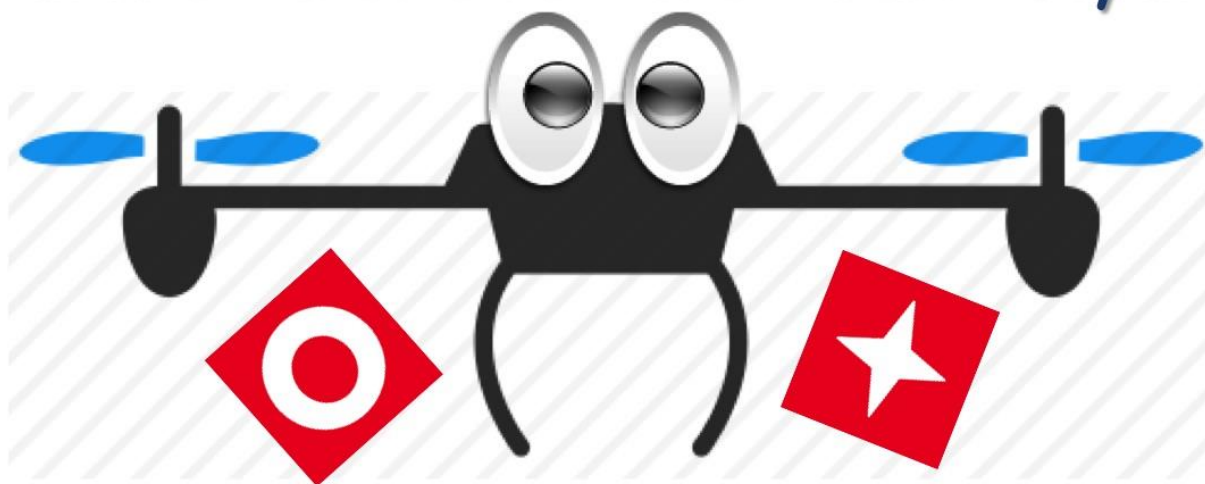


...We believe we can fly...



Drone Vision Cup 2015

Endorsed by



REGOLAMENTO

Il seguente documento formalizza le regole del contest Drone Vision Cup 2015, ideato e promosso dal MIVIA Lab, Laboratorio di Macchine Intelligenti per il riconoscimento di Immagini, Video e Audio, nell'ambito del corso di Modelli e Sistemi per la Visione Artificiale, anno accademico 2014-2015, della Laurea Magistrale in Ingegneria Informatica. Titolari di cattedra sono i docenti M. Vento e G. Percannella.

Descrizione generale

L'obiettivo del contest consiste nell'implementazione, attraverso metodi di visione artificiale, di un agente intelligente il cui task è quello di riconoscere due tipi di marker inquadrati dalla telecamera a bordo del drone. I marker sono posizionati in maniera sparsa all'interno di un'arena che ricopre un'area totale massima di 30x30 m. Al termine di ogni volo, il video registrato sarà reso disponibile a tutti i concorrenti, i quali potranno effettuare l'elaborazione specificando, per ogni marker identificato, il tipo e l'istante della detection. I risultati della video-analisi dovranno essere consegnati in un formato predefinito.

La competizione consta di tre sessioni, ognuna delle quali caratterizzata da un crescente livello di difficoltà, relativamente al percorso e alla velocità del drone, nonché per la presenza di eventuali oggetti distrattori di colore simile a quello dei target. Alla fine di ogni sessione, gli arbitri calcoleranno i punteggi ottenuti dai gruppi, con le modalità elencate nel seguito del presente documento.

Il gruppo che totalizzerà il punteggio maggiore si aggiudicherà la Drone Vision Cup 2015.

Specifiche tecniche

Per garantire ai concorrenti la possibilità di calibrare i propri algoritmi di video-analisi in maniera adeguata, è opportuno specificare una serie di vincoli che saranno rispettati durante la competizione.

In particolare vigono le seguenti norme:

- la dimensione massima dell'arena è di 30x30 m
- la dimensione massima dei marker è di 50x50 cm
- la dimensione minima dei marker è di 20x20 cm



- l'altezza massima raggiunta dal drone durante il volo è di 10 m
- la velocità massima di volo del drone è di 6 m/s
- la velocità minima di volo del drone è di 2 m/s
- i marker sono di due tipi, di colore bianco su sfondo rosso (Figura 1)

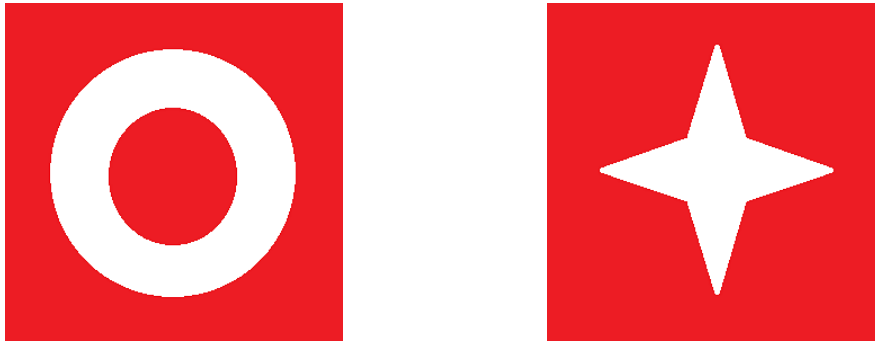


Figura 1 - Marker della competizione

Percorso

Il percorso che il drone seguirà durante il volo autonomo è specificato all'inizio di ciascuna sessione. Per ogni waypoint si specificherà l'altezza, il tempo di stazionamento e la velocità con cui il drone deve muoversi verso il waypoint successivo.

Specifiche sui video

Il video della singola sessione viene acquisito con frame rate costante pari a 25 fps con risoluzione 640x480. Siano t_0 l'istante temporale in secondi associato al primo frame del video e t_f quello associato all'ultimo frame del video; si definisce durata del video, misurata in secondi, la quantità $T = t_f - t_0$.

Note realizzative

Tutti i gruppi dovranno consegnare il proprio software eseguibile su piattaforma Linux senza alcuna dipendenza, se non quella relativa ad OpenCV. L'applicativo dovrà essere lanciato da linea di comando, specificando come parametro il path del video da elaborare.



Punteggio

La valutazione della qualità dell'algoritmo di visione artificiale viene effettuata utilizzando un indice detto **F-Score**. Questo indice prende in considerazione due elementi fondamentali: **Precision** e **Recall**. La Precision è il rapporto tra il numero di true positive generati dal sistema e il numero totale degli allarmi generati dal sistema; la Recall è il numero di true positive rilevati dal sistema e il numero di allarmi che il sistema avrebbe dovuto generare.

L'F-Score verrà quindi valutato nel seguente modo:

$$F - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Pertanto in questa competizione il punteggio parziale di ogni gruppo, ovvero il punteggio totalizzato nella singola sessione, è l'F-Score dell'algoritmo rispetto al task assegnato. Se l'applicativo termina la sua esecuzione prima che il video sia finito, al gruppo verrà attribuito un **valore nullo** dell'F-Score per la specifica sessione.

Al termine di ogni sessione, in base ai risultati conseguiti, saranno calcolati i valori di Precision e Recall

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

dove TP è il numero di true positive, FP è il numero di false positive (falsi allarmi generati dal sistema) e FN è il numero di false negative (allarmi che il sistema avrebbe dovuto generare), e si procederà con il calcolo dell'F-Score per ogni gruppo.

Il punteggio finale totalizzato da ogni gruppo sarà calcolato come media aritmetica tra gli F-Score ottenuti nelle tre sessioni della gara:

$$Punteggio = \frac{\sum_{i=1}^3 F - Score_i}{3}$$

Il gruppo che totalizzerà il punteggio più elevato sarà proclamato vincitore della Drone Vision Cup 2015.



Calcolo TP, FP e FN e penalizzazioni

Al termine di ogni sessione gli arbitri della competizione rivedranno il video acquisito dalla telecamera a bordo del drone e annoteranno, per ogni marker inquadrato, il tipo (cerchio o stella) e il secondo di ingresso nella scena. Il documento prodotto dagli arbitri costituirà la groundtruth della sessione, ovvero la tabella con cui saranno confrontati i risultati conseguiti da ogni gruppo. Inoltre gli arbitri prenderanno nota della durata del video T, definita precedentemente.

Il formato del file *NomeGruppo_NumeroSessione.xml* con i risultati dei concorrenti per ogni sessione è il seguente:

```
<results>
  <marker>
    <id>1</id>
    <type>Cerchio/Stella</type>
    <time>Secondo di ingresso nella scena</time>
  </marker>
  <marker>
    <id>2</id>
    <type>Cerchio/Stella</type>
    <time>Secondo di ingresso nella scena</time>
  </marker>
  ...
  <marker>
    <id>N</id>
    <type>Cerchio/Stella</type>
    <time>Secondo di ingresso nella scena</time>
  </marker>
</results>
```



Affinchè il calcolo degli indici di Precision e Recall sia effettuato in maniera trasparente, è necessario definire univocamente le modalità di identificazione dei true positive (TP), dei false positive (FP) e dei false negative (FN).

Un marker si considera "entrato" nella scena quando è completamente all'interno di essa; analogamente un marker è considerato "uscito" allorquando esso esce dalla scena in tutte le sue parti.

Gli arbitri valuteranno i risultati utilizzando i seguenti criteri:

- Un marker rilevato con un errore massimo di ± 2 secondi e dello stesso tipo rispetto ad uno presente nella groundtruth, è considerato un true positive.
- Non sono ammessi allarmi duplicati sullo stesso marker. Qualsiasi altro marker rilevato all'interno della stessa finestra temporale del primo e per il quale non sono presenti altre corrispondenze nella groundtruth, è considerato un false positive.
- Un marker rilevato con un errore massimo di ± 2 secondi ma di tipo diverso rispetto ad uno presente nella groundtruth, è considerato un false positive.
- Un marker rilevato con un errore superiore a ± 2 secondi, a prescindere dal tipo, è considerato un false positive.
- Ogni marker presente nella groundtruth che non ha alcuna corrispondenza con i marker presenti nel file XML con i risultati, è considerato un falso negativo.

Oltre all'accuratezza si richiede che l'algoritmo sia in grado di elaborare le immagini in real-time. Ciò significa che il tempo di elaborazione del filmato in secondi T_{el} deve essere inferiore o uguale alla sua durata T . Ogni gruppo che non rispetterà questa specifica riceverà una penalizzazione di un numero di FP pari a $(T_{el} - T) / 3$. In pratica ogni gruppo sarà penalizzato con un FP ogni 3 secondi (75 frames) di ritardo sull'elaborazione rispetto alla durata del video T . Questi falsi positivi contribuiranno a ridurre l'F-Score totalizzato dal gruppo nella corrispondente sessione.

