Avoid a fight in venues with Nonlutte app

Student Name and Number

Phillips Owie / 100711252

Christiano Varella /100761213

Samantha Yushan Tan/ 100731528

Sai Krishna

Table of Contents

SCENARIO

The situation repeats itself too many times and it exceeds individual motives and focus on group attention: fights. Generally, one of the reasons a fight occurs is by the simple fact of how many people are in a venue. Big crowds of 1,000 people have all that is necessary for a fight to break out. Due to sociologically reasons, only one fight may happen in such crowds. More than 1,000 people per venue, more fights are likely to occur. Venues that hold less people, from 50 to 200 people, may also experience one fight per night.

With the possibility of having one fight per gathering in Mardi Gras, sport victory parades, street festivals, parties, clubs and other entertainment venues and not knowing when a fight may start, it's crucial for mayors and venues owners to prevent fights from ever happening in order to have all patrons safe and with the intention of always keep coming back.

PROBLEM SUMMARY

Venue owners cannot count how many fights they have seen every day at their grounds. It happens when nobody is expecting and there are always risk to safety of the crowd. And specially for venue owners, they also have to deal with damages to the property. In clubs, for example, table, chairs, lamps, glasses can all become part of the brawl. The crowd that is not involved in the fight, start to leave and quickly one night of profit becomes one night of expenses. Normally clubs, for example, are only busy on the weekends, which makes a fight affect their month income.

PROBLEM STATEMENT

Provide a solution that integrates with the surveillance cameras of the venue, and with expression detection, advise security guards to move towards the possible "epicenter" of the fight, in order to prevent it from ever happening.

APPROACH

Owning an entertainment venue or holding big crowds on the street takes a lot of investment. For this kind of business to keep on thriving, it's mandatory that fights do not happen in any time.

Just like in the movies, where a crime can be avoided, our solution **nonLutte (noFight)** will capture video, identify faces and anger/fear expressions, which will trigger an alert for security personnel to move closer to where a fight may happen.

KEY PERFORMANCE METRICS

KPI	Definition
- Percentage of expressions detected	Measures the expression detected against the number of faces detected
- number of expressions detected with a combination of anger and fear	This measures the number of possible fight scenarios avoided
- online real-time availability	Measures availability of
- number of fights avoided	Measures number of fights avoided by a combination of faces with anger and fear detected
- application processing time	Measures how fast it processes images to detect a combination of anger and fear
- Customer satisfaction index	Measures customer satisfaction per event covered through feedback rating system

PROJECT FEATURES AND ADVANTAGES

Expression recognition technology has been implemented in many commercial applications and has a very good market prospect. Emotion recognition is a person's subjective or objectively hidden mental state. Detecting this mental state has nothing to do with a person's privacy, but hopes to better use this technology to serve society and individuals.

Emotional intelligence analysis technology uses human biological principles to perform non-contact intelligent analysis of the potential emotions of people in video images, that is, to capture micro-vibration images of the subject's head and neck through a camera, and use related algorithms to collect micro-vibrations of the head and neck muscles. Inductive, statistical, and comparative analysis of the sampled data yields the corresponding emotional index, including attack power, stress, tension, suspiciousness, balance, self-confidence, self-control, energy, reaction speed, and degree, and display it in the form of data reports and graphics user. There are currently four major characteristics:

1. Quick and easy. Check each person for less than 5 seconds

2. accurate and objective. Different thresholds can be set for objective screening.

3. Friendly shielding. No contact and no interference.

4. Intelligent early warning. Real-time dynamic early warning function.

USER INTERFACE

The interface will present the main cameras that are being used at the venue that cover the entire area. Whenever the application detects an angry face, a message will be displayed to the user stating the security point closest to the location of the camera was alerted.

DATA ASSUMPTIONS, LIMITATION & CONSTRAINTS

Assumptions

Suppose there are six discrete basic emotions (anger, fear, disgust, surprise, joy, sadness). This data set contains a total of 200,000 48 * 48 grayscale images with a total of 6 expressions.
For example: 0 = anger, 1 = disgust, 2 = fear, 3 = happy, 4 = sad, 5 = surprised, and 6 = normal.

Limitation

1. The total size of the two pictures of face matching input is less than 8MB.

2. The picture size is less than 8MB. Due to the large picture, the delay will be longer and the amount of picture information is not large. It is recommended to be less than 1MB.

3. The resolution of the picture is less than 4096 * 2160, and the face pixels in the picture are larger than 80 * 80. It is recommended to be more than 120 * 120.

In order to ensure the recognition effect, the following requirements are recommended for face pictures:

- Illumination is greater than 200lux, no reflection, strong light and shadow.
- The face is unobstructed, the whole is clear, and there is no blurring of motion.
- The side face does not exceed 30 °, the elevation angle is less than 15 °, and the deflection angle is less than 15 °. The human face in the picture remains upright.

Constraints

Constrained by many factors such as technology and cost, there are some constraints on facial expression recognition services. System-level constraint restrictions:

- Only supports JPG, PNG, JPEG, BMP format pictures.
- Each user can use 10 facial databases for free, and each facial database has a capacity of 100,000 facial features.
- The system does not save user pictures or videos.

METHODOLOGY

We are going to use SVM and Random Forest Classifier algorithms for classifying the emotions of the person into one of the binary classes (Anger / Fear). Our model converts the video into vectors of video level feature extracts and to introduce knowledge transfer a largescale emotion-centric auxiliary image set is being used to classify the emotions. This entire process in been classified into five different stages, they are

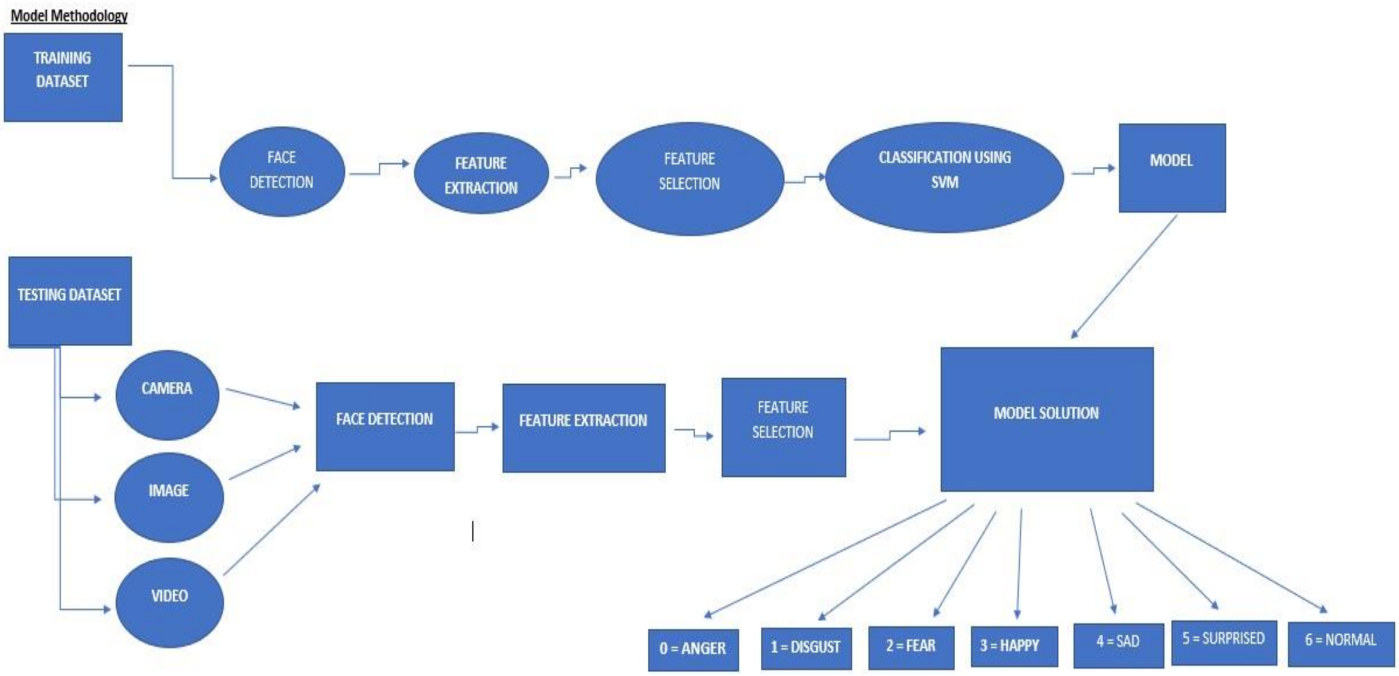
Face frontalization: We synthesize the front view of the face, once the face has been detected the system will track the face as best as it can to recognize the emotion in each frame of the video during this process the resolution is reduced to less than 4096 * 2160.

Preprocessing: Several preprocessing techniques will be applied like contrast adjustment, intensity adjustment to remove the noise and improve the image and eliminate the suppresses unwanted distortions or enhances.

Feature Extraction: Unprocessed faces has lots of amount of data in it and feature extraction is required to decrease it to smaller sets of data called features. This dimensionality reduction process efficiently represents the interesting emotions of the face in the form of a vectors. This approach is used to reduce the feature representation and complete the tasks of matching the emotions quicker.

Feature Selection: This process is concerned with choosing the subset of features of the face which are required for our model to classify the expression into different categories. Having to many features may raise the complexity of classifying and having improper features may result in improper classification.

Classification: During this process the facial expression of a person recognized is classified into binary classification as fear or anger. Accordingly, the nearby guards can be alerted.



TESTING PLAN

Introduction

The final model called Nonlutte, it's a solution that takes live video images as captured using security cameras as input to detect facial emotions using image detection technology to extract features which is then classified into one of the following; anger, disgust, fear, happy, sad, surprised and normal.

Objectives

This document describes the plan for testing the Nonlutte Emotion detection solution with the following objectives.

- Identify the mode architecture and data
- Identify the test requirements
- Describe the testing strategy
- Describe the test output
- List the deliverable elements of the test activity

Test Scope

The purpose here is to test the feasibility and acceptance of our proposed solution, as it is critical that all systems work perfectly alongside the Nonlutte solution. The test would measure the performance and interface with the following subsystems

1. Response time for registration and logon
2. Existing camera functionality and integration with the Nonlutte solution
3. Ability to classify live images correctly
4. Response time to notify appropriate security personnel

Requirements for Test

Data Integrity Testing

- Verify access to cloud storage to store image data for continuous training of the model
- Verify accuracy and consistency of live data

Function Testing

- Verify integration with existing security camera
- Verify integration with security alert mechanism

User interface testing

- The desktop user interface shall be windows 10 complaint
- The systems shall be easy to use and the target market are event center's with large crowd of people

Performance Testing

- Verify response time between detecting required expression and when security alert is activated
- Verify response time for remote login

Load Testing

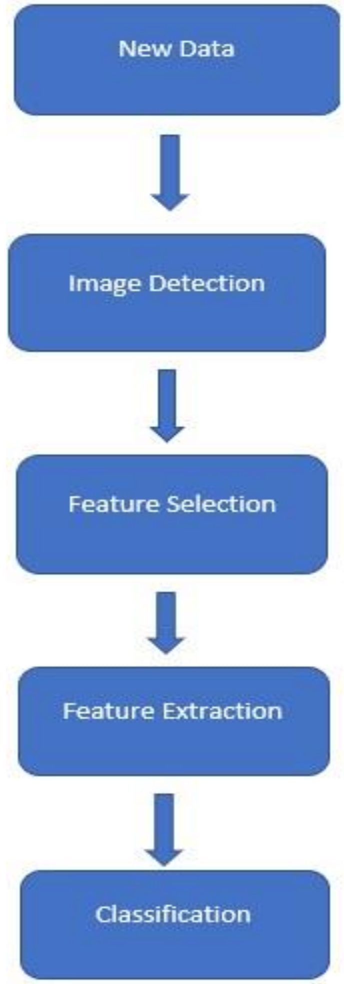
- Verify system response time when processing large data input

Test Strategy

The goal here is to produce a continuous learning solution of the highest possible accuracy level and in order to ensure that the model works as intended, the following strategy is adopted for testing.

- A subset of the dataset is set aside to test all possible combination of images to estimate how well the model performs
- The test data is fed into the model
- The model then detects the faces and assign a bounding box to each image
- Feature selection is performed and feature extraction pulls the features from the image required to classify emotions
- The classifier then produces the required output of a combination of Anger and Fear

- Each classified combination of Anger and Fear sends an alert to the closest security personnel indicating the location of the classified outcome.



Test Outcome Milestone

- Verify that the system alerts security personnel of the location of a possible fight based on the combination of the classified emotion.

GITHUB LINK

https://github.com/chrisvarella/visualai

MODELLING

Model Structure

In order to solve the problem statement, the Haar Feature-based Cascade Classifier for Object Detection model will be used.

The model will be trained with multiple faces with expressions including anger and fear – those are called positive examples. After the classifier finishes training, it's ready to start classifying images and it does so by searching for angry faces by a sliding window. The process involves multiple stages that are applied to a specific area until rejection or all stages pass. The classifiers are decision-tree classifiers with at least two leaves.

Software Tools

This project relies on the python language and its computer vision libraries, plus images of faces with expressions. Our IDE will be Anaconda's Spyder and Jupyter Notebooks.

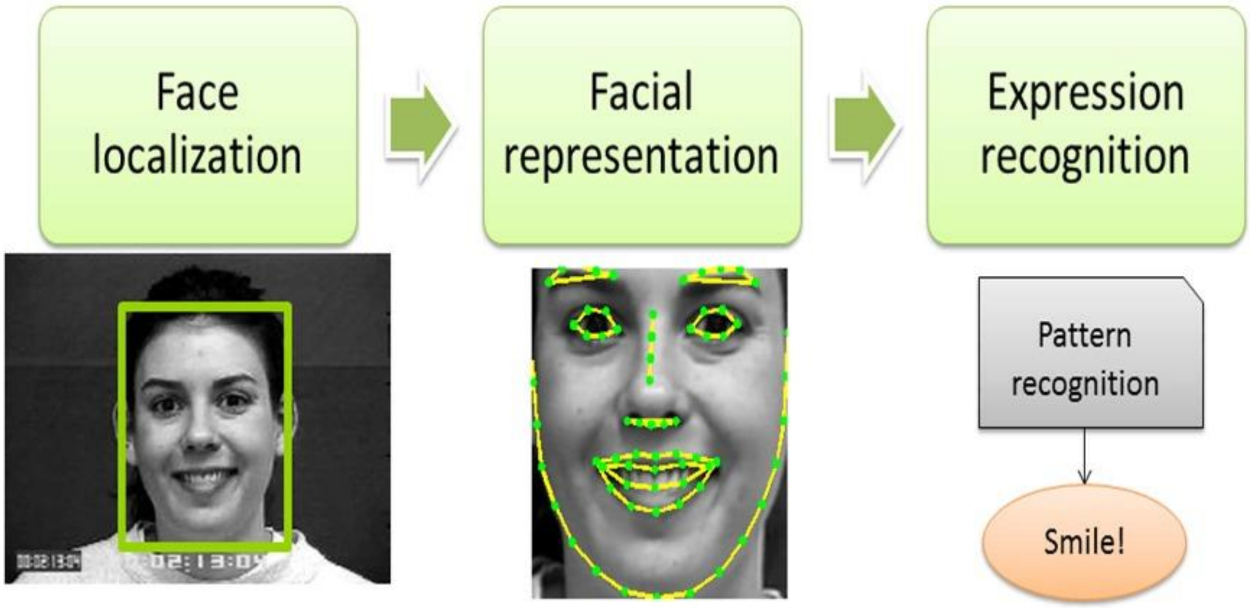
Model Calibration

For evaluation process and calibrating the performance of the model of emotion recognition system we have five different calibration techniques. They are, standard calibration using linear transformation of scores and second order models (9 and 25 points), 2D mapping with interpolation, a mapping with developed model describing eye image formation process and polynomial calibration method.

Exact calibration in real time is required for emotion recognition system from the surveillance video camera and the cost of likelihood ratio is used as a calibration performance metric. According to the research conducted, results indicate that the scores of linearly calibrated face recognition are less misleading in likelihood ratio interpretation. These calibration techniques are also used to improve the verification performance of a face recognition system

Model Architecture and Software Pipeline

Our model was to classify different emotions of a person and based on the emotion we will try to predict the situation and alert the staff if required. In this process we will be using the video data from the surveillance camera and this data is preprocessed. When the face is identified then the feature extraction is performed and an expression is recognized. Based on the previous patterns the model classifies the expression as smile or anger. This is the pipeline used for the flow of data in our model.



MODEL CREATION AND EVALUATION CRITERIA

Model selection involves selecting the right statistical model, given the data, that simply addresses the subject matter problem. This is often the most critical part of any analysis, but given that we know what our data assumptions, limitations and constraints are, we can then assess as our criteria for model selection and evaluation.

Traditionally, in solving this problem using a classification model would involve image detection from an input, feature extraction to extract facial components and landmarks from face region, then passed through a classifier such as Naïve Bayes, SVM and Random forest algorithms to produce classified recognition results using the extracted facial features. This focuses on specific label that are predefined, ignoring the underlying imbalance that it could learn from.

In order to get a higher accuracy, we choose to apply deep learning approach because this technique enables an end to end learning to occur in the model pipeline directly from the input images. Using the CNN deep learning model to pass the image through a filter collection in the convolution layers to produce a feature map which is combined with fully connected networks to produce facial recognition belonging to each class.
To reduce computational time in training the model, we decided to use transfer learning method, by using a pre-trained models that have been used for another facial recognition task to jumpstart the development process for this analysis. Combining Transfer learning with CNN deep learning approach would help address the following;

Assumption – The pre-trained model has saved model weights that have been trained with larger datasets classifying more than 6 different expressions. So this would work well in classifying six expressions for our dataset.

Limitation – Using transfer learning with CNN would tackle the size limitation with size consistency using the Haar cascade function from OpenCV to face detection and cropping of the image size.

Constraints – Using transfer learning helps minimize the constraint of technological cost and computational time in training the model.

CNN MODEL ARCHITECTURE

- Convolutional Layer – In convolutional neural networks, this layer is the feature defining layer. For every input of 2 dimensional or 3 dimensional image fed into this layer, convolution gives the image pixel features which are only a function of the pixel to form the feature map. This layer also output a smaller image after carrying out dimensionality reduction to extract the key features from the image.
- Applying the activation function
- Pooling layer
- Densely connected layer

MODEL VALIDATION AND EVALUATION

For validating our model, we would split our dataset into the following three subsets.

Training set: Used to train the model

Validation set: Used to estimate and arrive at the best model, as well as tune the model hyper parameters.

Test Set: Used for final testing of the approved model.

Under Fitting: This signifies the model is too simple and to address this we can increase the model complexity by adding more feature, layers to our network to try a different activation function.

Over Fitting: This signifies our model is too complex and we increase regularization parameters or reduce the layers to simplify the features extracted.

MODEL EVALUATION

Here we evaluate how well our model will generalize when exposed to out of sample or new data. Looking at our problem statement, it is important for our model to accurately predict emotion in real time to avoid fights at events. But accuracy would not be the right metric to evaluate our model on as accuracy could also include a lot of false positives which could cause unnecessary pressure on the solution and security personnel. Rather, we would use precision, recall and F1 to balance our evaluation process of true positives and false positives. True positives are data points classified correctly as positive while False positives are data points the model classifies incorrectly as positives but are actually negatives.

Precision – This is the ability of our model to identify only the relevant datapoints to classify an emotion into the right class. It is defined as the number of true positives divided by the number of true positives plus the number of false positives. False positives are datapoints that were wrongly classified as positive which are actually negative.

Recall – The ability of the model to find all relevant datapoints within a dataset. It's the number of true positives divided by the number of true positives plus the number of false negatives.

F1 score – F1 score measures the harmonized mean between precision and recall taking both metrics into account.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

MODEL SELECTION SCORECARD

OBEJECTIVE	MEASURE	TARGET	VALIDATION METRIC
	Financials		
Early detection of fights using emotion detection	Number of fights avoided / number of successful events / No loss experienced from fight disruptions	April 16, 2020	F1 score – Harmonized mean of Precision and recall
	Customer		
Ensure a peaceful hosting of events / Increase customer confidence at events	Accuracy in deploying security personnel to areas where suspected emotions have been detected	April 16, 2020	Precision of system deployment of security to resolve correctly detected emotion
	Internal		

Reduce Loss from disrupted events due to fights	Reduction in Losses from events with fights	April 16, 2020	40% decrease in fights at events / Safe events – increased patronage
Learning and Growth			
Innovate through further learning using data	Improved personalized services using correctly detected emotion	April 16, 2020	Measured against number of correctly predicted emotions