Miglioramento del sistema di annotazione WATSS con tecniche di Computer Vision

Lorenzo Cioni
lore.cioni@gmail.com

1. Introduzione

L'obiettivo di questo elaborato di migliorare, anche tramite tecniche della *Computer Vision*, il sistema di annotazione web *WATSS*[2].

WATSS, abbreviazione per *Web Annotation Tool for Surveillance Scenarios*, un sistema di annotazione web per la creazione di un groundtruth di scenari di sorveglianza. Il sistema consente infatti di annotare persone all'interno dei singoli frame di un video, assegnandogli una posizione (determinata tramite una *bounding box*), una identit (tramite *avatar*), la parte visibile e le orientazioni del corpo e dello sguardo. E' inoltre possibile associare pi persone ad un medesimo gruppo e il punto di interesse presso il quale la persona si trova.

Uno degli obiettivi quello di introdurre nel sistema un meccanismo di predizione delle annotazioni, andando a generare, a partire da una o pi annotazioni consecutive di una stessa persona, una serie di *proposals* per i frames successivi. L'elaborazione dell'immagine a questi scopi viene effettuata tramite l'utilizzo di OpenCV, una libreria open source, nativa per C++, per la Computer Vision e l'Image Analysis.

Viene poi presentato uno studio e l'implementazione di un sistema che consente di sfruttare la geometria della scena per proporre delle annotazioni possibili.

Nelle sezioni successive viene presentata inizialmente un'analisi comparativa tra i vari sistemi di annotazione che vanno a costituire l'attuale stato dell'arte. Viene poi presentata la parte relativa alle migliorie apportate al sistema e le tecniche di Computer Vision utilizzate. Infine viene presentata un'analisi di usabilit a posteriori, mettendo in evidenza anche eventuali sviluppi futuri per l'applicazione.

2. Stato dell'arte

In questa sezione viene presentata un'analisi comparativa di alcuni dei pi famosi sistemi di annotazione esistenti. Lo scopo di questa ricerca quello di individuare i le caratteristiche comuni ai vari strumenti e le loro limitazioni. Da questa analisi poi stata stilata una lista di requisiti che portano WATSS ad essere in accordo con gli altri sistemi introducendo allo stesso tempo nuove caratteristiche.

L'analisi dei sistemi si incentrata principalmente su 3 strumenti open source esistenti: *LabelMe*, *ViPER-GT* e *VATIC*. Per ciascuno dei sistemi stata stilata una lista di caratteristiche offerte ed evidenziate le eventuali limitazioni. Infine presentata un'analisi anche con il sistema WATSS.

2.1. LabelMe

LabelMe[3] un sistema Web che consente l'annotazione di oggetti all'interno di immagini. Le singole annotazioni sono effettuate mediante la definizione di aree poligonali nell'immagine e l'assegnazione di una label. Il tool offre la possibilit di indicare se un oggetto annotato occluso o meno da altri oggetti presenti nella scena (non consente per di individuare la parte occlusa o visibile).

Le annotazioni possono essere annidate, possibile dunque etichettare oggetti che sono inclusi gli uni negli altri.

In aggiunta alle annotazioni di oggetti il sistema consente di annotare intere aree dell'immagine: questo reso possibile andando inizialmente a delimitare una porzione di immagine ed associando ad essa una label. In questo caso l'area cos definita viene colorata interamente ed necessario stabilire se si tratta di un'area interna o esterna.

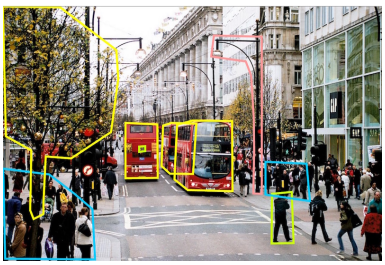

Figure 1. Annotazione di un'immagine tramite LabelMe

In fase di esportazione delle annotazioni viene generata una struttura in formato XML che possibile importare nuovamente in un'altra immagine.

Il sistema non prevede la possibilit di generare *proposals* per le annotazioni, tutto il lavoro a carico dell'utente.

2.2. ViPER-GT

ViPER-GT[4], acronimo di *Video Performance Evaluation resource*, un sistema di annotazione per video e la generazione di un *groundtruth*.

Il sistema consente di annotare un video indicando cosa contenuto nella scena, definendo un insieme di *classi* per ciascuna tipologia di contenuto. L'annotazione avviene manualmente da parte dell'utente che definisce delle *bounding boxes*, regioni di interesse dell'immagine, andando ad associare a ciascuna di esse una classe di appartenenza ed alcune metadati aggiuntivi, come ad esempio titolo, dimensione, etc. Le regioni possono essere di forme diverse: cerchi, ellissi, rettangoli e generici poligoni (definiti dai loro vertici).

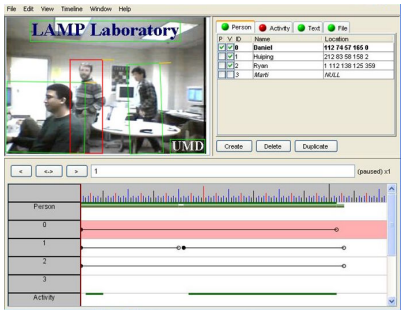

Figure 2. Interfaccia utente del tool ViPER-GT

Le annotazioni effettuate vengono visualizzate in una *timeline*: allo scorrere dei frame le annotazioni presenti nella scena corrente vengono evidenziate.

Il tool mette a disposizione anche un sistema di predizione delle annotazioni inserite basato sull'interpolazione lineare di pi frame consecutivi. Questo metodo risulta molto efficace se si fornisce un numero di annotazioni, chiamate *ancore*, adeguato; con poche ancore definite la predizione risulta essere molto approssimativa.

2.3. VATIC

VATIC un software di annotazione di video distribuito ai fini della ricerca nell'ambito della Computer Vision che consente la creazione di grandi dataset video. Il tool utilizza il sistema di crowdsourcing *Mechanical Turk* di Amazon.

Il sistema consente l'inserimento manuale di annotazioni per ciascun frame del video, definite mediante delle bounding box rettangolari.

Il tool dispone di una serie di plugin aggiuntivi che ne aumentano le potenzialit:

- *Tracking integration* per il tracciamento di oggetti in movimento nella scena
- *Sentence annotation* per l'annotazione di frasi e parole
- *Labeling time intervals* per l'annotazione di intervalli temporali
- *Human action labeling* per l'annotazione di azioni umane nella scena

Figure 3. Interfaccia utente del tool VATIC

Il tool pensato principalmente per l'object detection nelle scene.

2.4. WATSS

WATSS[2], *Web Annotation Tool for Surveillance Scenarios*, un sistema web per l'annotazione di dataset. Il tool stato sviluppato per consentire l'annotazione del dataset *MuseumVisitors*[1] come parte del progetto MNEMOSYNE e rilasciato successivamente open source.

Il tool permette l'annotazione di persone e oggetti nella scena mediante la definizione di bounding boxes. In caso di occlusione possibile poi indicare poi, mediante la definizione di una seconda bounding box, la parte visibile della persona o oggetto annotati.

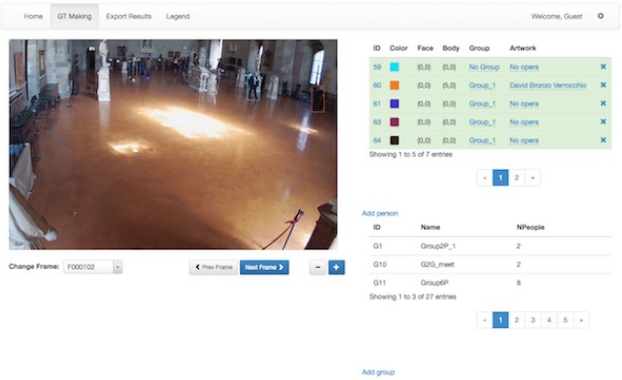

Figure 4. Interfaccia utente del tool WATSS

A ciascuna annotazione corrisponde un'*identit*: della persona annotata viene generato un avatar che consentir di annotare la stessa persona negli altri frames del video. Per ciascuna annotazione inoltre possibile indicare inoltre l'*orientazione del volto e del corpo* e il *punto di interesse* presso cui si trovano nella scena (nel caso del museo i punti di interesse sono rappresentati dalle varie opere d'arte).

In caso di presenza di gruppi di persone, possibile indicare il gruppo di appartenenza definendo il nome dello

stesso, cos da poterle poi riassociare in seguito anche nelle altre camere.

Il tool consente l'annotazione da parte di pi utenti e la gestione di pi camere.

3. Obiettivi

L'obiettivo principale del presente elaborato di migliorare alcune caratteristiche del sistema di annotazione WATSS, utilizzando alcune tecniche di Computer Vision al fine di agevolare la creazione di annotazioni.

3.1. Interfaccia utente

Da un'analisi del sistema precedente, emergeva che alcune operazioni effettuabili tramite interfaccia risultavano essere poco intuitive per l'utilizzatore. In particolare, dai risultati del *System Usability Scale (SUS)* presentato in [2], si ritiene che non sia molto semplice imparare ad usare il sistema.

Le modifiche all'interfaccia grafica sono state dunque apportate con il fine di rendere pi chiaro per l'utente le varie funzioni messe a disposizione dal sistema, a partire dalla schermata iniziale devono essere chiare fin da subito le sue caratteristiche e potenzialit.

3.2. Creazione e modifica delle annotazioni

La parte fondamentale del sistema la fase di creazione e raffinamento delle bounding box all'interno della scena. Queste sono definite mediante una serie di rettangoli associati ad opportuni metadati.

Essendo dunque una fase fondamentale, deve essere semplice ed immediato per l'utente poter interagire con le annotazioni, modificandole e inserendole senza difficolt. Il precedente sistema presenta alcune difficolt, non consentendo una rapida modifica all'utente, rendendo l'azione di inserimento delle annotazioni leggermente complicata e difficile da gestire.

Obiettivo per questo aspetto quello di introdurre un nuovo sistema di creazione e modifica delle annotazioni.

3.3. Timeline

Una delle caratteristiche mancanti in questo tool, presenti invece in molti altri sistemi di annotazione, una *timeline*. Questa ha come scopo principale quello della navigazione tra i vari frames e la visualizzazione temporale delle annotazioni inserite.

Con questo strumento infatti possibile visualizzare la durata di permanenza di una stessa persona in pi frames consecutivi, consentendo all'utente di avere maggiore controllo sulla annotazioni inserite ed andare a correggere eventuali mancanze.

3.4. Predizione delle annotazioni

Dato il gran numero di frame da annotare, pu risultare molto utile avere a disposizione un sistema di *predizione* delle annotazioni future in base ad una selezione corrente. Nel sistema implementato un semplice meccanismo di predizione che ripropone una stessa bounding box nel frame successivo che pu essere *approvata* con un click da parte dell'utente.

Mediante tecniche di Computer Vision si vuole fornire dei *proposals* per la posizione e la dimensione della stessa persona nei frame successivi. La predizione verr valutata mediante la combinazione di pi tecniche, come ad esempio la stima del moto, un *pedestrian detector* ed una stima mediante filtro di *Kalman*.

La fase di generazione dei proposals deve integrarsi nell'interfaccia, in particolar modo nella timeline.

3.4.1 Geometria della scena

Un altro tipo di predizione pu essere effettuata inoltre conoscendo la geometria della scena. Questo reso possibile se nota la calibrazione delle telecamere con cui sono stati scattati i frames e se le telecamere sono fisse.

Utilizzando le informazioni spaziali possibile ad esempio prevedere l'altezza di una persona data la sua posizione nella scena, consentendo cos, ad esempio, di ridimensionare automaticamente la bounding box in base alla posizione in cui si vuole inserire.

4. L'interfaccia

L'interfaccia utente del sistema stata rivisitata ed adattata alle nuove esigenze.

La parte principale costituita dal frame video su cui andremo ad aggiungere annotazioni e visualizzare quelle gi esistenti. I due pannelli laterali per le persone e gruppi presenti nella scena sono stati organizzati in modo tale da consentirne una rapida navigazione ed uso. Come nella precedente versione dell'interfaccia, il pannello *People* mostra la lista delle annotazioni presenti nel frame visualizzato, ordinate in base all'identificativo delle persone. Per ciascuna persona viene mostrato il *gaze* del corpo e della faccia, il *gruppo di appartenenza* ed il *punto di interesse* presso cui si trovano.

4.1. Inserimento di una annotazione

Tramite il pulsante *Add person* presente nel pannello delle annotazioni possibile aggiungere una nuova annotazione al frame.

Come mostrato in Figura 4.1, in fase di creazione possibile indicare se si vuole aggiungere un'annotazione rappresentante una nuova identit ancora non presente nel

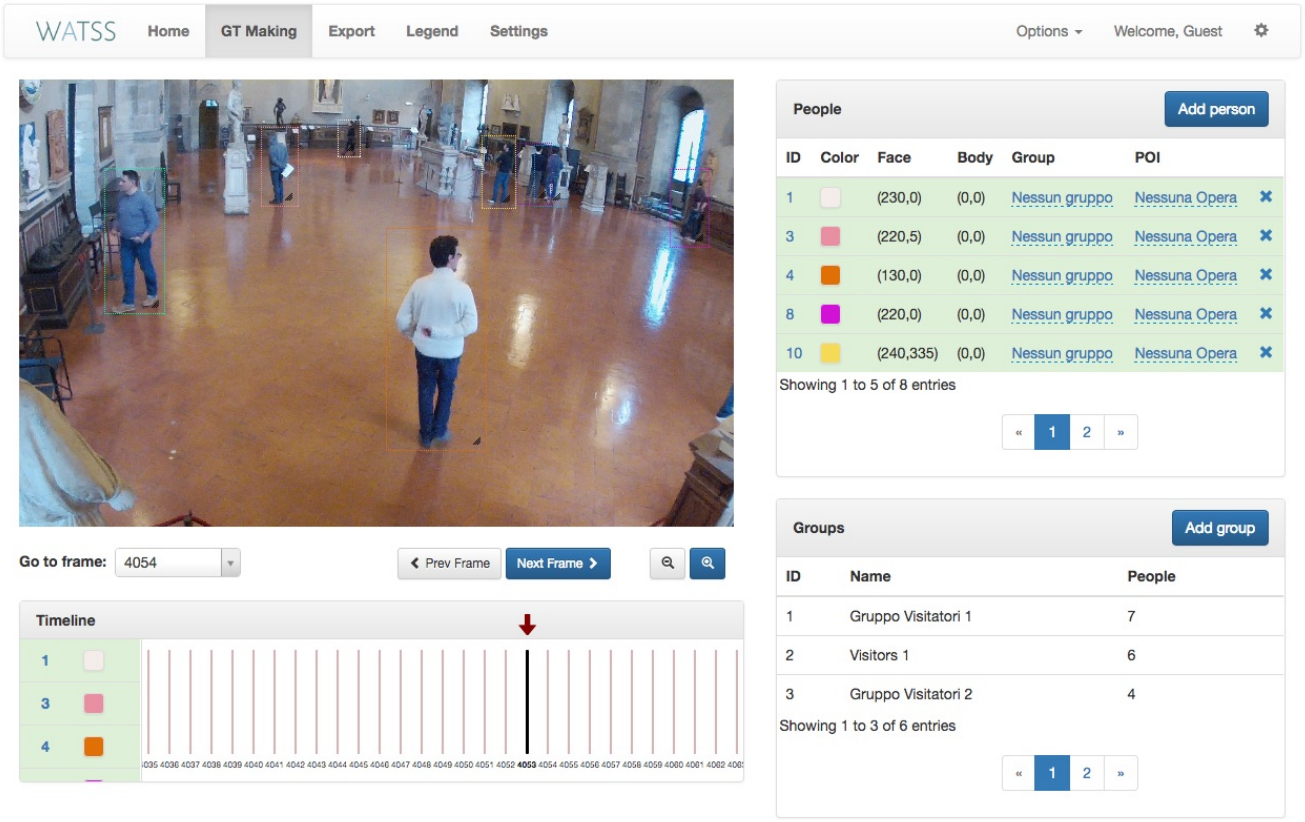

Figure 5. Interfaccia utente del tool WATSS

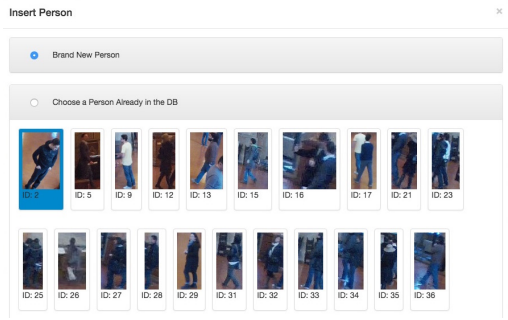

Figure 6. Aggiunta di una nuova annotazione

database oppure se si vuole aggiungere una nuova istanza di un'identit gi presente e di cui viene mostrato un *avatar*.

Una volta selezionata l'opzione desiderata, la fase di creazione diversa in base all'attivazione della *geometria della scena* o meno. Per l'attivazione e la disattivazione della geometria sufficiente spuntare l'opzione presente nel men *Options* della barra principale di navigazione.

In caso di geometria della scena *disattivata* la nuova annotazione verr creata con la tecnica *click and drag*: l'utente clicca nel frame nel punto in cui vuole iniziare la sua selezione e tiene premuto spostando il mouse finch la bounding box visualizzata non della dimensione desiderata. A quel punto, rilasciando il click, la bounding box verr inserita nella scena.

Se invece attiva la geometria, questa verr sfruttata per stimare l'altezza di una persona presente nella scena in base alla sua posizione nella stessa. La bounding box verr automaticamente attaccata al puntatore del mouse e, muovendosi nella scena, sar ridimensionata in base alla sua posizione. Una volta scelta la posizione, per effettuare l'inserimento sar sufficiente effettuare un click nel punto desiderato.

In entrambi i casi, la procedura di inserimento pu essere interrotta premendo il tasto *ESC*.

5. Implementazione

6. Conclusioni

Conclusioni dell'elaborato

References

[1] F. Bartoli, G. Lisanti, L. Seidenari, S. Karaman, and A. Del Bimbo. Museumvisitors: A dataset for pedestrian and group detection, gaze estimation and behavior understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 19–27, 2015.

[2] F. Bartoli, L. Seidenari, G. Lisanti, S. Karaman, and A. Del Bimbo. Watts: A web annotation tool for surveillance scenarios. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 701–704, New York, NY, USA, 2015. ACM.

[3] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173, May 2008.

[4] C. Vondrick, D. Patterson, and D. Ramanan. Efficiently scaling up crowdsourced video annotation. *Int. J. Comput. Vision*, 101(1):184–204, Jan. 2013.