Chalmers University of Technology

CHALMERS
UNIVERSITY OF TECHNOLOGY

Introduction to data science and AI

DAT405

Facial and speech recognition

An ethics essay

Authors:
Pauline Näsblander
Albin Ekström

March 9, 2022

Innehåll

1 Introduction

Facial- and speech recognition are widely used techniques in many different areas. While face detection can be used for unlocking phones, finding missing people, law enforcement, etc. speech recognition allows for devices to understand spoken commands resulting in smart assistants like Alexa or Siri.

Addressing facial recognition first works by reading the geometry of your face from a picture, noticing key factors such as the distance between the eyes and from forehead to chin. Then the mathematical representation of your facial features is compared to a database of known faces [1]. Test issued by the National Institute of Standards and Technology in April 2020 said that the best performing facial recognition algorithm had an error rate of only 0.08 % [1].

Speech recognition works by digitizing speech samples into discrete segments of spectrograms and then dividing them further into time steps using a short-time Fourier transform. Every spectrogram is then analyzed and an algorithm predicts the probability of all words in the programmed language and also considers the previous word and the likeliest next word based on knowledge on the language at hand before it decides what was said [2].

2 Face recognition

The face is a unique part of the human body and is strongly connected to people's identity. Therefore facial recognition yields extensive power by being able to identify people [3]. While face recognition allows for incredible socially beneficial possibilities, it will also, almost inevitable, lead to privacy intrusions as you can identify people wherever and whenever making it almost impossible to remain anonymous.

One heavily debated possibility of the face recognition technique (FRT) is to use it in law enforcement. This would enable law enforcement agencies to match photographs of suspects with surveillance photos [4]. This in turn makes it easier to identify and convict criminals which could result in a safer society. However, there are some concerns to be raised when debating face recognition uses in law enforcement purposes, not least the invasion of citizen privacy. While advocates of FRT usage argue that the privacy losses induced by this technology are minimal and out weighted by large security benefits, the opponents argue that the privacy loss is more significant than the advocates see and that the increase in security is

overestimated. In addition, the opposition argues that an incorrect match could lead to an innocent being arrested which would be a huge violation of individual freedom, and that the usage of face recognition in law enforcement is therefore not only a question of privacy intrusion [4].

Using FRT on cameras recording public places and the citizen's passing allows for, in addition to more efficient police work, the finding of missing people. In 2019 there were almost 200 000 missing children in India and in the time between introducing face recognition and the year 2019 the police have found 2930 missing children [5]. However, having FRT applicable in public areas also enables the government to keep track of all citizens. The Chinese state uses face recognition as one element in China's tracking efforts where they are using social credit scores to reward and punish their citizens [6].

The Chinese facial recognition system keeps track of almost all citizens including people as young as 9 years old. In addition, the Chinese government is accused of using this database and the face recognition technique to commit barbarity towards Uyghur Muslims [7]. As this harmful exploitation of FRT to persecute certain ethnicities is, if true, a result of decisions made by the Chinese government political restrictions on the matter would not be made by this government. According to Madiega and Mildebrath [8] there are very few legally binding rules applied to FRT in China thus far.

Addressing political regulations of facial recognition, the EU has produced some regulations of FRT in order to protect certain fundamental rights but different actors have questioned the current EU-frameworks effectiveness in doing this [8]. The above-mentioned lack of political regulations in some countries and lack of effectiveness in existing regulations in others compels one to ask themselves whether the ultimate responsibility of the effects of FRT must lie with the manufacturer's choice to release such a product.

3 Speech Recognition

Already in the early 60s, IBM introduced their first speech recognition program that could recognize 10 words [9]. Today speech recognition, also known as automatic speech recognition (ASR) has become a part of basically everyone's life. We find it in technologies such as smartphones, smart speakers, cars, and in areas such as healthcare, sales, and security. There are generally two types of ASR, the first is the usual speech-to-text recognition, the second is voice recognition which specializes in connecting a voice to a specific individual.

As technology has gotten better AI, Machine Learning, and Big Data have made ASR invaluably much better than it was at the beginning of the 60s. Today Google's speech-to-text algorithms support over 125 languages with additions and improvements for several different languages (e.g. English, Spanish, French, and Russia) [10]. Researchers think that the ASR market will be worth USD 24.9 billion by 2025 [9].

3.1 Is my phone listening to me?

We've all heard something like "When I talked with my friend yesterday about our hike this weekend. I immediately saw that all the ads on my Facebook page were about sleeping bags and other hiking equipment. I've never googled anything about hiking.". Is it reasonable to believe that our phone is listening to us all the time? Today the answer is *yes*, but it's a bit more complex.

Your phone (and all other "smart ASR") uses what's called profanity filtering to listen to words or phrases (e.g. "Hello Siri", "Hi Alexa", "Ok Google") that will trigger the whole ASR algorithm. In order to catch the trigger words the phone has to listen all the time. But it doesn't record everything it hears, it only activates when a hot word is spoken, but when you agree to the terms and conditions you often allow the service to record what you say after e.g. "Hi Siri" is said. Apple then selects a small portion of all recorded conversations to analyze and improve their algorithms, but sometimes Siri gets activated at inappropriate times e.g. when people having sex or private business conversations. If you've agreed to the terms some of these conversations can/will be passed on to third-party analyzing companies [11]. For most people, the ethical dilemma isn't that Apple record your conversations with Siri to improve their services. It's when the data is sold to third-party companies beyond your control without ever knowing what's being sent and if the conversation is inappropriate. Talking with Siri or Google Assistant is no other than typing in the search bar [11]. Your data is collected and then used to improve services but most times sold, that's why the services are "free". Another problem is that there is no active denial from these companies that they aren't spying on you without knowing. As Jacob Leon Kröger and Philip Raschke write "the spying fears were not disproved so far, neither by device manufacturers and ecosystem providers nor by the research community." [12].

3.2 Security Agencies and Governments

National Security Agency, NSA in the USA is the world's largest eavesdropping organization with over 40 000 employees and their super computers handle over a

quarter of a billion encrypted and not messages/conversations every day all over the world [13]. They've been known for decades to spy on domestic and foreign powers, governments and authorities. Today we carry around the best spying equipment ever invented in our pockets everywhere we go all the time; our phones. And that's the world we choose to live in, we were not forced to carry around our phone, we want to [14].

On the other, side NSA claims to have stopped terror attacks, helped other governments with political issues, and prevented war due to their eavesdropping operations. E.g. NSA claims their technology "was instrumental in helping to identify Saddam Hussein after the invasion of Iraq" [15]. Some even claim that NSA knows who you are only by the sound of your voice [15]. The Swedish Security Service also uses eavesdropping to identify and prevent terror gatherings and/or attacks [14], but not so often through its own sources but through the NSA services [13], because the regulations are much stricter here in Sweden than the USA. However, after the revelations/leaks from Snowden 2013 when he handed documents to the Guardian about, inter alia, NSA's high refined ASR, more regulations has come to place and NSA is not allowed to follow the same path they did before, according to legislative changes [15].

Referenser

[1] S. Symanovich. (2021) What is facial recognition? how facial recognition works. [Online]. Available: https://us.norton.com/internetsecurity-iot-how-facial-recognition-software-works.html

[2] S. Linguae. (2021) How does speech recognition technology work? [Online]. Available: https://summalinguae.com/language-technology/how-does-speech-recognition-technology-work/

[3] E. Selinger and B. Leong, "The ethics of facial recognition technology," January 2021.

[4] P. Brey, "Ethical aspects of facial recognition systems in public places," vol. 2, pp. 97–109, May 2004.

[5] M. Wendorf. (2019) Facial recognition technology is being used to find missing children. [Online]. Available: https://interestingengineering.com/facial-recognition-technology-is-being-used-to-find-missing-children

[6] C. Campbell. (2019) How china is using "social credit scores" to reward and punish its citizens. [Online]. Available: https://time.com/collection/davos-2019/5502592/china-social-credit-score/

[7] A. Ng. (2020) How china uses facial recognition to control human behavior. [Online]. Available: https://www.cnet.com/news/in-china-facial-recognition-public-shaming-and-control-go-hand-in-hand/

[8] T. Madiega and H. Mildebrath, "Regulating facial recognition in the eu," September 2021.

[9] I. C. Education. (2020) What is speech recognition? [Online]. Available: https://www.ibm.com/cloud/learn/speech-recognition

[10] G. Cloud. (2022) Language support. [Online]. Available: https://cloud.google.com/speech-to-text/docs/languages

[11] K. Bareckas. (2021) Is my phone listening to me? [Online]. Available: https://nordvpn.com/blog/is-my-phone-listening-to-me/

[12] R. P. Kröger J.L. (2019) Is my phone listening in? on the feasibility and detectability of mobile eavesdropping. [Online]. Available: https://www.researchgate.net/publication/334214258_Is_My_Phone_Listening_in_On_the_Feasibility_and_Detectability_of_Mobile_Eavesdropping

[13] J. Guillou, *Den som dödade helvetets änglar*. Piratförlaget, 2022.

[14] S. Radio, "Du kan inte gömma dig," 2018. [Online]. Available: https://open.spotify.com/episode/3hk7nvlum28cCkHG6KbTcT?si=33725cfb51f3447b

[15] A. Montag. (2018) The nsa knows who you are just by the sound of your voice—and their tech predates apple and amazon. [Online]. Available: https://www.cnbc.com/2018/01/20/the-nsa-can-recognize-you-by-just-your-voice-predating-apple-amazon.html