December 2021

UNIQUE GESTURE TRIGGERED MESSAGE/SOS ALERTING USING REAL-TIME VIDEO ANALYTICS

Nagendra Kumar Nainar

Gonzalo Salgueiro

Hugo Latapie

UNIQUE GESTURE TRIGGERED MESSAGE/SOS ALERTING USING REAL-TIME VIDEO ANALYTICS

AUTHORS:
Nagendra Kumar Nainar
Gonzalo Salgueiro
Hugo Latapie

ABSTRACT

The United States government, as well as other countries, are actively working on implementing Smart Cities. While there are various use cases that are being explored for Smart Cities, such as environment monitoring, transportation, etc. using advanced technologies to improve the overall safety of inhabitants, it is also important to improve the overall security of the inhabitants. There are various standardized signaling/alerting techniques, such as signaling "SOS" or the like using physical gestures etc. that are often used to signal emergency situations. Proposed herein are techniques to utilize advanced technologies, such as Deep Fusion Video analytics, along with facial behavioral and gesture analysis techniques, to configure unique, per-user-based "Gesture to Signal Mappings" for that can be used to trigger SOS/emergency and/or other types of alerts/actions upon detecting gestures of a given user.

DETAILED DESCRIPTION

Deep Fusion Video Analytics introduces additional intelligence to camera feeds using deep learning algorithms that can be used for various use cases such as detecting crowd density, detect road accidents and alert officials, tailgate detection, etc. As proposed herein, such techniques can be extended to facilitate providing unique, per-user-based "Gesture to Signal Mappings" (GSMs), for various emergency signals that may be signaled by individuals.

As shown in Figure 1, below, there are various internationally agreed upon sign language techniques that can be used for alerting/signaling SOS messages using physical gestures.

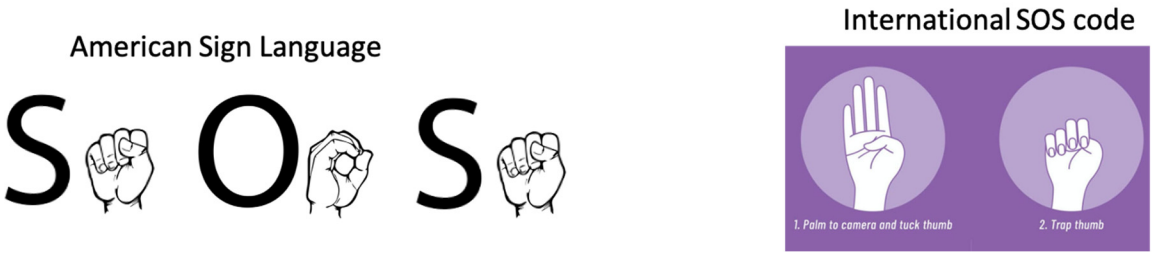

Figure 1: Standardized Emergency Gestures

Alternately, in some instances, domain-centric signaling may be used to signal an emergency situation. For example, in a hospital setting, placing a sticker on the bottom of a specimen sample container might be used to signal that a person is in danger and needs help.

One challenge with signaling an emergency using the above approaches is that, in many cases, the offender may be aware of such signaling and so using such mechanisms may put the victim in more danger.

In order to address such potential issues, this proposal provides techniques that leverage Deep Fusion Video analytics along with facial behavioral and gesture analysis in order to provide per-user-based unique "Gesture to Signal Mappings" (GSMs) that can be used to trigger SOS and other message alerts.

Broadly utilizing techniques herein, a user can create multiple gestures and map them to unique signal/message alerts (e.g., raising the left eyebrow of a user three times can be mapped to an SOS alert, etc.). Using OpenFace or other similar facial recognition tool, a model can be trained based on facial recognition inputs and gesture recognition inputs, such that various user-specific measurements can be learned in order to create GSMs on a per-user basis. The GSMs can be provided to a camera/security system which, upon detecting a particular GSM, can use the mapping to identify and trigger the corresponding signaling/alert action (e.g., alert officials, send message to a parent, etc.).

For the facial and gesture recognition training, facial coordinates and gestures can be captured to train a deep neural network (DNN) –based model, as illustrated in Figure 1, below.

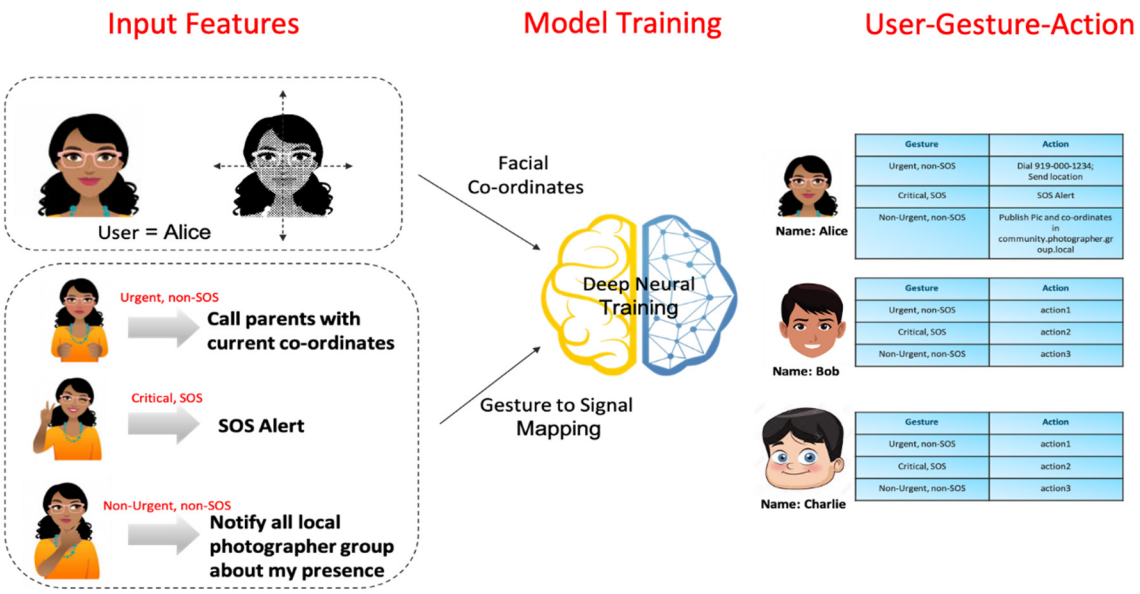

Figure 1: Facial and Gesture Recognition Training

As shown in Figure 1, facial coordinates are used to train the model. This is similar to how mobile phones are trained for facial recognition-based device unlocking. For each user, one or more gestures are used to further train the model. The gestures may be related to different fixed postures (e.g., closed fist near the face) and/or can be a continuous motion (e.g., raising an eyebrow twice/thrice, opening and closing a fist, etc.). These are only a few examples, as other gestures can be envisioned. The DNN model is trained based on various such gestures and the gestures are mapped to a given user.

For gesture to signal mapping, each gesture for each user can be mapped to different types of semantics, as illustrated, for example, with reference to Figure 2, below.

Name: Alice

Gesture	Intent	Action
Gesture1	Urgent, non-SOS	Dial 919-000-1234; Send location
Gesture2	Critical, SOS	SOS Alert
Gesture3	Non-Urgent, non-SOS	Publish Pic and co-ordinates in community.photographer.group.local

Name: Bob

Gesture	Intent	Action
Gesture1	Urgent, non-SOS	action1
Gesture2	Critical, SOS	action2
Gesture3	Non-Urgent, non-SOS	action3

Figure 2: Per-User Gesture to Signal Mappings

As shown above in Figure 2, gestures can be mapped to signal intent and an associated action. Gestures can be used to signal an SOS or can be used for other types of signaling. The gestures are unique to each user and so can be difficult for potential offenders to detect for a user signaling a particular alert.

Following configuration of such mappings, a user may use the relevant gesture at any point in time such that the gesture can be captured via a smart camera system using the Deep fusion analytics. Consider an illustrative environment, as shown below in Figure 3, within which techniques of this proposal may be realized.

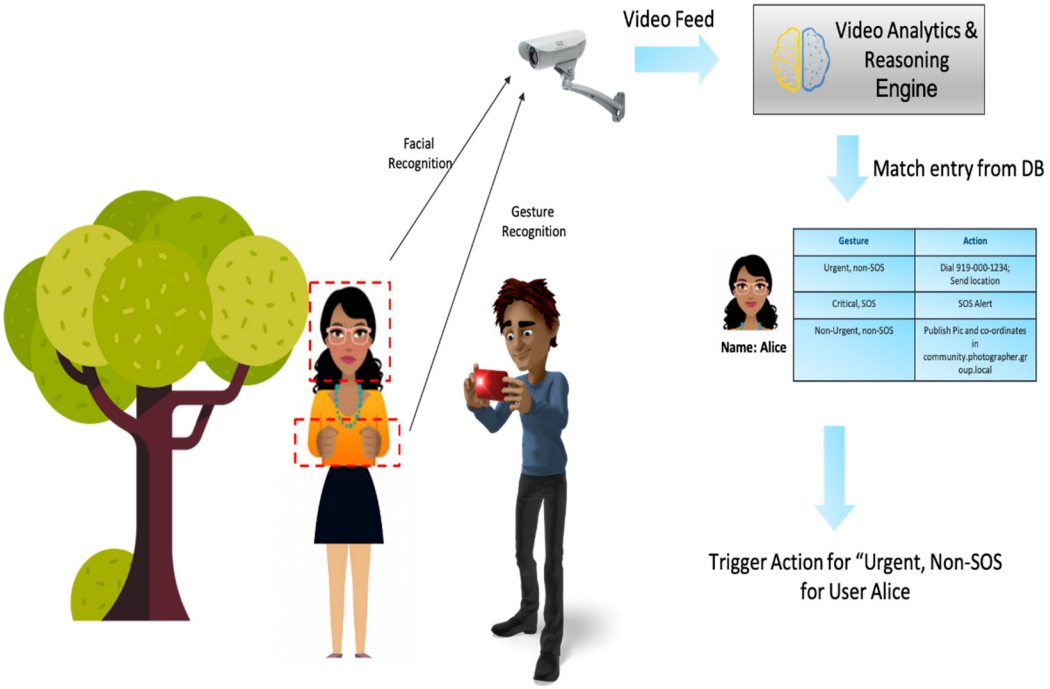

Figure 3: User-triggered Alert Environment

Within the environment as shown in Figure 3, user facial coordinates are used to identify the user and the relevant gestures are used to identify what the user is signaling. Upon matching a known gesture for the user, the relevant action is then triggered.

In one illustrative example, such techniques can be used to dynamically signal any video recording (privacy) consent. For example, consider a scenario in which a user enters a video hall in which cameras are continuously recording. The user, at any point in time, may use the gesture to agree/disagree recording consent. In such a scenario, the video recording can blur the face of the user.

Accordingly, techniques proposed herein provide for the ability to utilize advanced technologies, such as Deep Fusion Video analytics, along with facial behavioral and gesture analysis techniques, to configure unique, per-user-based "Gesture to Signal Mappings," which can be used to trigger SOS/emergency and/or other types of alerts/actions upon detecting gestures of a given user.