October 17, 2019

Face Recognition for Fast Information Retrieval and Record Lookup

Joseph Edwin Johnson Jr.

Rafael Blanes

Daniel Sheng

Arjun Narayanan

Face Recognition for Fast Information Retrieval and Record Lookup

Abstract:

This publication describes systems and techniques directed to taking a red, green and blue (RGB) image of a customer's face, identifying landmarks on the face, creating a face box region, thumbnailing the face box region, sending the thumbnail to a face recognition model where face data is embedded into a vector, and using the vector in a multi-dimensional tree searching algorithm to quickly look up or retrieve information or a record relating to the customer in a database. The face image capture may be performed using standard RGB camera technology to enable broad, cost-effective business adoption. Two-factor authentication may be employed to address potential RGB image spoofing. Embedding the vector may be implemented on local technology to minimize network data transmissions of face images and increase lookup speed.

Keywords:

RGB image, face recognition, machine-learned model, vector, hash, float, multi-dimensional tree search, k-d tree search, record lookup, two-factor security, authentication

Background:

Personal identification and authentication technology/processes continue to improve and gain adoption for different business purposes. For example, some face recognition models today have quite low false acceptance rates and/or false recognition rates with respect to discerning one individual face from another, making the models fairly precise and secure.

Due to cost and availability of hardware, facial recognition is frequently performed by using a standard red, green, and blue (RBG) camera to obtain a simple RGB image of a face. However, RGB image facial recognition can be fairly easily spoofed unless special precautions

are taken. For instance, utilizing anti-spoofing techniques that include using near infrared (NIR) cameras to get both two-dimensional (2D) and three-dimensional (3D) information, using a skin classifier to detect spoof masks, using an eyes open model to detect that the customer is not sleeping or trying not to look, using a gaze model to make sure the customer is paying attention, and the like.

However, deploying such dedicated anti-spoofing measures is expensive and requires special hardware (*e.g.*, NIR flood illuminators, dot projectors, a secure hardware layer to prevent spoofing, etc.) so is not financially conducive to small business operations or generally self-contained environments like kiosks. Similarly, other biometric sensors for identifying or authenticating persons, such as dedicated fingerprint sensors, voice sensors, or eye/retina scanners, etc., are also expensive or otherwise not widely available.

But in today's competitive environment, businesses need affordable yet fast identification and authentication for their customers. For example, it would be beneficial for a medical office or a retail store to deploy an affordable identification and authentication system that quickly and automatically identifies and authenticates a customer (*e.g.*, patient, client, user) and takes an appropriate action (*e.g.*, pulls the appropriate records for medical processing, associates discount reward card opportunities for purchases).

Description:

This publication describes systems and techniques directed to face recognition using a machine-learned model trained to classify captured images to identify a customer's face, embedding face data into a vector (*e.g.*, 128 floats) like a hash, and using the vector in a multi-dimensional tree searching algorithm (*e.g.*, k-dimensional tree (k-d tree)) to quickly look up or retrieve information or a record in a database. The face image capture may be performed using

standard RGB camera technology to enable broad, cost-effective business adoption. Simple two-factor authentication may also be employed to address potential RGB image spoofing. Embedding the vector may be implemented on modern local technology to minimize network data transmissions and increase lookup speed.

Figure 1 illustrates an example of this system and technique in a typical business context, including by using a standard, cost-effective RGB camera (*e.g.*, a webcam) to capture a customer's face image in coordination with using a machine-learned model trained to classify captured images to identify a customer's face deployed to a computer-readable medium (CRM) of a user equipment, such as a workstation, tablet computer, smartphone, and the like.

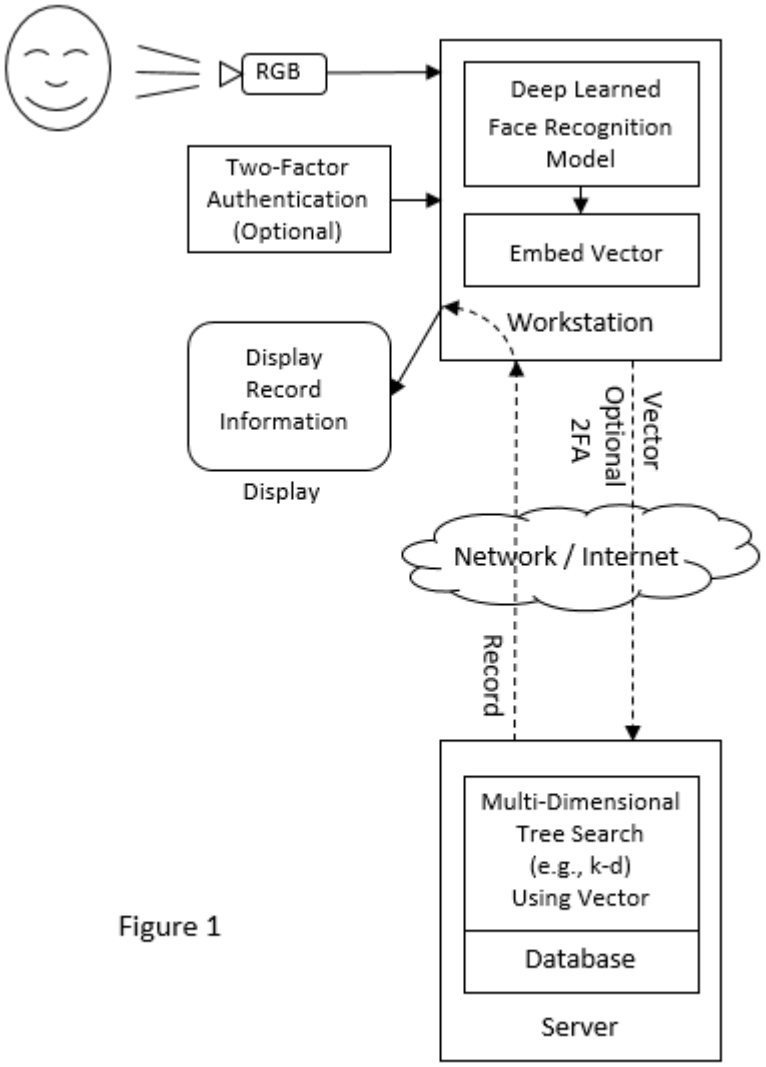

Figure 1

Example business settings may be in a doctor's office, mall shop, retail store, kiosk stand, or similar environment where either a strong standalone biometric is not required and/or where a two-factor authentication mechanism is sufficient for protecting the customer's information. It is beneficial for these businesses to avoid requiring that the customer manually fill out tedious forms such as in a doctor's office to "check in," requiring a customer to rummage through personal belongings to pull out an identification card (*e.g.*, a rewards card), or manually input a phone number or other personally identifiable information for obtaining discounts at a retail store or

grocery store. In this context, the webcam is the first step to capture the customer's facial image, possibly at a strategic location such as entering the business, to begin the process to identify and authenticate the customer quickly in order to associate the person with digital information or records on file with the business.

In this shared computer database scenario where an RGB camera is available with a machine-learned model trained to classify captured images to identify a customer's face from RGB camera images, the customer's face can be used to perform fast information retrieval, *e.g.*, detect that this is a specific person, with great precision (*e.g.*, at least 1/50k false accept rates). However, because an RGB image can be spoofed (by simply holding up a picture of someone else, for instance), without more advanced anti-spoof countermeasures, it is not necessarily well suited for an authentication mechanism by itself. But an RGB image can be used as a quick record lookup (*e.g.*, identifying that this is likely the person in question) in place of a username or any other biometric on record.

If the business desires a more secure identification beyond the RGB face recognition, a second factor authentication may be implemented (*e.g.*, pass phrase, personal identification number (PIN)). On the other hand, the second factor/pin authentication would not be required if stronger biometric detection scenarios are employed, such as advanced hardware (*e.g.*, NIR flood and dot projectors) and trusted anti-spoof models (*e.g.*, an office or kiosk equipped with depth/NIR and anti-spoof/mask recognition, eyes open/gaze detected, etc.). Two-factor authentication may also be unnecessary in business settings where the security of the identification and authentication models are naturally reinforced by a human proctor observing the transaction (*e.g.*, proctors/operators not allowing the customer to perform the business transaction or operate the

kiosk using fraudulent efforts such as wearing a custom silicone mask trying to imitate another person, or showing a photo of another person, etc.).

The image lookup process starts with taking an RGB image of the customer's face, where landmarks on the face (such as eyes, ears, mouth, nose) are identified and used to create a face box region, which is then thumbnailed (perspective distortion corrected, face roll normalized, brightness normalized, etc.). The thumbnailed image is then sent to the machine-learned model trained to classify captured images to identify a customer's face (preferably a deep-learned face recognition model having low (*e.g.*, 1/50K or better) false accept rates or false recognition rates with respect to discerning one individual face from another, making the model precise and secure).

Next, the machine-learned model embeds a vector (*e.g.*, 128 floats) with the captured facial data to identify that person's face. This vector now serves like a hash to be used for fast record identification processing in a multi-dimensional tree search.

If the machine-learned model can be run efficiently and quickly on the local workstation itself to embed the vector, then the larger amount of RGB face data would not have to be transmitted over the network to a server. Only the vector would be transmitted -- meaning no customer image is sent off the device over the network, thereby improving personal security, minimizing network transmission data, and increasing record lookup speed. Additional security considerations may include encrypting the vector during upload over the network, using SSL over https, and/or employing a second factor security mechanism (*e.g.*, PIN, passphrase).

The speed of running inference (the face identification) on the machine-learned model might take several seconds to run on a standard central processing unit (CPU), but many business affordable modern higher-end computers or workstations are sold with graphic processing units (GPUs) which may have parallel computing cores that can speed up the inference stage of the

machine-learned model (taking it from seconds to milliseconds). Additionally, in higher-end platforms, an artificial intelligence (AI) accelerator application-specific integrated circuit may be deployed, which would be dedicated hardware for running inference over fixed point models, providing maximum performance without the cost of an expensive and dedicated high-end graphics card needed to run the floating point models quickly. Thus, if the local workstation processing power is sufficient for running the machine-learned model, then there is no need to transmit the customer's image over the network.

Once the vector is transferred over the network to the server, the vector is then processed with a k-d tree (where k is 128 in this case) in an example system, or similar hash or multi-dimensional tree search algorithm, to efficiently and quickly lookup the associated personal record or other information in the database at scale. The identified record and information is then returned back over the network to the local workstation to be displayed or used by the business as appropriate.

Figure 2 illustrates an example of where the local workstation's processing power is insufficient, then the customer's image may be transmitted over the network to be processed by a machine-learned model deployed on a higher powered server (*e.g.*, a remote server or cloud processing environment). In this context, additional security considerations may include encrypting the RGB data during upload over the network, using SSL over https, or employing a second factor security mechanism (*e.g.*, PIN, passphrase).

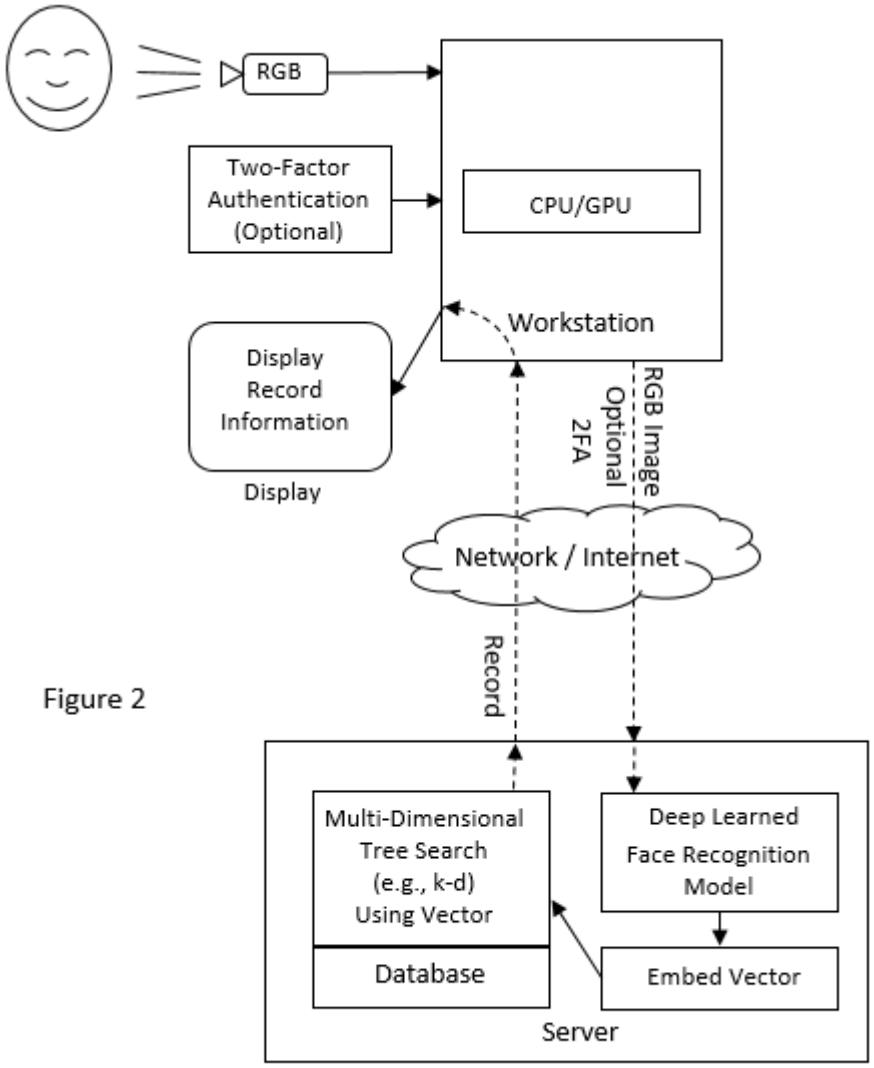

Figure 2

This authentication as a service business model (*e.g.*, Single-Sign-On using RGB + optional pin/passphrase), enables a fast returned authenticated and authorized identification (ID) of the customer who just signed in (with their face + optional pin/passphrase). That ID is then used by the business to quickly bring up the customer's relevant records, or other secure content. Alternatively, this process can be used by businesses in many other scenarios, for example, to simply log users into their secure workstations, as methods of authenticating the user for payroll systems (*e.g.*, clock them in at work), or for reward card recognition scenarios, etc.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (*e.g.*, photographs, information about a user's social network, social actions or activities, profession, a user's preferences), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

References:

[1] US 20150262496 A1. Multimedia educational content delivery with identity authentication and related compensation model. Date of Filing: March 16, 2015.

[2] US 20150023602 A1. Fast recognition algorithm processing, systems and methods. Date of Filing: July 15, 2014.

[3] Pabiania *et al.*, Face Recognition System for Electronic Medical Record to Access Out-Patient Information, Jurnal Teknologi, Vol. 78, No. 6-3, December 13, 2015. https://jurnalteknologi.utm.my/index.php/jurnalteknologi/article/download/8935/5312.

[4] Face Recognition System, Edom Technology Co, Ltd., July 16, 2019. https://www.edomtech.com/en/solution/ins.php?index_m_id=7&index_id=38.