Comprehensive Design and Analysis Project
Design Document

IT16129740 Y.C. Tittagalla

Fraud Detection Based on Human Body Language Patterns

Bachelor of Science (Honors) in Information Technology
Department of Software Engineering
Sri Lanka Institute of Information Technology
Sri Lanka
May 2019

Table of Contents and Index

1 Introduction.........4

1.1 Purpose.........4

1.2 Scope.........4

1.3 Definitions, Acronyms, and Abbreviations.........4

1.5 Overview.........5

2 Overall Descriptions.........5

2.1 Product perspective.........6

2.1.1 System interfaces.........8

2.1.2 User interfaces.........8

2.1.3 Hardware interfaces.........8

2.1.4 Software interfaces.........8

2.1.5 Communication interfaces.........9

2.1.6 Memory constraints.........9

2.1.7 Operations.........9

2.1.8 Site adaptation requirements.........9

2.2 Product functions.........10

2.3 User characteristics.........10

2.4 Constraints.........10

2.5 Assumptions and dependencies.........11

2.6 Apportioning of requirements.........11

3 Specific requirements.........11

3.1 External interface requirements.........11

3.1.1 User interfaces.........11

3.1.2 Hardware interfaces 12
3.1.3 Software interfaces 13
3.1.4 Communication interfaces 13
3.2 Architectural Design 14
3.2.1 High level Architectural Design 14
3.2.2 Hardware and software requirements with justification 14
3.2.3 Risk Mitigation Plan with alternative solution identification 15
3.2.4 Cost Benefit Analysis for the proposed solution 15
3.3 Performance requirements 15
Processor 16
GPU 16
RAM 16
Memory 16
3.4 Design constraints 17
3.5 Software system attributes 17
3.5.1 Reliability 17
3.5.2 Availability 17
3.5.3 Security 17
3.5.4 Maintainability 17
4 References 17

1 Introduction

1.1 Purpose

The purpose of this DD document is to provide a detailed description about how we can identify frauds using human behavior patterns and body language and how we have incorporated it with machine learning to identify the frauds in real time. This document include system interfaces, hardware interfaces and site adaptation requirements.

Also this DD will indicate the functional and non-functional requirements and it will compare the system that is currently being built with the systems that are already in the market. And also this document will describe about the intended goal to be achieved and the benefits and objectives of this study which will describe the expectations and purposes for doing this research.

The intended audience of this DD document is mainly researchers and developers who would like to know more about how this component of the system is being built how it'll contribute to the whole system. This will also give an overview, high level architecture of this component.

1.2 Scope

This document covers the requirements for the one of the main components of the whole fraud detection system. The whole system is consisted of four parts and this document will cover fraud detection using human gait and body language.

This part of the system will only identify fraudsters using body language and the system will only give a suggestion to the user(s) of the system that this person is more abnormal than other persons in the vicinity. It will not brand any person as fraudster.

The data are fed in to this component via CCTV footage and it'll be analyzed in real time to give the user(s) of the system a feedback of a fraudulent activity. This system will also actively track the person between multiple cameras.

1.3 Definitions, Acronyms, and Abbreviations

CCTV	Closed-Circuit Television Camera
BLOB	Binary Large Object
ML	Machine Learning
DD	Design Document
3D	3 Dimensional
PCA	Principal Component Analysis
DVR	Digital Video Recorder

FOV	Field Of View
GPU	Graphics Processing Unit
RAM	Random Access Memory
PCI	Peripheral Component Interconnect
SSD	Solid State Drive
CPU	Central Processing Unit
LAN	Local Area Network
WAN	Wide Area Network

1.5 Overview

The main goal of this component in the system is identifying fraudsters using their body language and in this DD we will describe how we will achieve this by using Machine learning (ML) and what the plans are for the future of this system.

The 1st section of the DD contains the purpose which will describe about the purpose of preparing the DD. The scope of the system which will describe the benefits, goals and objectives of implementing the application and the overview which explains how DD is organized and what the rest of the DD contains.

2nd section the overall description will include the product perspective that will compare the application with the other competitive products. System, user, hardware, software and communication interfaces. Summary of the main functions of the application. The user characteristics that describe what kind of people will use the system. Constraints that describe all the conditions that may limit developers' options. And also this section will include the assumptions and dependencies done during the development.

The 3rd section contains the specific requirements describes the design characteristics and external behavior of the system in technical terms and in a more detailed manner than in the section two. Which will be referred by the developers. This section includes all user interfaces in detail and hardware, software interfaces and explains the functional and non-functional requirements of the system.

2 Overall Descriptions

Recognition of human body language may be tricky since each different person has their own unique body language but when a person is subjected to fear and anxiety kind of emotions we all have natural instinctive way of acting and we use these social cues to differentiate the fraudsters.

We cannot identify a fraudster by only examining one's body language so in the whole system we have other components which will help us give the final prediction. After the systems has been installed within a certain company it'll also re-evaluate the results based on the inputs given by the user(s) about the fraudsters that the systems identifies so that the system will continue to learn even after deployment and will get more familiar with the environment that it has been deployed in.

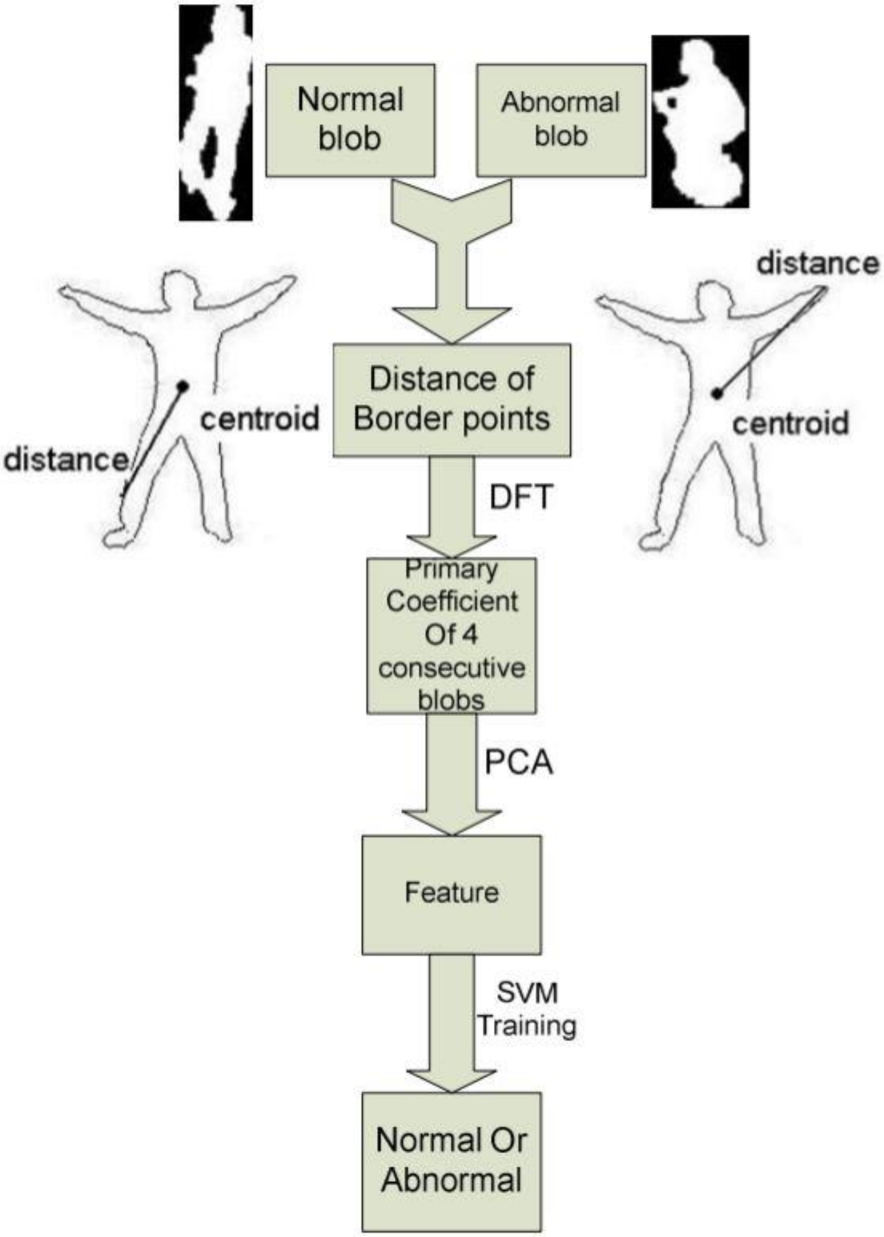
This kind of software is beneficial in many scenarios in the world even though this was mainly targeted towards fraud detection the body language detection component can be used in various places since it detects anomalies that are happening. This kind of system can also be applied to the current situation that the Sri Lanka is going through we can scan for threats in the hotels, schools and offices since even though no matter how hard a person tries to hide his true intentions it'll be portrayed by his body language.

Introducing this into new environments are also easy since most of the places already have CCTV cameras setup and we only have to introduce this software component.

2.1 Product perspective

Existing working commercial systems for this instance cannot be found at the moment most of the work done in this area are research-oriented projects and since this project is also research oriented we will compare the system to be built with other research projects that have been done and what are the pros and cons of this system compared to other researches that have been done.

In most of the research that have been carried out in the video based anomaly detection only focus on one or two human behavior patterns to distinguish normal behaviors from abnormal behaviors in [1] they used a BLOB to identify humans but in it they also differentiate the foreground and the background of the footage they receive but the problem was the images it was hard to identify every single anomaly. The researchers used human part segmentation and PCA(Principle Component Analysis) to choose the most suitable feature in the scene. This modeling approach for unusual events' detection is good only if these abnormal events are well defined and there are enough training data.



Identification Of Body Language

In [2] it reviews a list of most popular methods that are currently used in human anomaly detection. Most of the research that are reviewed does not take account the 3D space around human so it cannot identify those movements but in this research we will be using dense-pose [3] to create surface-based representation of the human body to represent 3D space.

2.1.1 System interfaces

Access for the companies CCTV footage as an input to the component.

2.1.2 User interfaces

There are no special UI that the user(s) can interact with in this component only the output will be given from this component to the main system. We can use the interface that is designed for fraud detection using transaction patterns if the user(s) only wishes to by a single component.

There can be some intermediate visual outputs given to the user just for the purpose of demonstrating how the system works like dense map of the human body recognition.

Dense pose human body estimation

2.1.3 Hardware interfaces

CCTV Cameras – To take the input of the surveillance video into the component.

DVR – To get the input from multiple CCTV cameras.

2.1.4 Software interfaces

The only communication this component will have is the communication between the components that resides in the system a direct communication with other three components of the systems to identify frauds which are,

(1) Voice based fraud detection component.
(2) Transaction based fraud detection component.
(3) Micro-facial expression based fraud detection component.

All these components will come together to give the final verdict of the person that has been analyzed by the system.

2.1.5 Communication interfaces

Web Browser will provide access to the dashboard when needed.

Access to the local server

2.1.6 Memory constraints

For internal memory system will not need that much of RAM capacity since most of the heavy work will be done by the GPU 8GB of RAM would be more than enough to run the system.

For External memory if we need to crosscheck possible fraudsters that were identified by the system earlier but no action were taken we can use the old CCTV footage that is already stored in the servers of the company normally a recordings will last at least one week before it's looped trough and written over again but if the requesting company doesn't need any past video crosschecking we can only use the real-time data to make our prediction. But if the user wants separate server just for storing old video data then it'll take some extra memory this can vary according to the number of cameras that has been used and the quality of the footage we get from the cameras.

2.1.7 Operations

All the operations in this component are unattended operations which means that after installing this component it'll run automatically without any user interactions and this operation will run until the system is shutdown or until CCTV footage are fed into the component.

If the footage we get is too noisy we can use denoising auto encoders to clear up the footage this can only achieve clarity for some extent only.

There are no backup operations in this component but most of the CCTV footages are backup by the companies for various security reasons so we can even analyze the footage that is not in real-time.

2.1.8 Site adaptation requirements

All the sites that require this component should have digital CCTV cameras installed in the site and system should be able to access these footage.

Need to have machine that is able to run the system.

2.2 Product functions

One of the major functions of this component is analyzing CCTV footage in real time. This part of the system is and internal component users does not need to interact with this component other than to feed data into it data also fed automatically after a secure link is established between the system and the CCTV camera network.

In this component users will also be tracked between cameras so if we start analyzing a user and he went out of fov (Field of vision) of the camera that he is being currently tracked and appeared in another cameras fov then the system will recognize that this is the same user as before and continue with its scanning.

Each person that has been analyze by this component will have a threat score and it will determine how much should this user be analyze by other components of the system this will help us reduce the computational power of the overall system.

2.3 User characteristics

Users of this system should not have any prior knowledge regarding this system or the underlying architecture of the system.

User should know that the results produced by this system is just a suggestion and it will not always be 100% accurate.

User(s) should have the authority to act upon the information that is outputted by the system or should know how to contact persons who know how to handle the situation if such occurs.

2.4 Constraints

The system should be modular so the system should work without having all the components, user(s) can buy the components that they think are valuable for their environment and system should give the output based on those components.

The footage that we receive form the CCTV cameras should be digital and non mono-colored.

The server that the software is running on should have at least gtx 1060ti 6GB GPU and processor that should not bottleneck this graphics card.

The accuracy will be limited to the video the component get fed.

2.5 Assumptions and dependencies

Assume that one server can handle all the customers in a peak hour.

Assume that fraudsters are not highly trained to hide their body language and emotions.

Assume we have clear analyzable footage from CCTV

2.6 Apportioning of requirements

The requirements mentioned in the 1st and 2nd section of the document are primary specifications of this component. In the 3rd section the functional and non-functional requirements are mentioned in detail. If any major defects are found according to the requirements the testing will be done and defects will be corrected. Application will be implemented by the developers in horizontal manner and no function will be completed at the middle of the development.

Essential requirements of this component are,

- Mapping humans into 3D space.
- Identifying anomalies in human behavior.
- Tracking person.

Desirable requirements of this component are,

- Classification of persons according to the level of suspiciousness.
- Learning from previous findings.
- Getting a heat signature map of the person.

There are no optional requirements for this component at the time this DD was created.

3 Specific requirements

3.1 External interface requirements

3.1.1 User interfaces

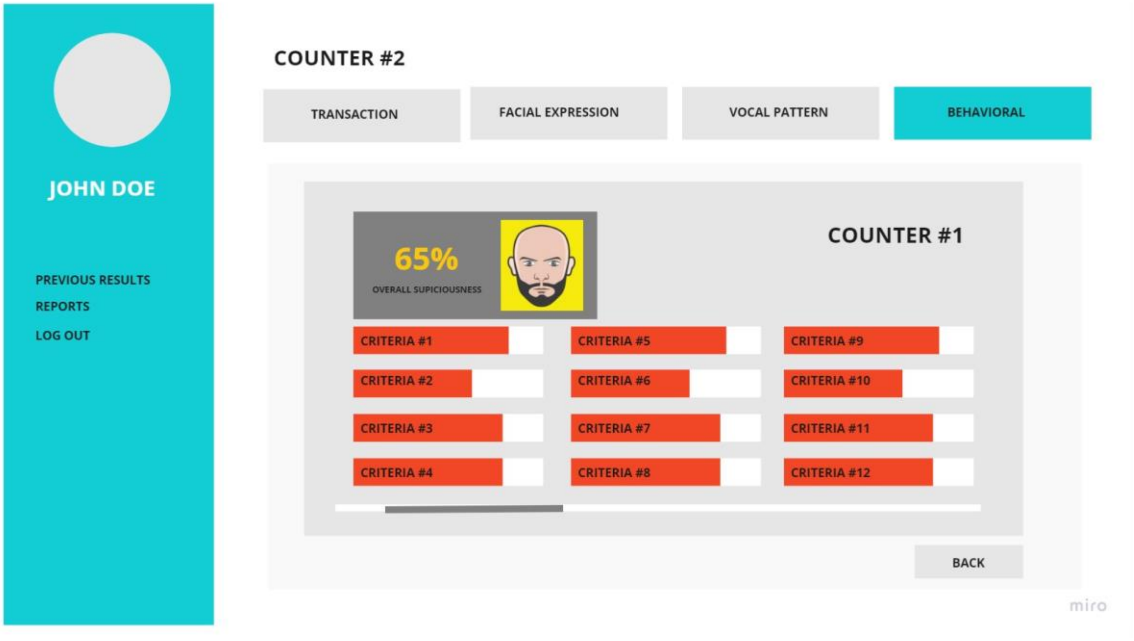
As mentioned in the section 2.1.2 this component does not contain any interfaces that the user can interact with but if the user chooses to only use this component for their entire flow then we can use the lobby page that will be developed to give the user a brief idea about what have been found so far and what are they.

Sources of input for this component is mainly the footage that we receive from the CCTV cameras and the output of this component will be complete threat analysis score of the person that is being analyzed by the component which is mainly based on human body language.

Timing of this component will began when the person enters the site and while and he is on the site component will do regular analysis and if the person leaves site then

The accuracy of this component vary from the quality of the footage , number of objects that are present in the scene this component have low level of accuracy but also have a very low level of computational power usage compared to the other components that are contained in the system.

When the customers enter the site they will be scanned for abnormal body language by clicking on the body language tab user(s) of the system can see for what abnormal behaviors they were detected for and the user can decide whether it is disability or an actual fraudster.
In the overview page user(s) can get an idea about the whole situation of the site and take necessary precautions accordingly.



3.1.2 Hardware interfaces

The hardware aspects that this component deals with are CCTV cameras and the local servers which the components resides in.

The server will need at least 16GB of RAM and nVIDIA GTX 1080Ti for analyzing videos in real-time. The processor should not bottleneck these other components as for the memory since there is no need of storing the data unless the client asks for it storage space can vary.

3.1.3 Software interfaces

This component will interact with Human voice analysis component which will classify the fraudster according to their voice patterns and micro facial expressions which will analyze the persons face to figure out whether the persons actions are unnatural or not and finally transaction patterns component which will analyze the persons transaction details to find out suspicious activities.

TensorFlow framework will be used for training the machine learning module. After the training phase it will get the data stream from bank data and predict the probability of frauds.

For the admin personnel of the bank, we will be implementing a web application using Angular which will have a dashboard for the who system. All the modules of the system could be accessed by the Angular based web application.

Since both the frameworks (TensorFlow and Angular) are node package-based java script frameworks, it will be easier to communicate and handle within the software applications.

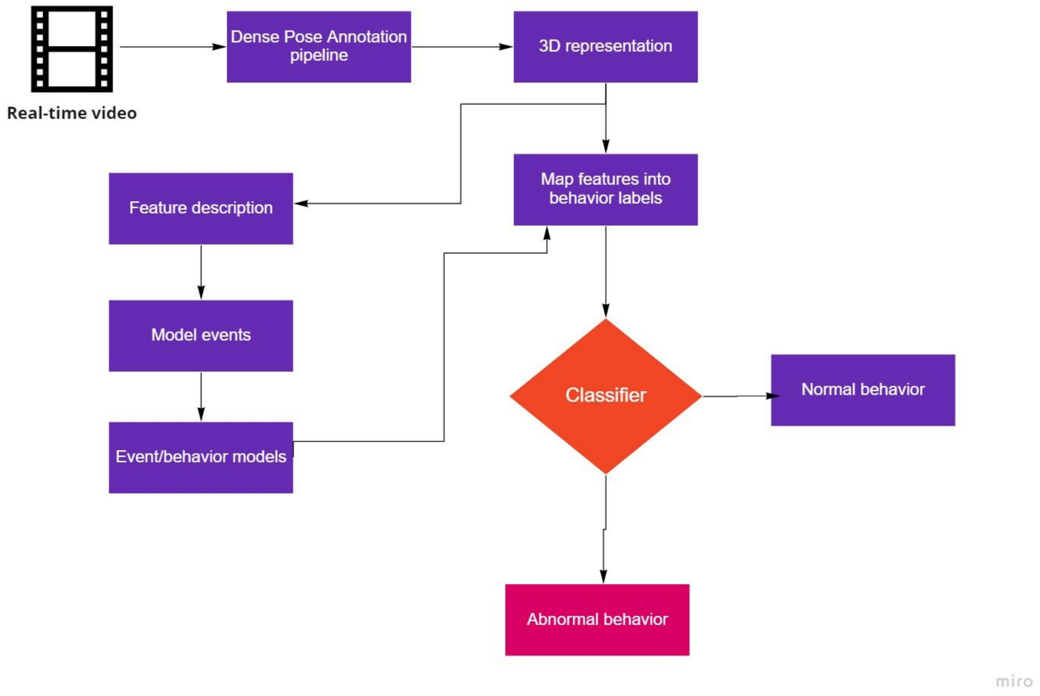
3.1.4 Communication interfaces

The whole system will reside inside intranet of a company that is a Wide Area Network (WAN) because we need to communicate between the other components which are also connected to the local network.

When a fraudster is detected the application will raise this to the user and via the local network it will inform the relevant parties to take action.

3.2 Architectural Design

3.2.1 High level Architectural Design



High level architecture diagram

3.2.2 Hardware and software requirements with justification

Hardware Requirements

CCTV cameras -Main hardware requirement for this component is CCTV cameras since without the data we cannot make use of this component, only with the data that has been provided through the cameras we can run the algorithm that will identify fraudsters.

nVIDIA gtx 1060ti 6GB graphics card – Since we are doing real time processing of the data that are fed into the component we will at least need a graphics card with much of VRAM and processing power.

16GB of RAM – since the data processed should be analyzed and while they are being analyzed by the other components the intermediate result should be stored in the RAM for the ease of access so we will need a higher RAM capacity.

3.2.3 Risk Mitigation Plan with alternative solution identification

There are several risks involved with component one of them being the failure of the system due to the footage being unclear this can be mitigated by using variation encoders to clear up the images that we are gathering.

We can also reduce the single point of failure risk by not only giving the output based on this component but also combining it with all the other components in the system we can reduce the risk of single of point of failure.

By analyzing past projects and researches we can identify what went wrong and what are the best/optimal algorithms to use for each of the components.

If the tracking of the person turns out be not as optimal for the solution we proposed we can create a temporary profiles for those who enters the site and can update their risk factor in those profiles of the person.

3.2.4 Cost Benefit Analysis for the proposed solution

There are lot of cost benefits of using an automated system to detect the frauds,
Since this system will learn from its' past mistakes this after some time this will very adaptive to the environment that this is deployed in and since this is an automated process it'll run 24/7 regardless of any interference.

The human error will be minimum and not susceptible to bribes there will be somewhat initial cost if the site does not have an already working CCTV camera system but it'll be more cost efficient that hiring agents to watch over the footage.

Can connect between multiple sites if there is a WAN connection between sites so we can communicate between them to identify large scale frauds.

3.3 Performance requirements

This component will analyze the footage in real time and give the user(s) of the system alert indicating that a fraud is happening then and there. The main key factory that will decide the overall system efficiency is the GPU.

Processor

AMD’s 1920X has 12 cores and 38MB cache and is more expensive vs. 1900X’s 8 cores and 20 MB cache. Earmarking 2 cores / 4 threads per GPU and the fact I might want the machine to double as a staging server later, 1920X gives me a little more breathing room.[4]

GPU

It’s hard to know how many GPUs you’ll need because some models take 10s of hours to train (Vision CNNs, Natural Language Processing LSTMs, etc). So, one of the best ideas is to start with 1 or 2 GPUs and add more GPUs as you go along.

Each GPU requires at least 8x PCIe lanes (it’s 16x officially, but there’s data for which 8x is good enough if you’re not running cross-GPU experiments). If we are not running SLI will need 4x PCIe lanes for the M.2 SSD (which plugs right in and is 5x faster than SATA3), and another 4x PCIe lanes for Gigabit ethernet. That’s a total of 40 PCIe lanes and will restrict your CPU choices quite a bit. Your CPU will dictate the motherboard you need. (For example: AMD Threadripper CPU = X399 chipset motherboard, Intel 7900X CPU = X299 chipset motherboard, etc).

This will only needed for the training of the model but when we deploy it in sites we can manage the system with much lesser GPU. [4]

RAM

When working with large/big datasets we might need to have them in memory. Size of the RAM decide how much of dataset you can hold in memory. For Deep learning applications it is suggested to have a minimum of 16GB memory Regarding the Clock, The higher the better. It ideally signifies the Speed—Access Time but a minimum of 2400 MHz is advised.

Always try to get more memory in a single stick as it will allow for further expansion in remaining slots.I have seen many people who get 4*8 GB RAM instead of 2*16 GB ending up using all 4 Slots and no room for upgrade just because they are bit cheap than the latter.[4]

Memory

Its always better to get a small size SSD and a large HDD. SSD’s are preffered to store and retrieve data that is actively used. On the other hand HDD should be used to store data that are to be used in future.

[4]

3.4 Design constraints

The main screen when the user(s) of the systems logs into will be navigated to the lobby area which will display the details of the components and the will be navigation to view the whole crowd that are in the site like a normal CCTV footage. Since these are not that essential interfaces designers are free to design any suitable design.

3.5 Software system attributes

3.5.1 Reliability

This module should be predicting the fraud possibilities at least at a rate of 60%. If it becomes lower than that, the commitment of this module to the system will be very low. Because from all the four modules, most reliable modules will be giving a higher percentage to the decision of fraud detection. For a competitive product, this should be in a higher level of reliability. This could be measured by confusion matrices and other techniques, by predicting and later resolving whether the predictions were correct or wrong.

3.5.2 Availability

This system's availability mostly relies on the LAN performance of the bank premises. If the LAN is available, all the information streams are open and the system will function as usual. The availability will also depend on the CCTV camera networks availability.

3.5.3 Security

Since cameras are not installed around sensitive information there is no issue even if someone did manage to get the CCTV footage but this camera network is also secured. However there is security concern of a normal physical attacks against the cameras which will render them useless. Other than that, LAN's can not be penetrated through outside parties, giving the application whole lot of security

3.5.4 Maintainability

The machine learning model is the only components that changes frequently and automatically. So, there is nothing programmers can do. The model will self-learn via the human patterns that are analyzed and the data coming in from the other components.

4 References

[1] O. ,. H. Q. Y. X. Xinyu WU, "A Detection System for Human Abnormal Behavior," 2005.

[2] M. I. a. K. W. Oluwatoyin P. Popoola, "Video-Based Abnormal Human Behavior," 2011.

[3] R. A. N. N. Guler, "DensePose: Dense Human Pose Estimation In The Wild," 2018.

[4] G. T. S, "https://medium.com/mlreview/choosing-components-for-personal-deep-learning-machine-56bae813e34a," 2017. [Online].