

SpinRetail: IA al Servizio del Retail

Cristiano Carlevaro^a, Roberto Larcher^a, Stefano Messelodi^b, Andrea Simonelli^b

SpindoxLabs^a, Fondazione Bruno Kessler^b

{cristiano.carlevaro,roberto.larcher}@spindox.it, {messelod,asimonelli}@fbk.eu

Abstract

Il mondo retail sta vivendo un periodo di forte evoluzione. L'e-commerce conquista grandi fette di mercato, soprattutto per certe categorie di prodotti come ad esempio l'elettronica. Per difendersi, il mondo dei retailer tradizionale sente forte l'esigenza di introdurre innovazioni in grado di aumentare i margini di profitto e la fidelizzazione della propria clientela. La soluzione che abbiamo sviluppato, sfruttando sensoristica appositamente progettata per lo scopo e algoritmi di deep-learning per l'analisi dei dati, cerca di dare una risposta a queste richieste. Grazie ad un sistema hardware-software integrato in un espositore è in grado di monitorare in tempo reale la merce raccolta dagli utenti all'interno dello store. Questo tipo di monitoraggio permette di implementare servizi ad alto valore per il retailer, come ad esempio la generazione di allarmi nel caso di prodotti in esaurimento o la notifica di tentati furti.

1. Introduzione

Il monitoraggio automatico di prodotti all'interno degli store è un tema di interesse e molti stanno cercando di dare una soluzione al problema. L'esempio più famoso è quello di Amazon Go [Wikipedia], un sistema in grado di monitorare i prodotti raccolti dagli utenti con tale precisione da potersi permettere di non avere personale umano che verifica quanto presente nei carrelli dei clienti. Tale soluzione tuttavia soffre di un problema: è estremamente costosa. La qualità della sensoristica utilizzata, il numero di telecamere disposte all'interno dello store e la capacità di calcolo necessaria per l'elaborazione dei dati raccolti non la rendono appetibile al mondo dei retailer. Questo settore è infatti caratterizzato da margini di profitto minimi e di conseguenza da una capacità di investimento in soluzioni innovative estremamente contenuta. Il progetto Spinretail¹ nasce proprio con questo proposito: realizzare un sistema di monitoraggio con finalità simili a quelli di Amazon GO utilizzando però un'infrastruttura con un

costo ridotto e quindi accessibile per i retailer tradizionali. La soluzione nella sua totalità sfrutta tre tipologie di sensori:

Telecamere: dislocate nei punti strategici dello store che sono gli ingressi, le casse e gli scaffali.

Bilance: disposte sui vari ripiani degli scaffali con il compito di monitorare in modo continuativo il peso degli oggetti posti sullo scaffale.

Sensori di posizione: disposti sul soffitto del supermercato per rilevare la posizione degli utenti e tracciarne gli spostamenti in due modalità: attraverso appositi tag disposti nei carrelli o tramite la localizzazione dello smartphone dei clienti.

Il presente contributo è focalizzato sulla analisi delle sequenze video fornite dalle telecamere, in quanto sfrutta pesantemente le moderne tecniche di intelligenza artificiale.

2. Metodologia

Le telecamere sono poste sugli ingressi, sulle uscite (casse), e sugli espositori. Nei primi due casi sono a soffitto, nel terzo sono installate sulla parte superiore dell'espositore (Fig.1). Si tratta di telecamere HD a colori, grandangolari, con l'obiettivo rivolto verso il basso. Ognuna cattura 15 frame al secondo ad una risoluzione di 1080p. Le immagini vengono compresse direttamente dalle telecamere ed inviate ad un'unità di elaborazione dotata di GPU.

Il sistema di visione adempie i seguenti compiti: determinare quali oggetti vengono presi/rilasciati dagli/sugli scaffali (*Event Detection*) e associare un id univoco ad ogni cliente visibile dalla telecamera (*Reidentification*).

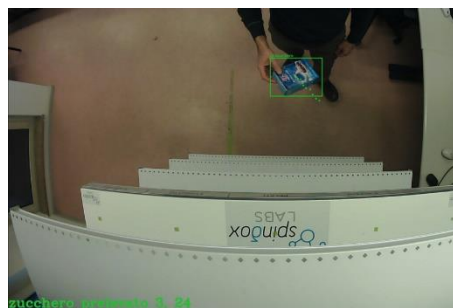


Figura1 - Frame ripreso dalla telecamera sull'espositore, con evidenziati risultati parziali dell'applicazione di Event Detection

¹ Progetto parzialmente finanziato da Provincia Autonoma di Trento tramite LP6/99

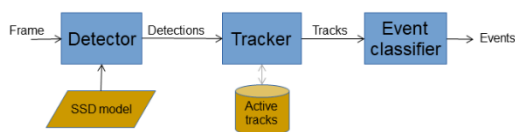


Figura 2 - Schema del modulo di Event Detection

2.1 Event Detection

La rilevazione degli oggetti presi e rilasciati viene realizzata tramite l'esecuzione in sequenza di tre moduli: object detector, object tracker, event classifier (Fig. 2).

Object detector: ha il ruolo di identificare se nel frame ripreso dalla telecamera compaiono uno o più oggetti tra quelli esposti sugli scaffali. Ad esempio, in Figura 1 - Frame ripreso dalla telecamera sull'espositore, con evidenziati risultati parziali dell'applicazione di Event Detection il detector ha rilevato la presenza di una scatola di zucchero. L'output del detector è un elenco di box che racchiudono gli oggetti identificati. Ad ognuna delle box viene associata un'etichetta con l'oggetto identificato ed uno score di confidenza della detection effettuata. Le detection vengono considerate anche se la loro confidenza è molto bassa, sono poi gli step successivi a dare robustezza al sistema eliminando le box errate. Il modulo è realizzato mediante una rete neurale (SSD [Liu et al., 2015]) che permette di localizzare e classificare oggetti appartenenti ad un insieme predefinito di categorie.

Object tracker: ha il ruolo di tracciare il movimento di un oggetto mentre si muove nell'area ripresa dalla telecamera. Per fare ciò analizza le box rilevate in un certo istante e, tenendo conto della posizione spaziale e della classe del prodotto, valuta la possibilità di associarle a quelle rilevate negli istanti precedenti costruendo così una o più tracce attive. Le tracce interrotte vengono passate al modulo successivo per la relativa classificazione. In Fig. 1 la nuvola di punti verdi rappresenta le 10 posizioni in cui lo zucchero è stato rilevato dall'object detector nei frame precedenti. Il problema della associazione tra detection successive viene modellato come un problema di ottimizzazione su grafo.

Event classifier: deve determinare se la traccia in ingresso corrisponde ad un evento di presa, di rilascio o a un falso allarme. In aggiunta fornisce la localizzazione dell'evento in termini di ripiano e di posizione in riferimento alla larghezza dello scaffale. La classificazione si basa sulla posizione spaziale delle box di inizio e fine traccia. Per determinare il ripiano interessato all'azione si analizzano alcuni frame che precedono e seguono l'evento per identificare le zone interessate da movimento e la loro interazione con i ripiani visibili nelle immagini. In Fig. il sistema ha rilevato che lo zucchero è stato prelevato dal terzo scaffale.

2.2 Reidentification

Il sistema di re-identificazione utilizza immagini provenienti da tre distinte zone dello store:

Ingresso del supermercato: grazie ad una telecamera posizionata all'ingresso del supermercato, per ogni persona che accede, viene creato un identificatore univoco, una sorta di "firma" legata all'apparenza della persona. La collezione di queste informazioni viene chiamata *gallery*.

Interno del supermercato: quando una persona viene rilevata da una telecamera posta sugli espositori, il sistema ne raccoglie alcune immagini per trovare la persona più simile tra quelle presenti nella *gallery*.

Uscita del supermercato: grazie ad una telecamera posizionata all'uscita del supermercato, vengono raccolte le immagini delle persone che escono dalla struttura che, una volta re-identificate, vengono rimosse dalla *gallery*. Quest'ultima conterrà così le sole persone presenti nel supermercato. Il sistema di re-identificazione non si basa sui dati biometrici, come l'immagine del volto, ma su caratteristiche globali quali struttura fisica e abbigliamento [Messelodi et al, 2015].

3. Discussione

Il lavoro qui descritto punta soprattutto all'economicità della soluzione anche a spese di una minor precisione rispetto ad alcuni prodotti sul mercato, più costosi e complessi. La precisione della soluzione è comunque sufficiente per poter implementare servizi a valore aggiunto per il retailer. A seguire presentiamo due esempi.

Segnalazione out of stock: essendo il sistema in grado di tener traccia dei prodotti, è possibile conoscere in tempo reale il numero di oggetti divisi per tipologia presenti nello store. Nel caso di un prodotto in via di esaurimento il sistema invia una notifica ad un operatore che può intervenire ed evitare disagi per i clienti e mancate vendite per lo store.

Notifica tentati furti: la capacità di associare i prodotti ai clienti che li hanno prelevati permette al sistema la creazione di un carrello virtuale associato al cliente. Anche se, allo stato attuale, le prestazioni non sono tali da realizzare un carrello virtuale completo, anche il riconoscimento di un sottoinsieme di prodotti può risultare utile per la prevenzione dei furti: se alla cassa un cliente non presenta uno dei prodotti a lui associati è possibile generare una notifica di tentato furto. Il gestore può decidere se procedere immediatamente ad una verifica oppure tenerne traccia ed intervenire solo al momento di una sua eventuale ripetizione.

Uno dei punti critici della soluzione sviluppata consiste nella capacità di individuare una modalità semplice e veloce di addestramento del modulo di object detection. Attualmente, per ogni nuovo prodotto da riconoscere è necessario fornire al modulo un insieme di immagini in cui il prodotto viene posizionato/prelevato dallo scaffale. Successivamente tali immagini devono essere annotate con posizione e categoria del prodotto. Fatta questa operazione per ogni nuovo prodotto può essere avviato l'addestramento del modello. Per migliorare ciò stiamo lavorando in diverse direzioni tra cui l'utilizzo di immagini prodotte sinteticamente a partire da modelli 3D degli oggetti.

Riferimenti Bibliografici

- [Wikipedia] https://en.wikipedia.org/wiki/Amazon_Go
- [Liu et al., 2015] Wei Liu et al. *SSD: Single Shot MultiBox Detector* (arXiv:1512.02325v5)
- [Messelodi et al, 2015] S. Messelodi, C.M. Modena, *Boosting Fisher Vector based Scoring Functions for Person Re-Identification*, Image and Vision Computing, Vol. 44, 2015