Demo Abstract: Automatic Face Recognition Adaptation via Ambient Wireless Identifiers

Chris Xiaoxuan Lu
Peijun Zhao
Department of Computer Science
University of Oxford

Bowen Du
Hongkai Wen
Department of Computer Science
University of Warwick

Andrew Markham, Stefano Rosa, Niki Trigoni
Department of Computer Science
University of Oxford

ABSTRACT

Face recognition is a key enabling service for smart-spaces, allowing building management agents to easily monitor 'who is where', anticipating user needs and tailoring their local environment and experiences. Although facial recognition, especially through the use of deep neural networks, has achieved stellar performance over large datasets, the majority of approaches require supervised learning, that is, to be trained with tens or hundreds of images of users in different poses and lighting conditions. In this paper, we motivate that this enrollment effort is unnecessary if the smart-space has access to a wireless identifier e.g., through a smart-phone's MAC address. By learning and refining the noisy and weak association between a user's smart-phone and facial images, AutoTune can fine-tune a deep neural network to tailor it to the environment, users and conditions of a particular camera or set of cameras.

CCS CONCEPTS

• **Human-centered computing** → Ubiquitous and mobile computing; • **Computing methodologies** → Machine learning;

KEYWORDS

Adaption of Learning Systems, Face Recognition

ACM Reference Format:
Chris Xiaoxuan Lu, Peijun Zhao, Bowen Du, Hongkai Wen, and Andrew Markham, Stefano Rosa, Niki Trigoni. 2018. Demo Abstract: Automatic Face Recognition Adaptation via Ambient Wireless Identifiers. In *The 16th ACM Conference on Embedded Networked Sensor Systems (SenSys '18), November 4–7, 2018, Shenzhen, China.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3274783.3275191

1 INTRODUCTION

Accurate and robust person identification is a key component of smart spaces, e.g., offices and buildings for determining who is where. Knowing this information allows a building management system to enable a wide range of ambient services. With recent advances in deep learning [2], face recognition systems enjoy increasing adoption in smart spaces, from personalization and recommendations systems to security and efficiency monitoring.

A vast amount of research over the past decade has gone into designing tailored systems for facial recognition and with the advent of deep learning, progress has accelerated. As an example of a state-of-the-art face recognizer, FaceNet [3] achieves extremely high accuracies (e.g., 99.5%) on very challenging datasets through the use of a low dimensional embedding, allowing similar faces to be clustered through their Euclidean distance. Although FaceNet and similar approaches can operate at near-perfect levels of performance, they suffer from two overarching issues in indoor (wild) environments. Firstly, viewing angles and image sizes from smart cameras show even higher levels of variability than experienced in the pre-trained network on public datasets. Secondly, and more critically, those pre-trained network are *supervised* training methods which use large numbers of images (e.g., 300) of the same person in *different scenes* and angles to train an accurate clustering and embedding. To build such a database without access to labelled data would require a significant amount of subject enrollment effort.

We seek a solution to automatically tailor a pre-trained face recognition model on public datasets to adapt to the new environments with *zero* user enrollment effort. To this end, we exploit the fact that users are typically colocated with their mobile devices e.g., smartphones and fitness monitors. To provide ubiquitous connectivity, these devices have some form of wireless interface e.g., BLE, WiFi, cellular. These provide a unique identifier, ranging from the hardware level (e.g., IMEI or MAC addresses) to the network authentication level (e.g., usernames). Our aim is to use these identifiers to crowdsource a set of faces and a set of identifiers to refine a pre-trained classifier with the goal of improving its performance over time. The main challenge to update the model is that the binding between an image and a wireless identifier is not reliable, as multiple users could be present in the same area, a user may not be carrying their device or they may have changed their device. The weak relationship between devices and user presence is also one factor that prevents the use of WiFi sniffing alone to identify people in smart spaces. In this work, we present `AutoTune`, a system which can be used to gradually improve the performance of facial recognition systems in the wild, with zero user effort, tailoring them to the visual dynamics of a particular smart space.

2 AUTOTUNE SOLUTION

2.1 Key Terms

Event: An event is the setting in which people interact with each other in a specific part of the environment for a given time interval. It is uniquely identified by three attributes: effective timeslot, location, and participants. An instance of an event might be a meeting in a conference room, exercise in a gym, or a lunch in the canteen.

SenSys '18, November 4–7, 2018, Shenzhen, China

ACM ISBN 978-1-4503-5952-8/18/11…$15.00
https://doi.org/10.1145/3274783.3275191

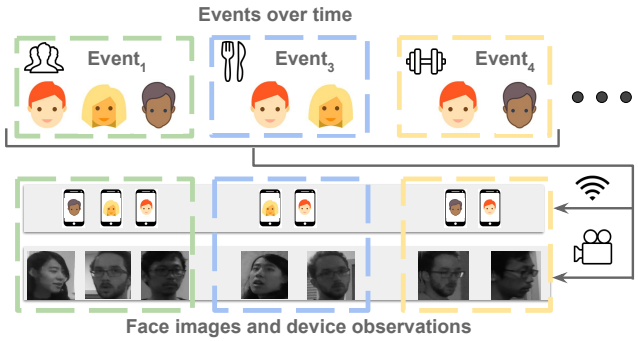

Figure 1: An event has two types of observations. The face observation are images cropped from surveillance videos and the device observation are sniffed MAC addresses.

Observations: A smart camera and a WiFi sniffer are assumed to sense the identities in an environment. With these two sensor modalities, two types of observations can be attached to an event: i) cropped face images detected by the cameras, and ii) device observations (MAC addresses) which give weak user attendance or presence information through WiFi sniffing. Fig. 1 illustrates the relationship between identity observations and events through examples.

Pre-trained Model: A pre-trained model is used to transform the face images into discriminative features or representations, thereby aiding accurate clustering or classification. In this work, a deep neural network (DNN) for discriminative face representation is trained by a model provider, e.g., Facebook or Google, who are able to access a large number of annotated face images.

2.2 Workflow

We are now in a position to introduce AutoTune, which assigns IDs to images from noisy observations of images and WiFi MAC addresses, and uses such learned ID-image associations to tune the pre-trained deep face representation model automatically. Our approach is based on two key observations: i) although collected by different modalities, both face images and device MAC addresses are linked with the identities of users who attend certain events; and ii) the tasks of model tuning and face-identity association should not be dealt with separately, but rather progress in tandem. Based on the above insights, AutoTune works as follows (see Fig 2). i) *Event Segmentation:* Given the face images and sniffed WiFi MAC addresses, AutoTune first segments them into events based on the time and location they were captured. ii) *Cross-Modality Clustering:* Then the face images are clustered based on their appearance similarity computed by the pre-trained face representation model, and also taking into account information on device attendance in events. iii) *Cluster Labeling:* Each image cluster should broadly correspond to a user, and the cluster's images are drawn from a set of events. AutoTune assigns each cluster to the user whose device has been detected in as similar as possible set of events. iv) *Visual Model Update:* Once images are labeled with user identity labels, AutoTune then fine-tunes the pre-trained face representation model. v) *User Attendance Update:* We further use the cluster labels to update our belief on which device (MAC address) has participated in each event. With the two models updated, we are in a position to iteratively repeat the steps of clustering, labeling, and model updates, until the changes in the user attendance model become negligible.

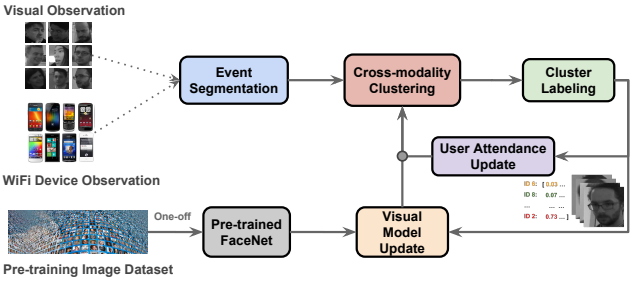

Figure 2: System Workflow. AutoTune consists of 5 steps i) Event segmentation ii) Cross-Modality clustering iii) Cluster Labeling iv) Visual Model Update v) User Attendance Update. AutoTune sequentially repeats the above five steps until the changes in the user attendance model are negligible.

Table 1: Evaluation Results on Two Testbeds

	Precision		Recall		F1 Score		Accuracy	
	UK	CHN	UK	CHN	UK	CHN	UK	CHN
AutoTune	0.90	0.85	0.87	0.91	0.87	0.88	0.98	0.91
WebCrawl	0.18	0.16	0.20	0.15	0.18	0.15	0.31	0.27

3 EVALUATION

We deployed AutoTune in UK and China. The UK testbed consists of a large office, a meeting room and a kitchen. The testbed in China is a common room in an university. 43 3-hour events and and 102 2-hour events are recorded in two different places. 13, 176 and 7, 495 face images of 31 and 36 people are used to examine AutoTune in two testbeds respectively. Tab. 1 shows the performance of AutoTune compared with the web-crawl approach [1].

4 DEMONSTRATION

An annotated video is available online[1]. The demonstration explains the intuition and workflow of AutoTune and demo results in real-world scenes. The video showcases the performance gain of AutoTune compared with baselines and allows the participants to develop a feel for the concept behind. A face verification system developed by AutoTune will be lively showcased to justify its usefulness in real-world application scenarios

REFERENCES

[1] Lacey Best-Rowden, Hu Han, Charles Otto, Brendan F Klare, and Anil K Jain. 2014. Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security* (2014).
[2] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. 2015. Deep Face Recognition.. In *BMVC*, Vol. 1. 6.
[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815–823.

[1] https://youtu.be/jRNwws8soHg