Engineering Universe for Scientific Research and Management
ISSN (Online): 2319-3069
Vol. XVII Issue I
January 2025

Data extraction through face detection and logical analysis

Aman Prakash[1], Ayush Kumar[2]
M.Tech Scholar, Department of Computer Science and Engineering, REC, Bhopal (MP)[1]
Assistant Professor, Department of Computer Science and Engineering, REC, Bhopal (MP)[2]

Abstract

The growth of massive data stores has led to the development of a number of automated processors that work to discover relationships in and between the data in those stores. These processors are often referred to by a number of names including data mining, knowledge discovery, pattern recognition, artificial and machine learning. Data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It is used to automatically extract structured knowledge from large datasets. The application of logic with data mining makes information understandable to human. Data mining can have many methods like association rules, classification, clustering. One of the methods of implementing association rules is Apriori algorithm. In this thesis, a Apriori System is built; it uses Apriori algorithm alone, then Apriori algorithm with the application of logic to find association rules. It will find the relationships among items stored in a supermarket to present knowledge about what are the most soled items and the relations among items. From the experimental results, it was found that functions filters the results and make number of extracted rules less than the number of rules extracted by applying Apriori algorithm only.

Keywords: ***Face detection, recognition, system, OpenCV, Eigenface.***

1. Introduction

Progress in digital data acquisition and storage technology has resulted in the growth of huge databases. This has occurred in all areas of human endeavor, from the mundane (such as supermarket transaction data, credit card usage records, telephone call details, and government statistics) to the more exotic (such as images of astronomical bodies, molecular databases, and medical records). After that interest was grown toward tapping these data, and extracting from them information that might be of value to the owner of the database. The discipline concerned with this task has become known as data mining. Data mining is the analysis of large observational data sets to find unsuspected relationships, and to summarize the data in novel ways that are both understandable and useful to the data owner [Dav01].

The amount of data stored in databases continues to grow fast. Intuitively, this large amount of stored data contains valuable hidden knowledge, which could be used to improve the decision-making process of an organization. For instance, data about previous sales might contain interesting relationships between products and customers. The discovery of such relationships can be very useful to increase the sales of a company. However, the number of human data analysts grows at a much smaller rate than the amount of stored data. Thus, there is a clear need for semi automatic methods for extracting knowledge from data. This need has led to the emergence of a field called data mining and knowledge discovery. This is an interdisciplinary field, using methods of several research areas (specially machine learning and statistics) to extract high level knowledge from real-world data sets. Data mining is the core step of a broader process, called knowledge discovery in databases, or knowledge discovery, for short [Ale01].

Data Mining Data mining is the process of extracting meaningful information from large quantities of data. It involves uncovering patterns in the data and is often tied to data warehousing because it makes such large amounts of data usable. Data elements are grouped into distinct categories so that predictions can be made about other pieces of data. For example, a bank may wish to ascertain the characteristics 2 that typify customers who pay back loans. Although this could be done with database queries, the bank would first have to know what customer attributes to query for. Data mining can be used to identify what those attributes are and then make predictions about future customer behavior [Sar05].

2. Data Mining and Logic

Knowledge discovery, whose objective is to obtain useful knowledge from data stored in large repositories, is recognized as a basic necessity in many areas, especially those related to business. Since data represent a certain real-world domain, patterns that hold in data show interesting relations that can be used to improve human understanding of that domain. Data mining is the step in the knowledge discovery process

that attempts to discover novel and meaningful patterns in data. The theory of sets can certainly help data mining to reach this goal. It is widely recognized that many real world relations are intrinsically . For instance, clustering generally provides a more suitable partition of a set of objects than crisp clustering do. Moreover, sets are an optimal tool to model imprecise terms and relations as commonly employed by humans in communication and understanding. As a consequence, the theory of sets is an excellent basis to provide knowledge expressed in a meaningful way [Mig03].
Polyester Resins

2.1 Data Mining and Fuzzy Logic

Data mining consists of finding interesting trends or patterns in large datasets, in order to guide decisions about future activities. There is a general expectation that data mining tools should be able to identify these patterns in the data with minimal user input. The patterns identified by such tools can give a data analyst useful and unexpected insights that can be more :carefully investigated subsequently, perhaps using other decision support tools [Jef06]. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration [Jia06]
Definition of Data Mining
Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques [Dan05].
Data Mining Usage
Data mining is used for a variety of purposes in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. For example, the insurance and banking industries use data mining applications to detect fraud and assist in risk assessment (e.g., credit scoring). Using customer data collected over several years, companies can develop models that predict whether a customer is a good credit risk, or whether an accident claim may be fraudulent and should be investigated more closely. The medical community sometimes uses data mining to help predict the effectiveness of a procedure or medicine

Limitations of Data Mining

While data mining products can be very powerful tools, they are not self-sufficient applications. To be successful, data mining required skilled technical and analytical specialists who can structure the analysis and interpret the output that is created. Consequently, the limitations of data mining are primarily data or personnel-related, rather than technology-related. Although data mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns. These types of determinations must be made by the user. Similarly, the validity of the patterns discovered is depending on how they compare to “real world” circumstances. For example, to assess the validity of data mining application designed to identify potential terrorist suspects in a large pool individuals, the user may test the model using data that include information about known terrorists. However, while possibly re-affirming a particular profile, it does not necessarily mean that the application will identify a suspect whose behavior significantly deviates from the original model. Another limitation of data mining is that while it can identify connections between behaviors and/or variables, it does not necessary identify a casual relationship. For example, an application may identify that a pattern of behavior, such as the propensity to purchase airline tickets just shortly before the flight is scheduled to depart, is related to characteristics such as income, level of education, and internet use. However that does not necessarily indicate 10 that the ticket purchasing behavior is caused by one or more of these variables. In fact, the individual’s behavior could be affected by some additional variable(s) such as occupation (the need to make trips on short notice), family status (a stick relative needing care), or a hobby (taking advantage of last minute discounts to visit new destinations) [Jef06, Fan06].

2.2 Data Mining Process

A data mining application usually starts with an understanding of the application domain by data analysts (data miners), who then identify suitable data sources and the target data. With the data, data mining can be performed, which is usually carried out in three main steps [Bin07]:

- Pre-processing: The raw data is usually not suitable for mining due to various reasons. It may need to be cleaned in order to remove noises or abnormalities. The data may also be too large and/or involve many irrelevant attributes, which call for data reduction through sampling and attribute selection. Details about data pre-processing can be found in any standard data mining textbook.
- Data mining: The processed data is then fed to a data mining algorithm which will produce patterns or knowledge.
- Post-processing: In many applications, not all discovered patterns are useful. This step identifies those useful ones for applications.

Various evaluation and visualization techniques are used to make the decision. The whole process (also called the data mining process) is almost always iterative. It usually takes many rounds to achieve final satisfactory results, which are then incorporated into real world operational tasks.

Data Mining and Knowledge Discovery in Database

Many people treat data mining as a synonym for another popularly used term, "Knowledge Discovery in Databases", or KDD. Alternatively, others view data mining as simply an essential step in the process of knowledge discovery in databases [Jia00]. Knowledge Discovery in Databases is the process of extracting interesting, nontrivial, implicit, previously unknown and potentially useful information or patterns from data in large databases. Data Mining is the most important step in the KDD 11 process and involves the application of data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the data. The KDD process entails the application of one or more Data Mining techniques to a dataset, in order to extract specific patterns and to evaluate them on the data [And05].

Laps and joints are not required. Corrosion resistance: Unlike metal, FRP does not rust away and it can be used to make long-lasting structures.

Knowledge discovery as a process is depicted in Figure 1 it consists of an iterative sequence of the following steps:

1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
5. Data mining (an essential process where intelligent methods are applied in order to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures)
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)failure pattern. Most often the strengthened beams failed in.

Steps 1 to 4 are different forms of data preprocessing, where the data are prepared for mining. The data mining step may interact with the user or a knowledge base. The interesting patterns are presented to the user and may be stored as new knowledge in the knowledge base. Note that according to this view, data mining is only one step in the entire process, albeit an essential one because it uncovers hidden patterns for evaluation [Jia06, Ode08].

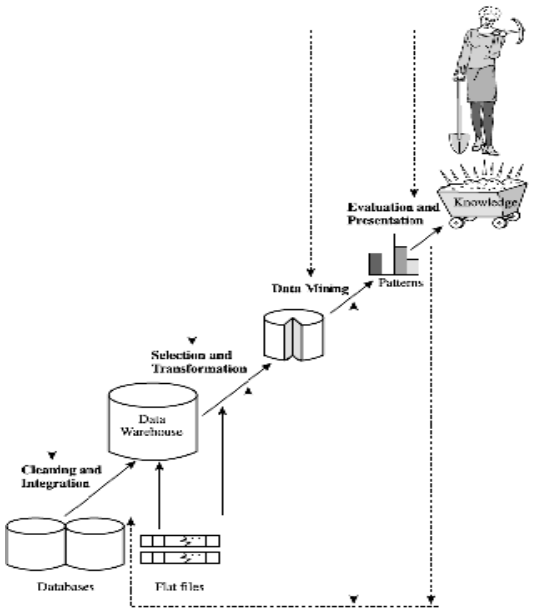

Figure 1 Data mining as a step in the process of knowledge discovery.

Architecture of a typical data mining system:-

The architecture of a typical data mining system may have the following major components (Figure 2.2) [Jia00]:

1. Database, data warehouse, or other information repository. This is one or a set of databases, data warehouses, spread sheets, or other kinds of information repositories. Data cleaning and data integration techniques may be performed on the data.
2. Database or data warehouse server. The database or data warehouse server is responsible for fetching the relevant data, based on the user's data mining request.
3. Knowledge base. This is the domain knowledge that is used to guide the search, or evaluate the interestingness of resulting patterns. Such knowledge can include concept hierarchies, used to organize attributes or attribute values into different levels of abstraction. Knowledge such as user beliefs, which can be used to assess a pattern's interestingness based on its unexpectedness, may also be included. Other examples of domain knowledge are

additional interestingness constraints or thresholds, and metadata (e.g., describing data from multiple heterogeneous sources).

4. Data mining engine. This is essential to the data mining system and ideally consists of a set of functional modules for tasks such as characterization, association analysis, classification, evolution and deviation analysis Many studies and researches have been introduced in the field of data mining, data mining with logic, the following are some of these researches:

a. A Data Mining Algorithm for Quantitative Values, [Tzu99]

This work attempted to propose a new data-mining algorithm to enhance the capability of exploring interesting knowledge from transactions with quantitative values. The proposed algorithm integrated the set concepts and the apriori mining algorithm to find interesting association rules from given transaction data. The rules thus mined exhibit quantitative regularity in databases and can be used to provide some suggestions to appropriate supervisors. The proposed algorithm can also solve conventional transaction-data problems by using degraded membership functions.

b. Mining association rules for classification problems, [Rue02]

This work proposed a learning algorithm, which can be viewed as a knowledge acquisition tool, to effectively discover association rules for classification problems. The consequence part of each rule is one class label. The proposed learning algorithm consists of two phases: one to generate large grids from training samples by partitioning in each attribute, and the other to generate association rules for classification problems by large grids. The proposed learning algorithm is implemented by scanning training samples stored in a database only once and applying a sequence of Boolean operations to generate grids and rules; therefore, it can be easily extended to discover other types of association rules for market basket analysis that can help managers design different store layouts and help retailers to plan which items to put on sale. The simulation results from the iris data demonstrate that the proposed learning algorithm can effectively derive association rules for classification problems.

c. Association Rules: General Model and Applications, [Mig03]

This work developed a general model to discover association rules among items in a (crisp) set of transactions. This general model can be particularized in several ways; each particular instance corresponds to a certain kind of pattern and/or repository of data. They describe some applications of this scheme, paying special attention to the discovery of association rules in relational databases. The proposed model has been tested on some of the applications, specifically to discover association rules in relational databases that contain quantitative data. The model can be employed in mining distinct types of patterns, from ordinary association rules to and approximate functional dependencies and gradual rules. They will be used in multimedia data mining and web mining.

d. Classification Based on Association Rule Mining, [Wei04]

He investigated the way to integrate association rule mining and classification. First, the framework of association rule mining is presented which incorporates set modeling in an association rule mining technique. He studied the impact of different aggregation operators on the rule mining result. The selection of the 4 operator should depend on the application context. Based on the framework of association rule mining, he proposed a heuristic method to construct the classifier based on the set of class association rules. He called this method the FCBA approach, where FCBA stands for Classification Based on Association. The objective is to build a classifier with strong classification ability .In the FCBA approach, the composite criteria of support and confidence is used as the rule weight to indicate the significance of the rule.

e. Extracting Association Rules for Distributed Association Rules, [Raw07]

In this thesis two proposed algorithms were introduced, they focused on the principle of mining knowledge over geographical distributed systems: The main proposed algorithm (extracting Association Rules for Distributed Association Rules (EAR4DAR) algorithm, aims to extract association rules for distributed association rules) instead of extract association rules from huge quantity of distributed data at several sites, and that is through collecting the local association rules from each site and storing them, these local association rules turn in series of operations to produce global association rules over distributed systems.

Secondary Proposed Algorithm: Association Rules_map (AR_map) algorithm aims to get association rules by using AND logic operation which is a suitable tool to represent association relations between items, since it's giving induction for finding relation or not. These new algorithms are saving the cost which is required to communicate over the network, cost of central storage requirements, and rate of required time for execution.

f. Utilize Data Mining to Find the Living Pattern of Customers in Hotels, [Zho07]

This study of finding the living pattern of customers in hotels adopts data mining that combine Apriori algorithm with different min-support and min-confidence and set theory to copy with the criteria, the yielded rules are not only useful to the hotel decision-makers, but also to those who want to do the business.

This study have yielded some association rules from quantitative data sets and taken the rules to the decision-makers of the hotels who are interested in, they have achieved some effects from the association rules. Someone who wants to operate a hotel, also gets some useful information from the rules. This approach can be used in other facets, such as the discovery to the shopping pattern of consumers in the supermarket according to different categories of commodities, etc.

An Evolutionary Data Mining Model for Concept Extraction, [Moh08]

In this work a method is proposed for extracting useful information from a relational database using a hybrid of genetic algorithm and data mining approach to extract user desired information. The genetic algorithm is employed to find a compact set of useful concepts with a good support for the output of data mining process. The output of the common mining system is constant. But sometimes users want more information from database, perhaps information with higher dimensions. So genetic algorithm is used to find information with more attributes. The genetic algorithm has one input which decides the dimension of output of system. The output of genetic algorithm is the number of linguistic values for each attribute in the database. This input is used to get information with lower/higher attributes.

Problem Formulation

The aim of this thesis is to build a system that uses data mining techniques to access dataset by applying specific algorithms for extracting desirable knowledge or interesting patterns from existing datasets for specific purposes, market basket data is used for this work. To improve the data mining work, logic is used to increase the flexibility for supporting supermarket managers in making decisions, it will reduce the number of association rules according to specified threshold for each algorithm.

The goal of this article is to provide an easier human-machine interaction routine when user authentication is needed through face detection and recognition. With the aid of a regular web camera, a machine is able to detect and recognize a person's face; a custom login screen with the ability to filter user access based on the users' facial features will be developed. The objectives of this thesis are to provide a set of detection algorithms that can be later packaged in an easily-portable framework amongst the different processor architectures we see in machines (computers) today. These algorithms must provide at least a 95% successful recognition rate, out of which less than 3% of the detected faces are false positives.

2. A general statement of the face recognition problem (in computer vision) can be formulated as follows: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Facial recognition generally involves two stages: Face Detection where a photo is searched to find a face, then the image is processed to crop and extract the person's face for easier recognition. Face Recognition where that detected and processed face is compared to a database of known faces, to decide who that person is. Since 2002, face detection can be performed fairly easily and reliably with Intel's open source framework called OpenCV. This framework has an in-built Face Detector that works in roughly 90-95% of clear photos of a person looking forward at the camera. However, detecting a person's face when that person is viewed from an angle is usually harder, sometimes requiring 3D Head Pose Estimation. Also, lack of proper brightness of an image can greatly increase the difficulty of detecting a face, or increased contrast in shadows on the face, or maybe the picture is blurry, or the person is wearing glasses, etc.

3. Face Recognition Using Deep Learning and OpenCV

The objectives of the thesis are listed below:

I. Evaluation of mechanical properties of FRP composites for flexural loading condition
II. Characterization of FRP composite materials and selection the suitable resin and reinforcement for the fabrication of composite bridge deck panel as a flexural member
III. Selection of proper geometrical profile for studies of GFRP bridge deck panels
IV. Fabrication of multi-cellular GFRP composite Structures by hand lay-up process

This part of the research is mainly based on the open source computer vision library called OpenCV. This part of the research is implementing OpenCV to perform face recognition. To build the collection, I performed face detection at first and then extracted facial embeddings from each face using deep learning and trained a face recognition model on those embeddings and then at last came up with the collection that finally recognize faces in both images and videos streams. While it was possible and easy to recognize faces using the OpenCV library, the fact that OpenCV itself is not responsible for identifying faces is a vibration. Along with OpenCV and deep learning, I have used another library called scikit-learn. This library is used in detecting faces, computing 128-d face embeddings to quantify a face, train a support vector machine (SVM) on the top of the embeddings to recognize faces in images and video streams.

How it works Deep learning is applied in two simple steps in order to build the OpenCV face recognition collection pipeline. The first step is to apply the face detection which will detect the presence of the face in an image or video stream but will not identify it and the second step is to extract the 128d feature vectors that quantify each face in an image or video stream. These vectors are also known as embeddings. One can perform fast and accurate face detection with OpenCV along with the pre trained deep learning face detector model that is shipped with the library.

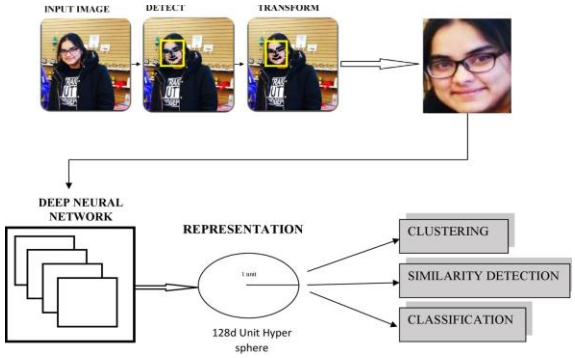

Figure 2: An overview of Face Recognition pipeline

Quantifying faces in an image is the responsibility of the model Facenet which I have discussed above. First step in the Face Recognition pipeline is to input an image or the video stream to the pipeline on which face detection is applied to detect the location of the faces in the image or the video stream. Along with face detection, it also computes facial landmarks to preprocess image and align the image. The second step after the face detection and cropping of the image is to input the cropped image to our deep neural network which calculates the 128-d embeddings to quantify the face itself.

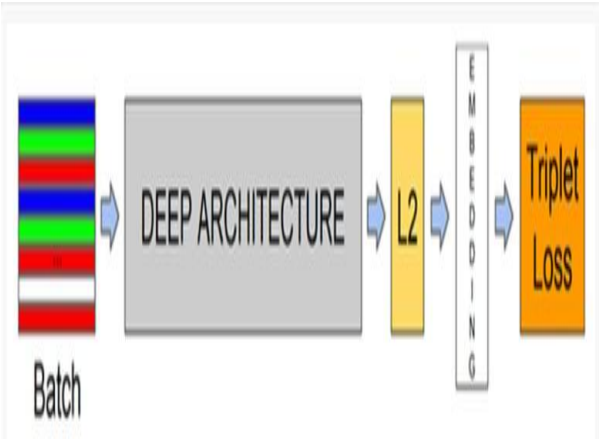

Figure 3: Deep Learning Face Recognition Model computing the Face Embeddings.

The dataset I have contains three people i.e. myself, my friend (Mayukh) and unknown (actress). Each class contains the total of six images (shown four) but I will suggest you to take at least 10-20 images per person you wish to recognize. Our face recognition dataset shown below.

Figure 4: Face Recognition dataset

3.1 Deep learning with tensor flow (creating the environment)

To create the environment for our face recognition collection, we need anaconda for python and CUDA toolkit. In anaconda, we either can create environment for the face recognition collection to run on the CPU or the GPU. On my windows, I was able to fetch 14FPS and on my Mac book Pro, I was able to fetch 16FPS.
After extracting the facial embeddings, we will train our face recognition model. Now we have to train our standard model so that it would recognize the person based on the embeddings extracted by the embedder. After loading the facial embeddings and encoding our labels, I have trained the SVM model for face recognizing. After I have trained our models, the pickle files for model and encoder will be generated.

```
# train the model used to accept the 128-d embeddings
of the face and # then produce the actual face
recognition
print("[INFO] training model...")

recognizer = SVC(C=1.0, kernel="linear",
probability=True) recognizer.fit(data["embeddings"],
labels)
```

Now is the time to recognize faces with OpenCV, so after loading the three models, I will load the images and detect faces. In the output, I am drawing rectangle around the face and placing the text above the face which includes name and the probability. The more the

images to train, the more the accuracy rate is.

Figure 5: The output shows that the face is of Tanveen with 53.83% probability.

Figure 6: The output shows that the face is of Mayukh with 46.65% probability.

3.2 Face Recognition in live video stream using live webcam.

The face recognition basically follows the same procedure except for feeding in the image to recognize, we have used OpenCV to start capturing videos using the live webcam by initializing the "Video Stream" object. Also, we have the system capturing the frames per second.

Video 1: Outputs to the face recognition using live camera.

Image Registration in Deep Learning

The process of aligning the two images or more that belongs to the same scene is known as Image Registration. One of the images is called the base image and the other images are called the input images. The base image is the reference to which the other images are compared. The objective of the image registration process is to bring the input image into alignment after applying spatial transformation to it. In image processing, the introductory step is the process of the image registration

registration. In medical field, the images that are created by different medical diagnostic modalities such as the process of Magnetic Resonance Imaging (MRI) or Single-Photo Emission Computed Tomography (SPECT) uses the image registration to compare various attributes in the images to see if the tumor is visible or not in the MRI or SPECT images. In computer vision, tasks like shape recovery, automatic change detection, automatic quality inspection, motion tracking, target template matching, are accomplished using image registration after alignment of the images has been achieved using different methods like Point based methods, Geometrical transformation, Surface or Intensity based method.

4. How it works

The process of face alignment is implemented using OpenCV and facial landmarks i.e. the input coordinates. The objective is to deform the input image and transform it to the output coordinate space after feeding in the input coordinates. The output coordinate space should consists of faces that are centered in the image, rotated in a way such that the eye lie in the horizontal line, and be scaled in a way that the sizes of the faces are almost identical. Thus, we can say that the process of face alignment is the type of data normalization. One can compare face alignment with the process of normalizing a set of feature vector via zero centering or scaling to unit norm prior to training the machine learning model. I have achieved higher accuracy from my face recognition model after performing this process. The reason behind me performing this normalization is the other facial recognition algorithms like Eigen faces, LBPs for face recognition, Fisher Faces and deep learning/metric methods. These methods aim at benefiting from applying facial alignment before even trying to identify the face.

5. Experiments and Result

The growing interest in computer vision of the past decade. Fueled by the steady doubling rate of computing power every 13 months, face detection and recognition has transcended from an esoteric to a popular area of research in computer vision and one of the better and successful applications of image analysis and algorithm based understanding. Because of the intrinsic nature of the problem, computer vision is not only a computer science area of research, but also the object of neuroscientific and psychological studies, mainly because of the general opinion that advances in

computer image processing and understanding research will provide insights into how our brain work and vice versa. Because of general curiosity and interest in the matter, the author has proposed to create an application that would allow user access to a particular machine based on an in-depth analysis of a person's facial features. This application will be developed using Intel's open-source computer vision project, OpenCV and Microsoft's .NET framework.

The goal of this article is to provide an easier human-machine interaction routine when user authentication is needed through face detection and recognition. With the aid of a regular web camera, a machine is able to detect and recognize a person's face; a custom login screen with the ability to filter user access based on the users' facial features will be developed. The objectives of this thesis are to provide a set of detection algorithms that can be later packaged in an easily portable framework amongst the different processor architectures we see in machines (computers) today. These algorithms must provide at least a 95% successful recognition rate, out of which less than 3% of the faces detected are false positive.

Facial recognition generally involves two stages:

- Face Detection where a photo is searched to find a face, then the image is processed to crop and extract the person's face for easier recognition.
- Face Recognition where that detected and processed face is compared to a database of known faces, to decide who that person is. Since 2002, face detection can be performed fairly easily and reliably with Intel's open source framework called OpenCV.

This framework has an in-built Face Detector that works in roughly 90-95% of clear photos of a person looking forward at the camera. However, detecting a person's face when that person is viewed from an angle is usually harder, sometimes requiring 3D Head Pose Estimation. Also, lack of proper brightness of an image can greatly increase the difficulty of detecting a face, or increased contrast in shadows on the face, or maybe the picture is blurry, or the person is wearing glasses, etc.

6. Conclusion

To improve the recognition performance, there are MANY things that can be improved here, some of them being fairly easy to implement. For example, you could add color processing, edge detection, etc.

You can usually improve the face recognition accuracy by using more input images, at least 50 per person, by taking more photos of each person, particularly from different angles and lighting conditions. If you can't take more photos, there are several simple techniques you could use to obtain more training images, by generating new images from your existing ones:

You could create mirror copies of your facial images, so that you will have twice as many training images and it won't have a bias towards left or right.

You could translate or resize or rotate your facial images slightly to produce many alternative images for training, so that it will be less sensitive to exact conditions.

You could add image noise to have more training images that improve the tolerance to noise. It is important to have a lot of variation of conditions for each person, so that the classifier will be able to recognize the person in different lighting conditions and positions, instead of looking for specific conditions. But it's also important to make sure that a set of images for a person is not too varied, such as if you rotated some images by 90 degrees. This would make the classifier to be too generic and give very bad results, so if you think you will have a set of images with too much variance (such as rotation more than 20 degrees), then you could create separate sets of training images for each person.

Reference

[1] [Ale01] Alex Freitas, "A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery", Postgraduate Program in Computer Science, Pontificia Universidade Catolica do Parana Rua Imaculada Conceicao, 1155. Curitiba - PR. 80215-901. Brazil, 2001

[2] [And05] Andreas L, "Agent Intellegence Through Data Mining", Symeonidis Pericles A. Mitkas, Springer, 2005.

[3] [Bin07] Bing Liu, "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data", Department of Computer Science, University of Illinois at Chicago, Springer-Verlag Berlin Heidelberg, 2007.

[4] [Cha98] Charu C. Agrawal and Philip S. Ya, "Mining Large Itemsets for Association Rules", Bulletin of the IEEE computer society Technical Committee on Data Engineering, 21(1) March 1998.

[5] [Dav01] David Hand, et al., "Principles of Data Mining", MIT Press 2001.

[6] [Dha03] Dhananjay R. Thiruvady, BComp, "Mining Negative Rules in Large Databases using GRD", M.sc thesis, School of Computer Science and Software Engineering at Monash University, November, Copyright by Dhananjay R. Thiruvady, 2003.

[7] [Dan05] Daniel T. Larose, "Discovering Knowledge in Data An Introduction to Data Mining", by John Wiley & Sons, 2005.

[8] [Da05] Da Ruan Guoqing Chen Etienne E. Kerre Geert Wets (Eds.), "Intelligent Data Mining Techniques and Applications", Springer, 2005.

[9] [Dan06] Daniel T. Larose, "Data Mining Methods and Models", by John Wiley & Sons, Inc, 2006.

[Fan06] Fan Jianhua & Li Deyi, "An Overview of Data Mining and Knowledge Discovery", 2006.

[11]
[12] [Ger02] German Florez, Susan M. Bridges, and Rayford B. Vaughn, “An Improved Algorithm for Data Mining for Intrusion Detection”, National Science Foundation Grant# CCR-9988524 and the Army Research Laboratory Grant# DAAD17-01-C-0011, 2000.
[13] [Jef06] Jeffery W. Seifert, ‘Data Mining and Homeland Security: An overview’, CRS Report for congress, January 2006, www.fas.org/sgp/crs/intel/R131798.pdf.
[14] [Jia06] Jiawei Han Micheline Kamber, “Data Mining: Concepts and Techniques”, Elsevier, 2006.
[15] [Lip05] Lipo Wang, Xiuju Fu, “Data Mining with Computational Intelligence”, Springer 2005.
[16] [Mig03] Miguel Delgado, et al., “ Association Rules: General Model and Applications”, IEEE Transactions on Systems, VOL. 11, NO. 2, April 2003.
[17] Singh, Harsh Pratap, et al. "AVATRY: Virtual Fitting Room Solution." 2024 2nd International Conference on Computer, Communication and Control (IC4). IEEE, 2024.
[18] Singh, Nagendra, et al. "Blockchain Cloud Computing: Comparative study on DDoS, MITM and SQL Injection Attack." 2024 IEEE International Conference on Big Data & Machine Learning (ICBDML). IEEE, 2024.
[19] Singh, Harsh Pratap, et al. "Logistic Regression based Sentiment Analysis System: Rectify." 2024 IEEE International Conference on Big Data & Machine Learning (ICBDML). IEEE, 2024.
[20] Naiyer, Vaseem, Jitendra Sheetlani, and Harsh Pratap Singh. "Software Quality Prediction Using Machine Learning Application." Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 2. Springer Singapore, 2020.
[21] Pasha, Shaik Imran, and Harsh Pratap Singh. "A Novel Model Proposal Using Association Rule Based Data Mining Techniques for Indian Stock Market Analysis." Annals of the Romanian Society for Cell Biology (2021): 9394-9399.
[22] Md, Abdul Rasool, Harsh Pratap Singh, and K. Nagi Reddy. "Data Mining Approaches to Identify Spontaneous Homeopathic Syndrome Treatment." Annals of the Romanian Society for Cell Biology (2021): 3275-3286.
[23] [Moh08] Mohammad Amin Rigi, et al. , “An Evolutionary Data Mining Model for Concept Extraction”, Department of Computer Eng.Ferdowsi University, Mashhad, Iran, IEEE, 2008.
[24] [Ode08] Oded Maimon, Lior Rokach, “Soft Computing for Knowledge Discovery and Data Mining”, Springer, 2008.
[25] [Pet98] Peter Cabena, et al., “Discovering Data from Concept to implementation”, Prentice-Hall Inc., 1998.