An AI Powered Re-Identification System for Real-time Contextual Multimedia Applications

Giuseppe Becchi, Andrea Ferracani, Filippo Principi, Alberto Del Bimbo

[name].[surname]@unifi.it

Università degli Studi di Firenze - MICC

Italy

ABSTRACT

In this demo we present a person re-identification system, based on cameras installed in an environment and featuring AI, that can be used in the design and development of human-centered multimedia applications intended to provide improved situational awareness and context-sensitive user-interfaces [1, 2]. Possible applications are related but not limited to user profiling and personalisation systems, multimedia recommendation, social context understanding and gamification, security, with objectives spanning from environment monitoring, cultural heritage fruition enhancement, retail trade promotion and assistive technologies provision in industry. In the context of the demonstration we are going to set up a system with several workstations equipped with cameras and contextual to paintings reproductions that simulate a museum exhibition in which users are tracked and re-identified at different locations. The data collected is used to enrich the user experience through a reactive voice interface that considers the user's visit, the artworks that most attracted the visitor's attention and the social context.

CCS CONCEPTS

• **Information systems** → *Collaborative and social computing systems and tools*; • **Computing methodologies** → **Cooperation and coordination**; • **Computer systems organization** → **Real-time systems**; • **Applied computing**; • **Human-centered computing** → *Human computer interaction (HCI)*;

KEYWORDS

Face detection, re-identification, age estimation, gender detection, sentiment analysis, cultural heritage, audio-guide.

1 INTRODUCTION

Nowadays, the ease of deployment and use of sensors and cameras capable of tracking people in indoor and outdoor environments, as well as the availability and effectiveness of recognition systems

capable of running in real-time at high frame-rates, especially in restricted domains and environments, allows to design smart and interactive applications able to profile users in particular contexts as well as to follow them, build localized profiles of behaviors and put them in relation to the behavior of other users present in the same space and at the same time. The main goal of our system is to provide the community with a modular framework that can be exploited to easily enable multimedia applications to rely on an efficient detection and re-identification system of people. Associated metadata such as visitors' gender, age and sentiment inferred by face expression recognition are also extracted using artificial intelligence models. This data can help the development of human-centered, personalized, social and situation aware multimedia applications.

2 THE SYSTEM

The framework[1] provides a real-time re-identification system designed to enroll people, to track their movements and re-identify them in real-time in a defined environment at different locations. The first task for the system is to understand if a subject has already been recognized and he is present in the database or if he must be 'enrolled'. The enrollment procedure is performed only if the subject is in favorable conditions (i.e. frontal and within a certain distance) with respect to the camera in order to have a more accurate acquisition. For each subject multiple views of the face are taken at different pitches from several angles in order to improve the accuracy of the re-identification. The system identifies the subject in real-time but it stores the detection in the database only if he stands for a sufficient interval of time in front of the camera (and of the painting). Frame detections are enriched with categorical and numerical features (i.e. age, gender, recognized emotions) that are then saved in the system in order to "build" a profile of the user and the history of his visit and of his emotional reactions to artworks. The system has been developed in Python as a rich internet application using Django[2] and NodeJS[3] for bi-directional communication and is composed by several modules that can be orchestrated in order to perform the following tasks: 1) face detection; 2) face recognition and matching; 3) face age estimation; 4) gender detection; 5) emotion recognition. To this end it provides a back-end that also allows an operator to set up, calibrate, localize and visualize cameras in the exhibition environment and to manage users and detections. Clients associated and equipped with cameras can be plugged-in the system and can easily communicate through REST APIs. All visual data are saved as descriptors in anonymized

---

[1]The Git of the project is available at https://github.com/giuseppebecchi/face-reidentification

[2] https://www.djangoproject.com/

[3] https://nodejs.org/

format in compliance with privacy guidelines. Contextual multimedia content can be provided and put in relation to each active camera plugged-in the system through the back-end interface in order to personalize the audio-visual output of the system at the different locations on the basis of the user's visit history. Extraction of face descriptors, face detection, tracking, face recognition and orientation (i.e. pitch, roll, yaw), age, gender and emotion estimation are provided on the client in real-time (FPS 5-10) using Tensorflow JS[4] and the Face API[5]. Face detection is implemented using a SSD (Single Shot Multibox Detector) based on MobileNetV1. A ResNet-34 like architecture is implemented for face recognition to compute a face descriptor (a feature vector with 128 values).

2.1 The demo

We demonstrate our system, as shown in the referenced video[6] through the design and implementation of a reactive personalized audio-guide in the context of a museum visit. The demo (in the video) simulates a hypothetical visit by two users at the Galleria Palatina Museum [7] in Florence, IT. Custom audio explanations are activated by micro speakers contextual to paintings for each user interested in an artwork and standing in front of it on the basis of analogies, similarities and differences with other artworks of the same artist or of the same subject, or otherwise related, viewed during the visit. The artworks shown in the demo are reproductions of three masterpieces, all preserved in the Galleria Palatina, i.e.: 1) "The Madonna del Granduca" by the Italian Renaissance artist Raphael, probably painted in 1505. 2) "Mary Magdalene" by the Italian baroque artist Artemisia Gentileschi, painted around 1616-1617, and 3) "The Penitent Magdalene", a painting of saint Mary Magdalene by Titian dating to around 1531-1533.

Figure 1: 1) "Madonna del Granduca" by Raphael. 2) "Mary Magdalene" by Artemisia Gentileschi. 3) "Penitent Magdalene" by Titian.

The flow of interaction between the users and the system is as it follows (in the video but we will replicate the setting during the demonstration at the conference): 1) User 1 approaches location 1 and stands in front of a reproduction of Raffello's "Madonna del Granduca". The user's face is detected and the visitor is labelled and enrolled in the system. The profile of his visit starts to be recorded (see Fig. 1.1); 2) User 1 approaches location 2 and stands in front of a reproduction of Artemisia Gentileschi "Mary Magdalene" (see

---

Fig. 1.2); 3) User 2 stands at location 3 in front of a reproduction of Titian's "Penitent Magdalene". The visitor is detected and enrolled as a new user in the system. Then he is joined by User 1 who is re-identified. Both their visit profiles are updated. 4) User 1, left alone, stands in front of the "Penitent Magdalene" by Titian (see Fig. 1.3), observing the painting. The system checks his level of sustained attention analysing the user gaze over a period of time and then it starts the audio-guide. The played audio compares the characteristics of the two Magdalene: the current, by Titian, and the other by Artemisia Gentileschi and present in the history of the visit of User 1. In the last scenario, 5) User 1, left alone, is observing the "Penitent Magdalene" by Titian. This time a different audio is started which illustrates to the visitor the statistics relative to the different emotive perception of the two artworks shown by the overall visitors, indicating the preference by age and gender on the basis of the emotions detected on the visitors' faces.

Figure 2: The system in action. The visitor is in front of the "Penitent Magdalene" by Titian. He is identified and enjoys the audio-guide explaining differences and similarities of this painting with another painting seen in the context of the visit: i.e. "Mary Magdalene" by Artemisia Gentileschi. The user is re-identified and the history of his visit is updated.

3 CONCLUSIONS

In this demo we presented a re-identification system demonstrated in the context of a real-time audio-based application contextual to a museum visit simulated for two users in the Galleria Palatina in Florence. The possibility of tracking the user along the visit, monitoring his reactions and interests together with those of the other visitors, in this way opening up to the social context, allows the design of digital fruition scenarios that can improve the experience and the process of user learning, also exploiting gamification paradigms.

REFERENCES

[1] Mica R Endsley and Daniel J Garland. 2000. *Situation awareness analysis and measurement.* CRC Press.
[2] Zinovia Stefanidi, George Margetis, Stavroula Ntoa, and George Papagiannakis. 2022. Real-Time Adaptation of Context-Aware Intelligent User Interfaces, for Enhanced Situational Awareness. *IEEE Access* 10 (2022), 23367–23393.