Scuola di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea Magistrale in Informatica

Tesi di Laurea

RICONOSCIMENTO DI AZIONI UMANE USANDO TECNICHE DI APPRENDIMENTO PROFONDO PER LA STIMA DELLA POSA

HUMAN ACTION RECOGNITION USING DEEP LEARNING TECHNIQUES FOR POSE ESTIMATION

ANDREA MOSCATELLI

Relatore: *Marco Bertini*

Anno Accademico 2018-2019

1

ABSTRACT - ITALIANO

I recenti progressi nel campo della visione artificiale hanno permesso alla comunità scientifica di spostarsi verso problemi ancora più articolati rispetto a quelli classici ed il riconoscimento di azioni corporee tramite l'analisi della posa umana sta attraendo recentemente una notevole attenzione. Il suo successo è senzaltro dovuto non solo agli ottimi risultati ottenuti, ma anche alla sua efficiente semplificazione della struttura umana riducendo di fatto i costi computazionali e le risorse necessarie allo stoccaggio dati.

In questo lavoro di tesi ci siamo dedicati al riconoscimento e alla classificazione di azioni umane tramite tecniche di apprendimento profondo per la stima della posa. A tale scopo abbiamo deciso di ideare un algoritmo che non avesse bisogno di informazioni iniziali complesse, come ad esempio la posizione dei giunti dei soggetti inquadrati, ma che attraverso l'uso dei soli video RGB fosse in grado di estrapolare tutte le informazioni necessarie.

Al fine di ottenere il miglior algoritmo abbiamo eseguito un serie di esperimenti strutturati secondo un procedimento ben preciso, che permettesse l'esplorazione rapida di ogni tecnica ideata affinando progressivamente i risultati per quelle più promettenti.

Quello che abbiamo ottenuto alla fine del processo sono due algoritmi con una discreta capacità di classificazione della azioni umane e per la loro semplicità anche un'elevata portabilità.

Il dataset utilizzato per l'addestramento degli algoritmi trovati è stato NTU-RGB+D di *Amir Shahroudy, Jun Liu, Tian-Tsong Ng e Gang Wang* e l'accuratezza dei nostri algoritmi misurate secondo le metriche proposte dagli autori del dataset sono le seguenti:

- il classificatore "*3BAR-NEXT-Detectron*" con un accuratezza Cross subject di $83,32\%$ e Cross view di $93,69\%$

- il classificatore "*3BAR-NEXT-Posenet*" con un accuratezza Cross subject di $79,06\%$ e Cross view di $87,79\%$.

La semplicità di questi due algoritmi sottintende un'elevata portabilità, inoltre pur facendo uso dei soli video RGB, le loro prestazioni sono comparabili a molti altri lavori scientifici facenti uso dalla posizione dei giunti fornita da dataset stesso.

2

ABSTRACT - ENGLISH VERSION

The recent advances in the computer vision field allowed the scientific community to move towards more complex problems than the classic ones and the recognition of human actions through the analysis of the human pose attracted recently considerable attention. Its success is undoubtedly due not only to the excellent obtained results, but also to its efficient simplification of the human structure, drastically reducing the computational and storage costs.

In this thesis work we focused on the recognition and classification of human actions through deep learning techniques for the estimation of the pose. To this end, we decided to create an algorithm that doesn't need complex initial information, such as the joints position of the subjects, and that through the use of the RGB video only is able to extract all the needed information.

In order to obtain the best algorithm, we followed a series of experiments structured according to a very precise procedure, which allowed us to a rapid exploration of the designed techniques, progressively refining the results for the most promising ones.

What we got at the end of the process are two algorithms with a good performance in classifying human actions and for their simplicity a high portability as well.

The dataset used to train the algorithms is *NTU-RGB+D* by *Amir Shahroudy, Jun Liu, Tian-Tsong Ng* and *Gang Wang* and the accuracy of our algorithms measured according to the metrics proposed by the authors of this dataset are the follows:

- the classifier "*3BAR-NEXT-Detectron*" with a Cross subject and Cross view accuracy of $83,32\%$ and $93,69\%$ respectively.
- the classifier "*3BAR-NEXT-Posenet*" with a Cross subject and Cross view accuracy of $79,06\%$ and $87,79\%$ respectively.

The simplicity of these two algorithms implies high portability and while using only RGB videos, their performance is comparable to many other scientific works which use the position of the joints provided by the dataset itself.