Studio di fattibilità per un sistema di highlights automatici di una partita di calcio

Fabrizio Zavanone, Jacopo Favaro

UNIVERSITÀ DEGLI STUDI
DI GENOVA

Sommario

Riassunto

Lo scopo di questo report è quello di descrivere il lavoro svolto nell'ambito del progetto di Transactional Systems & Data Warehouse per l'anno accademico 2017/2018. Verranno presentate le tecniche scelte per l'analisi del video, le tecnologie utilizzate, le idee prese in considerazione e successivamente scartate, nonché le motivazioni dietro tali scelte. Verranno inoltre presentati i risultati ottenuti per l'analisi della partita Italia – Francia, finale del mondiale del 2006.

1. Introduzione

Attualmente il lavoro di creazione di highlights sportivi viene eseguito manualmente da un incaricato che, durante lo svolgimento del match, annota tutti gli avvenimenti ritenuti importanti che accadono. Successivamente, un ristretto numero di questi avvenimenti verrà scelto per creare gli highlights finali del match. Il nostro progetto si propone di esplorare la fattibilità di un sistema automatizzato per la creazione del documento prima discusso in tempo reale, in modo da alleggerire il lavoro umano dietro questa procedura. Il nostro sistema è in grado di riconoscere con buona precisione eventi relativi ad ammonizioni, espulsioni, primi piani, goal ed inizio e fine di tempi di gioco riuscendo a collocarli correttamente in ordine temporale; il tutto rimanendo completamente real-time su un feed video diretto o appartenente ad un qualsiasi formato.

2. Feature rilevate

Dato l'enorme numero di possibili avvenimenti durante una partita di calcio, è da subito stato chiaro che fosse impossibile, con i mezzi a nostra disposizione, eseguire un'analisi completa al 100%. Abbiamo quindi deciso di concentrarci su pochi eventi facilmente riconoscibili e tentare di rintracciarli all'interno di una partita. Le tecniche scelte si sono rivelate soddisfacentemente precise ed il loro carico computazionale sopportabile dalla macchina.

2.1 Estrazione di cartellini gialli e rossi

Abbiamo implementato una nostra logica, affidandoci alle predizioni di una rete neurale basata su YOLO ed al rilevamento del colore sfruttando la libreria OpenCV. YOLO [1] (You Only Look Once) è un sistema che ha ottenuto risultati "stato dell'arte" nell'analisi in tempo reale di immagini. Al contrario del sistema R-CNN basato su classificatori, necessita unicamente di una valutazione per frame risultando circa 1000 volte più veloce. Per tradurre la logica di darkent [2] (libreria in C alla base di YOLO) in Python, ci siamo avvalsi di darkflow [3], che fornisce un equivalente Python appoggiandosi alla libreria Tensorflow. Per l'analisi è stata necessaria la creazione di un dataset proprietario, che consta di 167 immagini di cartellini gialli e 192 immagini di cartellini rossi. Il sistema basato unicamente sul dataset presentava numerosi falsi positivi, dovuti al colore della maglia dell'arbitro, alle bandierine dei guardalinee o a colori particolarmente sgargianti di borracce e persone nel pubblico. Per ovviare a questo

problema, un sistema di controllo ritaglia il frame utilizzando il bounding box dell'oggetto identificato e ne analizza la composizione cromatica con la libreria OpenCV. Nel caso vi sia discordanza cromatica tra il colore che dovrebbe avere il cartellino rilevato dalla rete e quello appena calcolato, la percentuale di sicurezza viene immediatamente abbassata. Un evento cartellino viene registrato solo se per più di 4 volte all'interno di una singola scena viene rilevato un oggetto cartellino con threshold di sicurezza maggiore del 55%.

Figura 1: **risultato dell'analisi di un cartellino all'interno di una partita.** A sinistra l'immagine come viene analizzata, a destra un esempio del crop e della sua maschera per il controllo cromatico.

2.2 Scene change recognition

Il problema del riconoscimento del cambio di scena, sebbene non direttamente utile al tagging di un evento, si è presto rivelato vincente come tecnica di controllo o conferma. Il controllo del cambio scena, infatti, permette di restringere i controlli per un possibile evento all'arco di una scena, riducendo il rischio di errore o falso positivo drasticamente. La logica adottata per risolvere il problema è basata interamente su OpenCV. Ogni 2 secondi il frame viene passato a una funzione che, dopo averlo diviso in tre diverse maschere (una per ogni colore principale R G B) esegue il rapporto tra i pixel non neri di tali maschere e del frame originale, trovando, così, la ratio della scena. Tramite chi-squared distance viene poi calcolata la distanza tra i valori appena ottenuti con quelli ricavati al ciclo precedente [4]. Se la distanza in questa maniera calcolata supera il valore di threshold 0.1, una nuova scena viene riconosciuta ed i relativi valori vengono aggiornati.

2.3 Real time face recognition

Per risolvere il problema della face recognition è stata utilizzata un'API per Python 3.3+ denominata 'face_recognition' [5]. Questa modello è stato scelto, oltre che per la sua affidabilità, soprattutto per la sua semplicità. Essendo lo scopo principale del progetto analizzare un video in tempo reale, abbiamo deciso di sacrificare precisione per migliorare le performance ove possibile. L'analisi delle facce rileva e salva tutti i primi piani all'interno di una scena, basandosi su un dataset proprietario che viene caricato come prima azione dal programma. Per il corretto funzionamento è sufficiente una sola immagine per persona dove compaia chiaramente il viso e la label utilizzata per tale viso sarà il nome dell'immagine stessa. A causa del peso computazionale di questa analisi è stato deciso, al fine di mantenere la componente real-time, di eseguire l'analisi su un frame ogni 3 e una faccia viene considerata presente se, all'interno di una scena, compare almeno 5 volte. Questo alleggerisce notevolmente il carico computazionale ed elimina molti falsi positivi dovuti alla non ottima qualità delle immagini nel dataset. Abbiamo riscontrato forti difficoltà nel riconoscimento di molti

giocatori di colore della Francia che, a causa di fisionomie simili, vengono spesso scambiati o non riconosciuti. Anche in questo caso, crediamo che un miglioramento del dataset, con immagini di qualità superiore rispetto a quelle reperibili in rete, si possano migliorare notevolmente le prestazioni ottenute. Inoltre, l'impiego di un hardware migliore permetterebbe di estendere l'analisi ad ogni frame senza inficiare sulla velocità dell'output.

Figura 2: esempio di riconoscimento facciale durante gli inni. L'immagine mostra il risultato a schermo del riconoscimento dei due giocatori italiani durante l'inno.

2.4 Goal detection e fine di tempo di gioco

Allo scopo di riconoscere un evento di tipo goal, abbiamo deciso di utilizzare le stesse tecniche del riconoscimento del cambio di scena e di individuazione del colore del cartellino, al fine di individuare la scomparsa e ricomparsa del tabellone grafico di gara nell'angolo in alto a sinistra dello schermo e la variazione di pixel bianchi del numero rappresentante il punteggio. La scomparsa del tabellone dalla grafica della partita per più di 12 secondi implica sempre un evento. Tale evento viene registrato come goal nel caso in cui la riapparizione avvenga entro i successivi 120 secondi e venga registrata una variazione di pixel nell'area contenente il numero del risultato. Nel caso in cui nessuna variazione venga registrata il sistema registra un evento di tipo mini spot (pubblicitario), mentre nel caso la scomparsa si protragga per più di 2 minuti verrà decretata la fine di un tempo regolamentare e il sistema registrerà automaticamente un evento consono tramite un sistema incrementale. La primissima comparsa del tabellone segnerà l'inizio della partita. L'estrazione del tabellone nell'angolo in alto a sinistra è stata impossibile da automatizzare, sebbene teoricamente possibile tramite il procedimento descritto in uno studio in cui ci siamo imbattuti. Vengono sfruttati l'eliminazione di pixel di colore verde appartenenti al campo di gioco, la differenza tra frame successivi e un processo denominato "*expansion processing*" [6]. La manipolazione necessaria ad adottare questa tecnica richiede una potenza computazionale a cui non potevamo accedere. In questo momento il tool ricava la regione del tabellone e quella dei punteggi delle due squadre tramite conoscenza pregressa della loro posizione relativa all'interno dello schermo, assumendo uno sfondo al più opaco o semitrasparente. Questo task è stato il più complesso da gestire, nonché quello che ha portato all'esplorazione di più idee fallimentari.

Figura 3: crop del tabellone e dei punteggi. L'immagine mostra il crop del tabellone e dei punteggi insieme alle maschere realizzate per l'analisi dei pixel bianchi per il risultato

3. Idee scartate

Gli obiettivi raggiunti nella versione finale sono stati frutto di un'evoluzione costante, durante la quale alcune idee sono state scartate. La bocciatura di un'idea può essere avvenuta per svariati motivi, tra cui l'inadeguatezza dell'hardware utilizzato o la nostra inesperienza nel campo. Per tale motivo proponiamo qui di seguito le idee che sono state scartate in questo specifico progetto.

3.1 YOLO per facce

Una possibilità esplorata è stata quella di utilizzare la rete basata su YOLO per identificare anche i volti dei calciatori. Questa strada si è presto rivelata impraticabile a causa dell'enorme numero di immagini che sarebbe stato richiesto per la fase di training. Abbiamo successivamente realizzato che utilizzare una rete trainata appositamente per rilevare le feature di volti avrebbe prodotto risultati decisamente migliori e l'idea è stata quindi scartata.

3.2 YOLO per replay

Uno delle prime idee è stata quella di utilizzare il replay per identificare un'azione saliente. Per fare ciò abbiamo tentato di individuare il simbolo della competizione, che prima di ogni replay fa da transizione sullo schermo. Il tentativo di riconoscere tale simbolo utilizzando un dataset si è però rivelato fallimentare: questo, infatti, appare per pochi frame spesso sfuocato e comunque in pessime condizioni di semi trasparenza. A causa degli scarsi risultati ottenuti, l'idea è stata scartata in favore di altre più efficienti.

3.3 Sequence recognition per goal

Uno studio trovato in rete [7], mostrava l'efficacia di utilizzare una rete convoluzionale trainata su azioni da goal per il riconoscimento combinato di feature spaziali e temporali. Questa strada, sebbene probabilmente più efficacie e generalizzabile di quella adottata, è stata scartata a causa dell'insufficiente potenza hardware a disposizione e della sua complessità. Viene qui menzionata come possibile futuro lavoro di miglioramento.

3.4 Audio recognition

L'analisi della telecronaca è stata inoltre considerata come possibilità in fasi iniziali, sia per il riconoscimento sia per la conferma di un avvenuto evento. Tali tecniche sono infatti comunemente usate in studi simili e, sebbene non abbiano raggiunto risultati sorprendenti, si sono comunque rivelate utili per l'interpretazione di quanto analizzato. Siamo convinti che tali tecniche possano migliorare la precisione del nostro lavoro ma, come per la sequence recognition, le limitazioni imposte dal nostro hardware hanno portato a scartare questa possibilità e a preferire altre opzioni computazionalmente più leggere.

4. Esperimenti

Proponiamo di seguito i risultati ottenuti durante alcuni esperimenti di test eseguiti durante lo sviluppo del tool discusso.

4.1 Riconoscimento cartellini gialli e rossi

VIDEO	QUALITA'	GIALLI	ROSSI
Italia – Francia 2006 (partita completa fino a fine supplementari)	Alta	3/3	3/2 È presente un falso positivo.
Portogallo – Olanda 2006 (compilation gialli e rossi, breve video)	Bassa	6/10	2/2

Il falso positivo individuato è dovuto ad un errore della rete, che confonde uno steward con pettorina rossa con un cartellino. Tutti i controlli effettuati sul colore non sono in grado di rilevare questo particolare errore, confermandolo così come cartellino rosso.

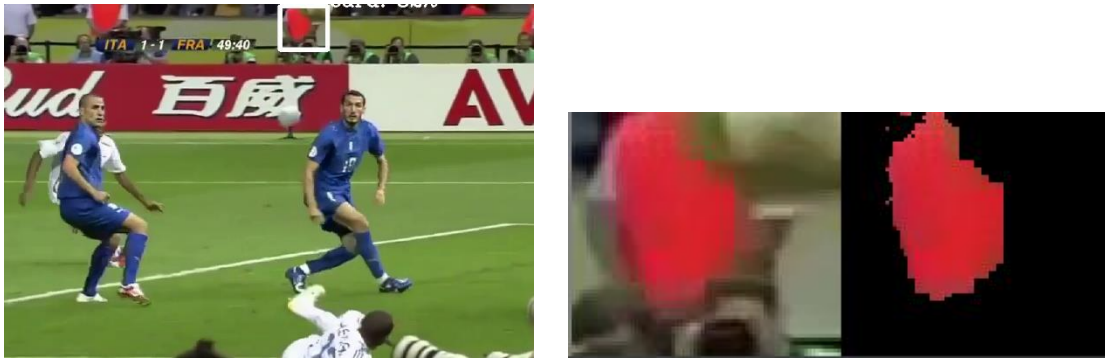

Figura 4: frame del falso positivo e relativa coppia crop-maschera. L'immagine a sinistra rappresenta un frame preso dalla partita dove compare il falso positivo (all'interno del bounding box), a destra il risultato dell'analisi del crop che conferma l'errore.

4.2 Riconoscimento facce

VIDEO	QUALITA'	SQUADRA	FACCE
Italia – Francia 2006 (inni nazionali)	Alta	Italia	11/16
Italia – Francia 2006 (inni nazionali)	Alta	Francia	4/14 Problema sul riconoscimento giocatori di colore.

Il problema del riconoscimento dei giocatori di colore è già stato brevemente citato in precedenza. Si tratta di un errore legato a fisionomie simili dei giocatori, unito alla scarsa qualità delle immagini che siamo riusciti a reperire.

4.3 Riconoscimento goal

VIDEO	QUALITA'	SQUADRA	GOAL
Italia – Francia 2006 (partita completa fino a fine supplementari)	Alta	Italia	1/1
Italia – Francia 2006 (partita completa fino a fine supplementari)	Alta	Francia	1/1

4.4 Riconoscimento fine/inizio tempo e mini-spot

VIDEO	QUALITA'	INIZIO TEMPO	FINE TEMPO	MINI-SPOT
Italia – Francia 2006 (video editato per verificare la nostra logica)	Alta	2/2	2/2	1/1

5. Conclusioni

Abbiamo discusso il tool sviluppato per la creazione automatica di highlights di una partita di calcio. Lo studio proposto mostra risultati promettenti: sebbene non facilmente generalizzabili, i test fatti hanno portato a risultati soddisfacenti mantenendo la possibilità di effettuare analisi in tempo reale anche su macchine dai componenti non ottimali. I pochi errori riscontrati possono essere facilmente migliorati da una maggiore potenza di calcolo, combinata con un dataset più ricco e strutturato.

References

[1] Joseph Redmon and Ali Farahadi . *You Only Look Once: Unified, Real-Time Object Detection*. https://pjreddie.com/darknet/yolo/ . 2016.

[2] Joseph Redmon. *Darknet: Open Source Neural Networks in C*. https://pjreddie.com/darknet/. 2016

[3] Trieu. *darkflow.* https://github.com/thtrieu/darkflow. *2017.*

[4] Yuan Jiang. *CS283 Final Project – Video Scene Detection.* https://www.youtube.com/watch?v=VG3oGwXGXQw*. 2015*

[5] Adam Geitgey. *The world's simplest facial recognition api for Python.* https://github.com/ageitgey/face_recognition. 2017

[6] Naoki Ueda, Masao Izumi. *Detecting Soccer Goal Scenes from Broadcast Video using Telop Region.* http://www.iaiai.org/journals/index.php/IEE/article/view/187/100. 2017

[7] Grigorios Tsagkatakis, Mustafa Jaber, Panagiotis Tsakalides. *Goal!! Event detection in sports video.* https://www.ingentaconnect.com/contentone/ist/ei/2017/00002017/00000016/art00004?crawler=true&mimetype=application/pdf. 2017