March 2022

Bootstrapped Identifier for Recognizing People in a Video

Anonymous

Bootstrapped Identifier for Recognizing People in a Video

ABSTRACT

This disclosure describes techniques that leverage the large volume of video captured by smart glasses to improve people identification. Bootstrapping off face recognition, the data collection capabilities of smart glasses or other devices are leveraged to train machine-learning models to utilize other input factors for people identification. Such factors can include, for example, gait, pose, silhouette, etc. The techniques generalize and expand people identification beyond face recognition, and, in doing so, hew closer to how humans recognize people.

KEYWORDS

- Smart glasses
- Face recognition
- Gait recognition
- Posture recognition
- Pose recognition
- Machine learning
- Bootstrap training

BACKGROUND

Many applications for smart glasses rely on the identification of people. In an example application, a user entering a crowded bar locates a friend using smart glasses that highlight the sought-after face from a group of several faces detected in the scene. To perform a task such as locating a friend, smart glasses generally rely solely on face recognition.

Smart glasses (and similar devices) obtain large volumes of visual data. For example, smart glasses can record video in a round-robin fashion, e.g., record continuously such that a most recent window of time is always available. Such continuously obtained visual data, however, is not utilized for people identification.

DESCRIPTION

This disclosure describes techniques that leverage the large volume of data captured by smart glasses (e.g., using built-in cameras) to perform superior people identification. Specifically, in addition to face recognition, people identification is performed based on signals such as gait (how people move), body posture or stance (height, pose, standing style, spinal curve), body shape or silhouette, etc.

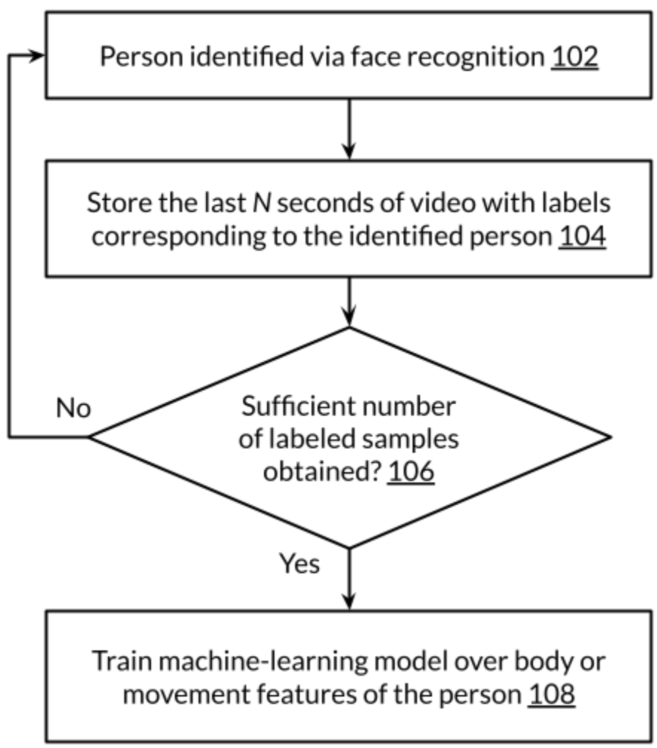

Fig. 1: Bootstrapped people identification using multiple signals

Fig. 1 illustrates bootstrapped people identification using multiple signals, per techniques of this disclosure. Face recognition is utilized to bootstrap the annotation of ground-truth samples as follows. When a person is identified via face recognition (102), the previous *N* (e.g., *N* = 5) seconds of video of the scene, as captured by smart glasses, is stored along with labels corresponding to the identified person (104).

Once enough samples have been obtained (106), e.g., 5-6 samples per person, a machine learning model is trained or fine-tuned over the entourage of labeled samples. The machine learning model learns an embedding, e.g., via large-scale classification or triplet loss, on the video sequence, where features such as body pose, gait, silhouette, etc., are extracted over time (108). Given the sequential nature of video samples, the machine learning model can be a sequence-based model such as a long short-term memory (LSTM) neural network.

The model is continuously trained and improved upon. Once enabled, it can utilize a broad variety of signals, e.g., face, gait, pose, silhouette, etc., to recognize people. For example, it can pick out in a crowded bar a friend whose back is facing the smart-glass user.

While the foregoing description refers to video captured via smart glasses and person identification via face recognition, gait recognition, body pose, or silhouette, the described techniques can be utilized to identify individuals based on any available video source. By utilizing these additional features, robust recognition is made possible, mimicking human recognition - humans recognize others based on multiple factors, not just face.

CONCLUSION

This disclosure describes techniques that leverage the large volume of video captured by smart glasses to improve people identification. Bootstrapping off face recognition, the data collection capabilities of smart glasses or other devices are leveraged to train machine-learning models to utilize other input factors for people identification. Such factors can include, for example, gait, pose, silhouette, etc. The techniques generalize and expand people identification beyond face recognition, and, in doing so, hew closer to how humans recognize people.