di  $200\text{ mm}$ . Utilizzando tale valore nei calcoli odometrici, abbiamo ottenuto valori largamente differenti dalle misurazioni. Abbiamo quindi dedotto sperimentalmente che il raggio del modello Webots è di  $3\text{ cm}$  più lungo.

L'IMU utilizzata ha 6 gradi di libertà ed è composta delle seguenti componenti:

1. giroscopio;
2. accelerometro;

Abbiamo ritenuto non necessario aggiungere il **magnetometro** in quanto in uno scenario reale sarebbe stato soggetto ad interferenze (significativamente più di un giroscopio), specialmente in un ambiente con molti oggetti metallici (quale potenzialmente lo scenario di utilizzo del nostro robot).

Il **lidar** del **TIAGo**, come indicato nel datasheet [11], ha un range di  $5.6\text{ m}$  ed un FOV di  $240^\circ$  (ricordiamo che agli estremi è parzialmente occluso).

## 8 Modello del moto e posizionamento

Il modello del moto è caratterizzato da rotazioni e traslazioni. Per ottenere l'angolo di rotazione ci basiamo sui dati ottenuti dal giroscopio, il quale misura il moto rotazionale fornendo una velocità angolare. Per ottenere l'angolo di rotazione effettuiamo quindi un'integrazione discreta dei campioni con interpolazione lineare del primo ordine. Per effettuare lo spostamento lineare utilizziamo il controllore PID (Proporzionale-Integrale-Derivativo) delle ruote fornito da Webots, che richiede l'angolo di rotazione corrente, il diametro delle ruote e fornisce l'angolo di rotazione necessario al fine di ottenere lo spostamento desiderato.

$$targetAngle = currentAngle + 2\pi \frac{distance}{2\pi \cdot diameter} \quad (1)$$

A causa di possibili imprecisioni calcoliamo la stima dello spostamento lineare utilizzando l'accelerometro. Al segnale dell'accelerometro viene applicato un integrale doppio per ottenere lo spostamento lineare.

Nell'immagine seguente viene mostrata la zona nella quale, se viene indicata dal **lidar** la presenza di un ostacolo, il **TIAGo** si ferma per ragioni di sicurezza al fine di evitare danni a persone e/o oggetti.

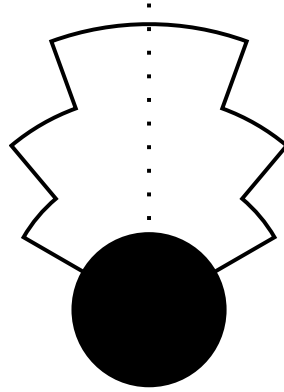


Figura 1: Collision detection

## 9 Object recognition

### 9.1 Campionamento delle immagini

Il FOV della camera è di  $57^\circ$ , di conseguenza per ricoprire  $360^\circ$  è stato necessario effettuare 7 campionamenti. Il settimo campionamento, come si evince dalla figura 2, è sovrapposto al primo per una porzione di scena pari a  $39^\circ$  coincidente col primo campionamento.

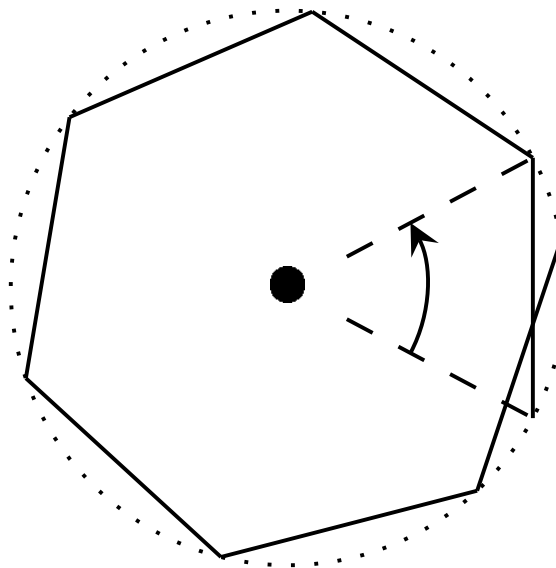


Figura 2: Campionamento delle immagini

### 9.2 Yolo

Al fine di riconoscere le persone è stato necessario utilizzare sistemi di **object recognition**. A tal fine abbiamo valutato le performance di YOLOv3 (you only look once), YOLOv3-tiny, HoG (Histogram of oriented gradients), HoG + SVG

(support vector machines) + NMS (non maxima suppression). In seguito a vari test su HoG abbiamo ritenuto essere problematica la larghezza delle bounding boxes fornite, in quanto, per motivazioni che verranno chiarite nel paragrafo successivo, vogliamo che queste ultime siano il più possibili vicine alla reale larghezza delle persone. YOLOv3, nonostante sia stato addestrato su foto di persone reali (e non modelli 3D) fornisce risultati soddisfacenti, in seguito al finetuning degli iperparametri parametri della rete. Tuttavia, considerando le caratteristiche hardware del robot mobile, abbiamo optato per l'uso di YOLOv3-tiny, il quale risulta essere significativamente più efficiente, sacrificando in termini di precisione ma comunque sufficientemente preciso per il nostro obiettivo. Ecco un paragone fra YOLOv3 e YOLOv3-tiny in termini di mAP (mean average precision) e FLOPS (floating-point operations per second), come illustrato dalla tabella seguente, i cui dati provengono dal sito di YOLO [12]:

Model	mAP	FLOPS	FPS
YOLOv3-320	51.5	38.97 Bn	45
YOLOv3-416	55.3	65.86 Bn	35
YOLOv3-608	57.9	140.69 Bn	20
YOLOv3-tiny	33.1	5.56 Bn	220
YOLOv3-spp	60.6	141.45 Bn	20

### 9.3 Scarto dei duplicati

Durante la fase di individuazione delle persone vengono individuate varie bounding box corrispondenti al medesimo individuo. Di conseguenza è stato necessario effettuare una fase di clustering al fine di scartare le bounding box duplicate. L'algoritmo di clustering utilizzato è DBSCAN (Density based scan) [13], i cui parametri principali sono **eps**, ovvero la massima distanza fra due punti affinché vengano considerati appartenenti a un cluster (da non confondere con la massima distanza fra i punti di un cluster), **min\_samples**, ovvero il numero minimo di punti affinché un cluster sia valido (nel nostro caso è uguale a 1 in quanto non vogliamo scartare ROI) ed infine la metrica di distanza. Come di evince dalla figura 3, la metrica utilizzata considera la lunghezza di un arco di circonferenza con raggio corrispondente alla distanza fra i due punti ed angolo  $\alpha$  corrispondente all'angolo fra i due punti rispetto alla posizione del robot.

Abbiamo ritenuto opportuno utilizzare l'implementazione dell'algoritmo fornita da **sklearn** [6].

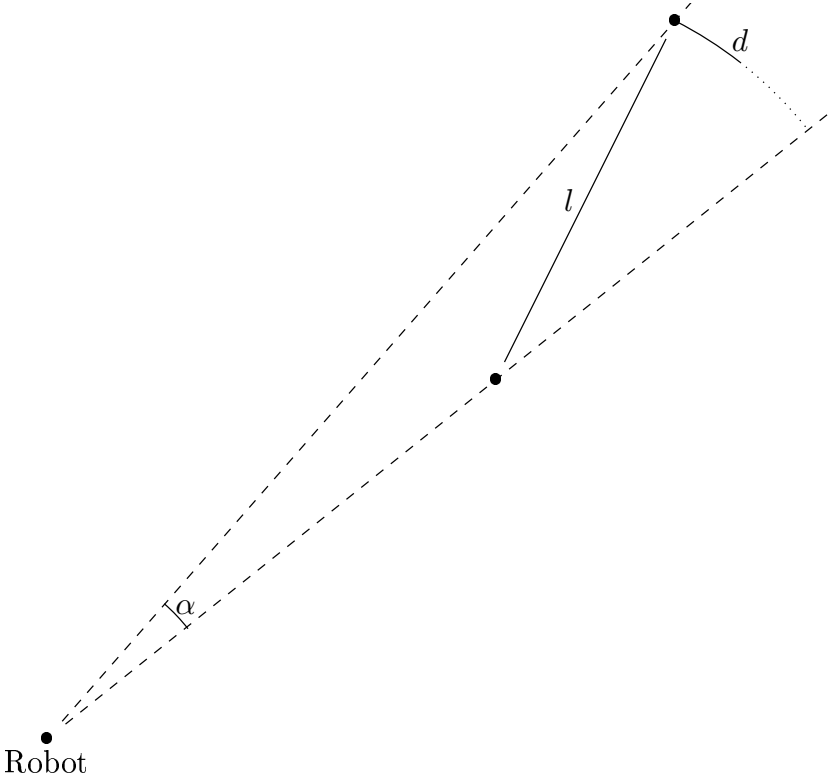


Figura 3: Non maxima suppression

## 10 Posizione dei target

### 10.1 Triangolazione

La triangolazione come metodo di individuazione delle persone, sebbene teoricamente possibile, presenta dei problemi nel nostro scenario. In primo luogo prendiamo in esame l'occlusione delle persone. Ad esempio, se due persone (5 e 2) sono una dietro l'altra lungo una retta immaginaria che le congiunge al robot (B), quest'ultimo non sarà in grado di individuare la persona 5. Inoltre, quando il robot effettua scan successivi, non sarebbe in grado di dedurre quali osservazioni derivano

dalla stessa persona. Infatti, poiché nel nostro scenario abbiamo più persone in una stanza, non è possibile dedurre quali intersezioni delle rette corrispondono ad osservazioni reali o se si tratta di intersezioni spurie.

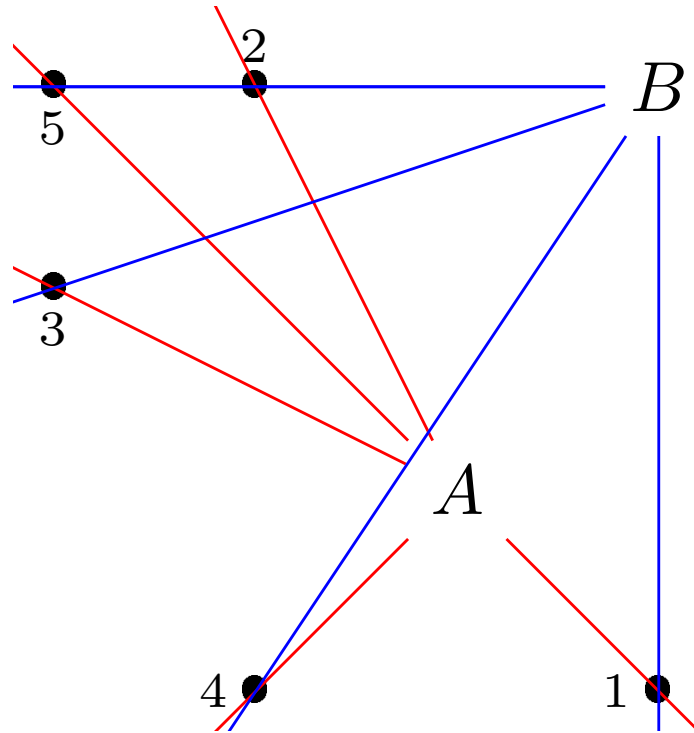


Figura 4: Triangolazione

## 10.2 Calcolo della distanza

L'altezza percepita dell'oggetto non è un indicatore affidabile della sua distanza dal robot in quanto parte dell'oggetto potrebbe essere occlusa o non presente nel frame. Inoltre classificatori quali HoG tendono a produrre ROI significativamente più alte dell'oggetto. La larghezza del torso, invece, è meno suscettibile a tali problemi, e non dipende dalla posizione (ad esempio seduto o alzato). Dobbiamo tuttavia ipotizzare che il torso abbia forma cilindrica introducendo quindi degli errori se l'obbiettivo non sta guardando la camera. La posizione orizzontale dell'oggetto relativa alla camera può influenzare la larghezza percepita diminuendola quando aumenta la distanza dal centro dell'immagine. Ipotizzando che la camera abbia un FOV (field of view) di  $2\alpha$  e sia distante  $d$  dall'oggetto, la massima distanza orizzontale che un punto dell'immagine potrebbe avere dal centro del piano dell'immagine sarebbe  $a = d \tan \alpha$  (Fig. 5).

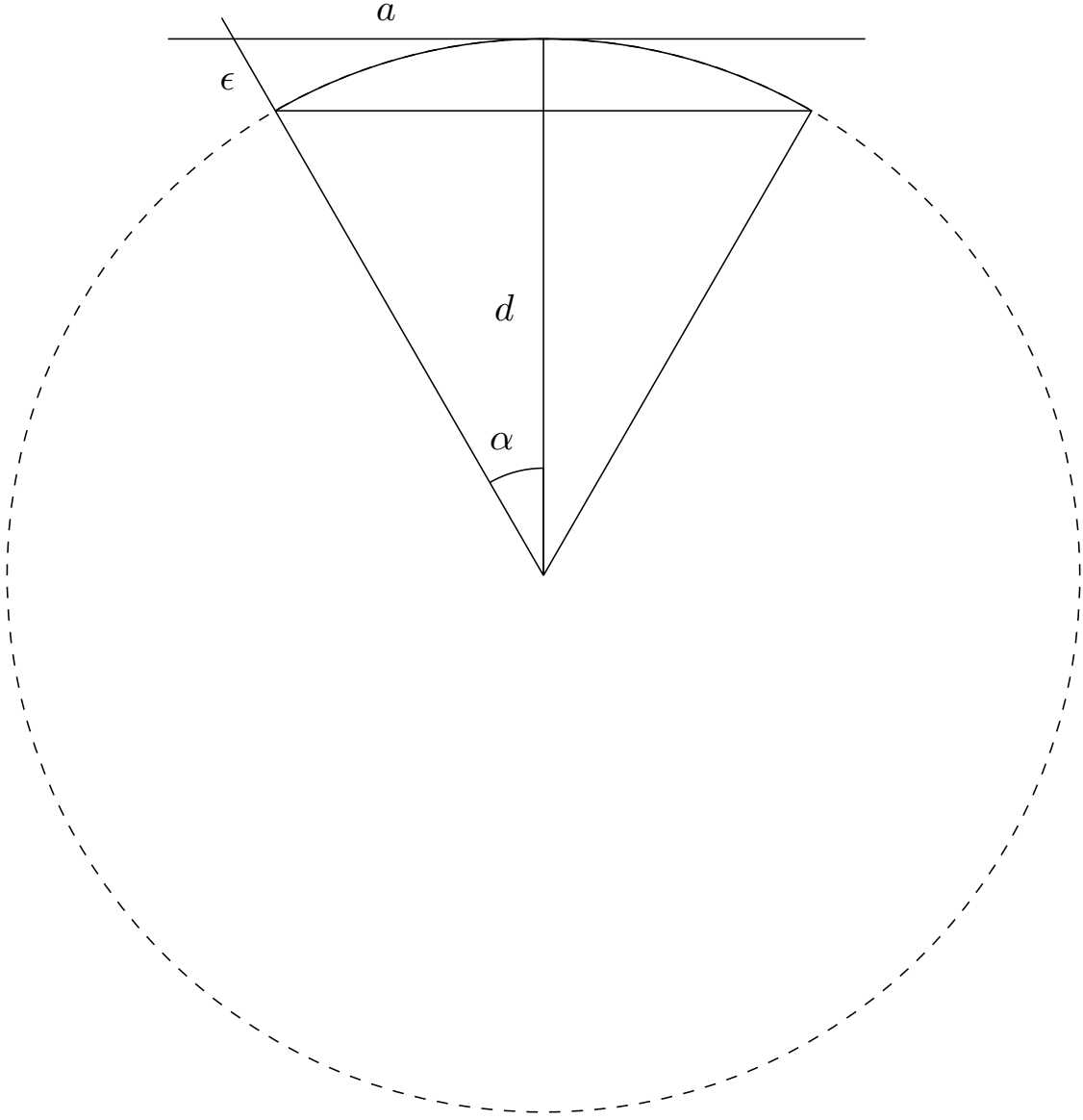


Figura 5: Errore nella stima della distanza con una linearizzazione a tratti della circonferenza

Ignorare la prospettiva significa effettuare un'approssimazione lineare del primo ordine e trattare il punto come se si trovasse su una circonferenza di raggio  $d$  centrata sulla camera. Di conseguenza consideriamo il punto come se fosse più vicino di quanto non sia realmente, commettendo l'errore mostrato nell' Eq. 2. Con una camera con FOV di 1 radiante quale quella del TIAGo il massimo errore causato dalla linearizzazione è quindi una sottostima del 13.9%.

$$\epsilon = \sqrt{a^2 + d^2} - d = \sqrt{(d \tan \theta)^2 + d^2} - d = d \left( \sqrt{\frac{1}{\cos^2 \alpha}} - 1 \right) = d (\sec \alpha - 1) \quad (2)$$

Sotto tali ipotesi possiamo quindi calcolare la distanza di un oggetto come mostrato in Eq. 3

$$object\ distance(m) = \frac{f(m) \times real\ width(m) \times image\ width(pixels)}{object\ width(pixels) \times sensor\ width(m)} \quad (3)$$

Poiché stiamo utilizzando un simulatore non è nota la larghezza del sensore da utilizzare per l'eq. 3. Abbiamo ovviato a tale problema posizionando il robot ed un oggetto dalle dimensioni note in posizioni note e abbiamo utilizzato questi dati insieme a delle misure in pixel nell' eq. 4. Abbiamo così stimato le dimensioni del sensore virtuale da utilizzare nei calcoli successivi.

$$sensor\ width(m) = \frac{f(m) \times real\ width(m) \times image\ width(pixels)}{object\ width(pixels) \times object\ distance(m)} \quad (4)$$

### 10.3 Modello probabilistico

Le relazioni geometriche sulla distanza degli oggetti esposte nella sezione 10.2 assumono che il target abbia la stessa larghezza in ogni posa e che sia possibile ottenere dal classificatore delle bounding box estremamente aderenti all'oggetto. Queste assunzioni nella nostra applicazione non sono rispettate, la distanza dell'oggetto sarà quindi soggetta ad errore

non trascurabile. Per questa ragione e per i problemi legati alla triangolazione abbiamo abbandonato la rappresentazione basata su oggetti e siamo ricorsi ad un filtro di occupazione bayesiano (BOF) [14].

**TODO: Inserire un paragrafo di collegamento tra BOF e dettagli implementativi**

Al fine di trasformare le osservazioni ottenute in una probabilità che le persone si trovino effettivamente nella posizione indicata è stato necessario definire una funzione densità di probabilità. La distribuzione normale, sebbene relativamente appropriata nel nostro scenario, è computazionalmente onerosa. La griglia di occupazione ha diverse migliaia di celle per cui va calcolata una pdf per ogni osservazione. Come si evince dalla fig. 6 con un numero così elevato di iterazioni i tempi di esecuzione non sarebbero accettabili, è stato quindi necessario ricorrere ad una approssimazione.

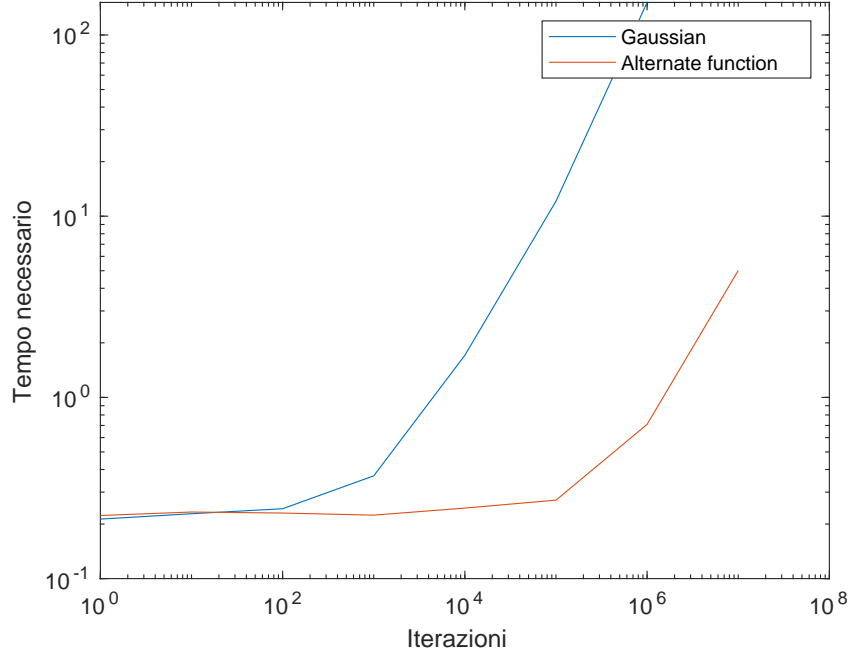


Figura 6: Benchmark funzione densità di probabilità (probability density function, funzione )

Un'approssimazione di largo uso è la distribuzione triangolare. Quest'ultima, tuttavia, presenta delle caratteristiche non desiderabili per il nostro caso d'uso. Una proprietà desiderabile della nostra funzione, difatti, sono le "fat tails", ovvero le code della distribuzione devono essere spesse e la distribuzione non deve tagliare nettamente. Quando, in seguito ad un'osservazione, non abbiamo alcuna osservazione di conferma nello scan successivo, non è infatti desiderabile che la probabilità di trovare una persona in quel punto scenda a zero. Inoltre, in una prima osservazione, un oggetto potrebbe essere occluso ed essere rilevato solo successivamente, quindi se in quella zona la probabilità di rilevare un target scendesse a zero non sarebbe possibile integrare correttamente questa nuova informazione. Ulteriori complicazioni sorgerebbero in caso di target in movimento.

Essendo la scansione una operazione estremamente costosa non è nemmeno possibile affidarsi esclusivamente all'aggiunta di rumore alla griglia di occupazione confidando in una eventuale convergenza.

Al fine di ottenere un'approssimazione di una gaussiana adatta al nostro scenario, è stato utilizzato un funzionale ispirato al guadagno del filtro di Butterworth.

$$p(\mathbf{x}|\mathbf{z}) = \frac{K}{1 + d(\mathbf{x}, \mathbf{z})^4} \quad (5)$$

Nell'equazione 5 è mostrato il funzionale utilizzato. In figura 7 si mostra un confronto della forma di questa funzione rispetto ad una gaussiana con media nulla e deviazione standard unitaria nel caso monodimensionale.

Nell'equazione 5  $K$  indica un parametro di scala per ottenere CDF unitaria.  $d$  è una funzione per il calcolo della distanza della cella  $\mathbf{x}$  dalla stima di posizione  $\mathbf{z}$ .

La distanza euclidea porterebbe a formare aree ad alta probabilità di forma circolare. Questo però non si adatta bene al nostro modello di errore del sensore: l'angolo dell'oggetto rispetto al robot è noto con una precisione molto elevata, mentre la maggior parte dell'incertezza si concentra nella distanza. Utilizziamo quindi la distanza di Mahalanobis per tenere conto del nostro modello.

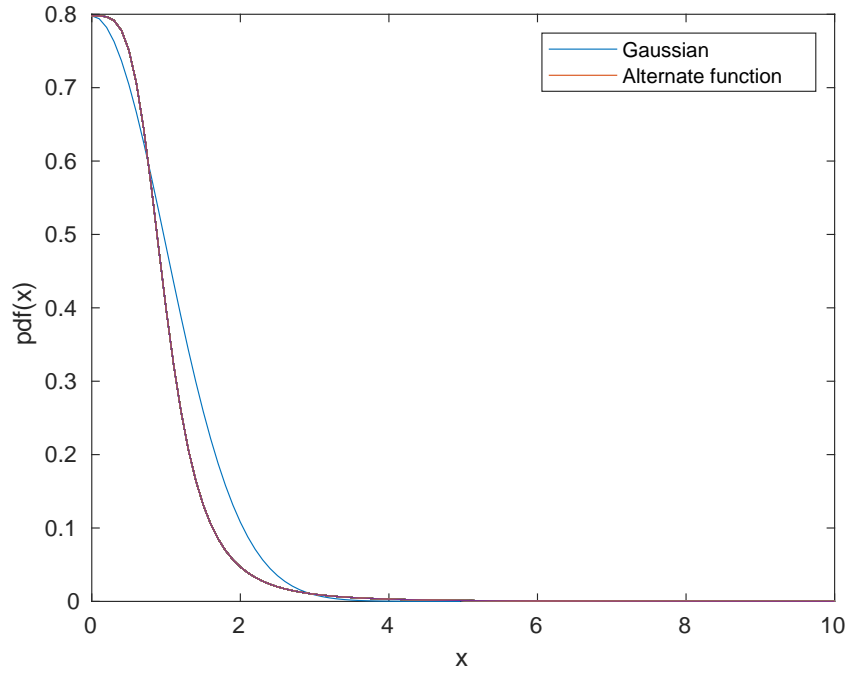


Figura 7: Forma pdf

Dato un insieme di osservazioni ottenute da una scansione  $Z_i$  aggiorniamo il belief precedente rispetto ad ogni cella della griglia di occupazione come mostrato nell'equazione 6.

TODO: check math to match slides

$$p(\mathbf{x}|Z_i) = p(\mathbf{x}|Z_{i-1}) \cdot \sum_{\mathbf{z} \in Z_i} p(\mathbf{x}|\mathbf{z}) \quad (6)$$

Al fine di individuare le zone con alta probabilità di contenere persone abbiamo utilizzato un approccio derivante dall'elaborazione delle immagini. In primo luogo separiamo i punti nella mappa in punti ad alta probabilità di essere occupati da persone e punti con bassa probabilità. Per effettuare questa sogliatura abbiamo utilizzato il metodo Otsu [15], il quale applica un thresholding automatico calcolato minimizzando la varianza dei valori di intensità all'interno delle classi e massimizzandola fra classi differenti all'immagine in input.

Da questa sogliatura otteniamo una mappa binaria in cui sono evidenziate le celle ad alta probabilità. Applicando l'algoritmo in [16], implementato in OpenCV, estraiamo le regioni ad alta probabilità contigue ed i loro contorni. Per ogni regione verrà selezionato come centro il punto con maggiore probabilità (Fig. 8).

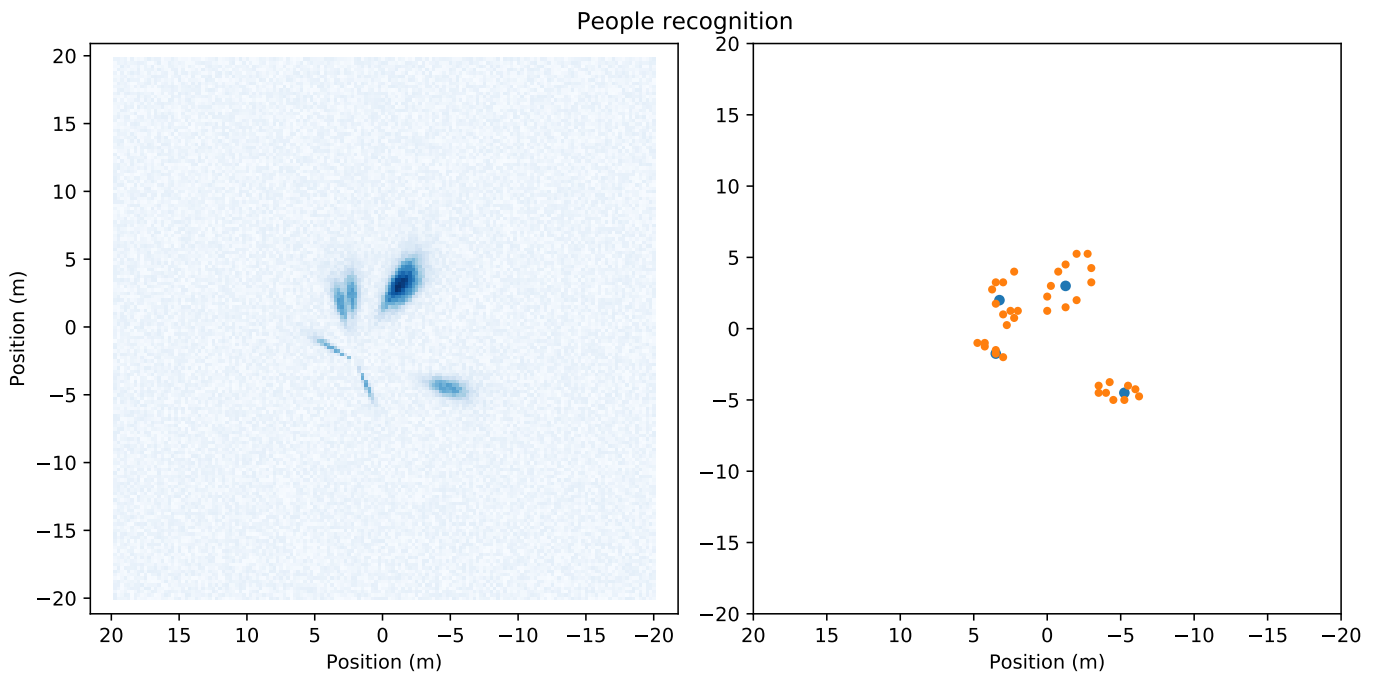


Figura 8: Segmentazione immagine: a sinistra i valori grezzi del BOF, a destra i contorni delle regioni (giallo) e i centri (blu)



## 11 Pianificazione del moto

### 11.1 Modalità di movimento

#### 11.1.1 Modalità esplorazione

Quando il robot entra in modalità esplorazione il suo comportamento è di esplorazione casuale della mappa. Il robot continua ad esplorare ed effettuare scan periodici fino a quando non viene rilevato un obiettivo. In tal caso il robot entra in modalità campi di potenziale. Se il robot incontra un ostacolo nel suo cammino ruota di  $90^\circ$  nella direzione opposta a quest'ultimo e continua l'esplorazione casuale.

#### 11.1.2 Bug mode

Quando il robot entra in modalità bug effettuiamo uno scan lidar. Poiché lavoriamo con valori discreti, confrontiamo i valori ottenuti con una soglia al fine di individuare le discontinuità nel profilo degli oggetti nel range. Analizziamo successivamente le discontinuità del segnale, corrispondenti ai bordi degli ostacoli, come si evince dalla fig. 9. Da tali punti viene calcolato l'angolo fra la retta che li congiunge con l'obiettivo. Il robot si muove infine verso il punto al quale corrisponde l'angolo minore, che ci farà allontanare meno dall'obiettivo. Se il profilo dell'ostacolo è una circonferenza, la retta fra il punto con angolo minimo e l'obiettivo è tangente all'ostacolo, da cui il nome dell'algoritmo Tangent Bug [17].

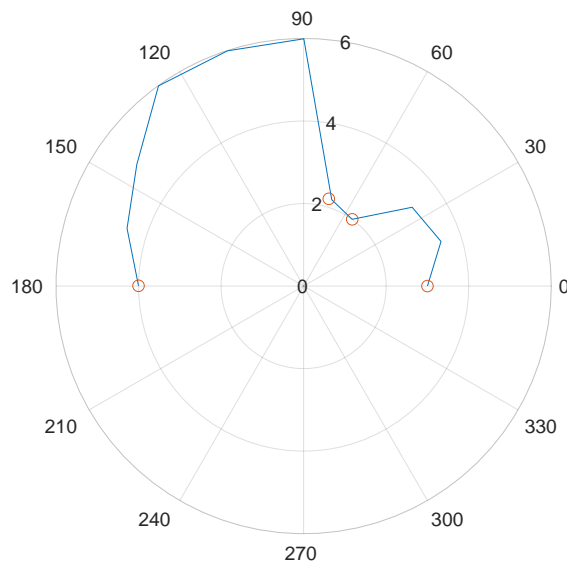


Figura 9: Profilo dell'ostacolo

### 11.1.3 Campi di potenziale

## 11.2 Pianificazione

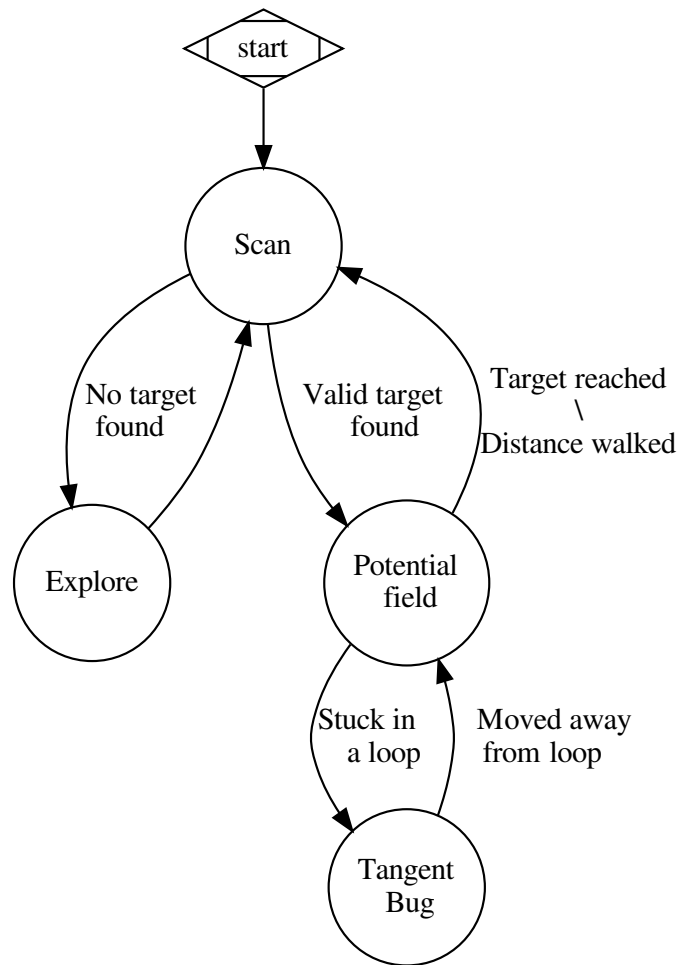


Figura 10: Automa a stati finito (Finite state automata)

## Riferimenti bibliografici

- [1] Hannah Devlin and Kate Lyons. Far side of the moon: China's chang'e 4 probe makes historic touchdown, Jan 2019.
- [2] Webots. Webots user guide: Tutorial 8: Using ros.
- [3] Ros documentation from ros.org.
- [4] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [5] A Rosebrock. Imutils.
- [6] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [7] Charles R. Harris, K. Jarrod Millman, St'efan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fern'andez del R'io, Mark Wiebe, Pearu Peterson, Pierre G'erard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with numpy. *Nature*, 585(7825):357–362, September 2020.
- [8] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [9] math - mathematical functions.
- [10] Webots. Webots user guide: Pal robotics' tiago iron.
- [11] Tiago datasheet.
- [12] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement. *arXiv*, 2018.
- [13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [14] MK Tay, Kamel Mekhnacha, Manuel Yguel, Christophe Coue, Cédric Pradalier, Christian Laugier, Th Fraichard, and Pierre Bessiere. The bayesian occupation filter. In *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, pages 77–98. Springer, 2008.
- [15] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [16] Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, 30(1):32–46, 1985.
- [17] I. Kamon, E. Rivlin, and E. Rimon. A new range-sensor based globally convergent navigation algorithm for mobile robots. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 1, pages 429–435 vol.1, 1996.