UNIVERSITÀ DI PISA

Business and Project Management A.A. 2021/2022

Candidates: Stefano Dugo and Gabriele Marino

Report and analysis on Face Recognition through NLP

1. Problem Presentation

The main topic we are going to analyze in this report is face recognition, a popular image processing technique that has been utilized in many applications, software and devices for multiple usages.

Face Recognition is an interesting topic because in the last years has gained more and more attention in many different application fields. It is becoming slowly and slowly an interesting technology in marketing: the global facial recognition market size was estimated to at USD 3.86 billion in 2020 and is estimated to grow even more in the next few years.

In this report we are going to expose the general workflow and themes behind face recognition systems, what are the main topics of the technology and how it can be relevant for business, showing some of the application areas of interest.
We will try to answer to two questions in particular:

● Which are the main topics and themes in documents related to technology?
For understanding how we can place face recognition software in the market we first need to find out what are the main aspects that characterize this kind of technology.

● Who are the users of technology?
We want to find the possible buyers or final users of face recognition software, which can be the final customers or firms that can make a certain usage of it through embedding this technology in their products.

By answering to these questions, we want to understand and explain how it is possible to sell in the market this kind of technology and why it can be useful to users.
For our analysis as we are going to explain in detail, we used the data that were made available to us and we tried to extract the information relevant for our presentation through NLP techniques, in order to show the main features of the face recognition systems.

The code, data and documents used are available in the following Github repository: https://github.com/gabrielemarino-gm/Business-ProjectManagement-project .

2. NLP and Analysis

In this chapter we are going to explain how we exploited NLP for going through our analysis on this technology, in order to extract all the key arguments and points of interest for developing and inserting into the market face recognition-related software.

2.1 Patent Extraction

We started by extracting information about face recognition technologies from the material we had available; after parsing all the available data we concluded that the data we needed could be extracted from the available patents.
For each patent we decided to extract information regarding explicitly “face recognition”:

```
import os
import re
import shutil
path = “path of the directory in which the data are in”
# getting all files from the directory given by the path
files = os.listdir(path)
# moving to the desired directory
os.chdir(path)
text = ""
cur_text = ""
for filename in files:
   file = open(filename, encoding="utf-8")
   cur_text = file.read(); # read file content
   # we search for a match for "face recognition" in the current file
   match = re.findall(r'\bface recognition\b', cur_text)
   if match: # this means: if match exists
      text += file.read() # all the text with the desired content is put together
      # we define the original path of the file and the path in which we will copy it to
      origin = “source directory”+filename
      target = “target directory”
      shutil.copyfile(origin, target) # copies the file from path origin to path target
```

2.2 Patent Processing

For each patent we first extracted the whole text and put it together:

```
import os
import re
path = “path of the directory containing the data”
# getting all files from the directory given by the path
files = os.listdir(path)
# moving to the desired directory
os.chdir(path)
text = ""
for filename in files:
   file = open(filename, encoding="utf-8")
   text+=file.read();
```

The patents have a markup format with tags, so we decided to extract the information of two specific tags: abstract and claims, which should contain the information we are interested in. For this purpose, we extracted the text between these two tags from the previously extracted text, using regular expressions:

```
p1 = ""
p1=re.findall(r'(?<=<abstract>\n)(?s:.*?)(?=\n</abstract>)',text)

p2 = ""
p2=re.findall(r'(?<=<claims>\n)(?s:.+?)(?=\n</claims>)',text)
```

For each of the following steps we will use as example the abstract text mining, but the steps are identical for the claims text mining as well.
The following step is **text cleaning**; for doing that we performed on both the sections the following techniques:

- Text lowering: for having a uniform text without upper case letters

```
lower_atext = ""
for abstract_text in p1:
#we pick each word and add to a variable, which will contain all the text
  lower_atext += abstract_text.lower()
lower_atext
```

- Whitespace removal

```
def remove_whitespace(text):
   return  " ".join(text.split())

lowera_text = remove_whitespace(lower_atext)
```

- Punctuation removal: we also decided to remove some punctuation and also the numbers, since in the claims they were used as indexes for lists of features of the technology, and were not necessary because not related to the data we are looking for

```
for char in '?.,!/;:()1234567890':
   lowera_text =
lowera_text.replace(char,'')
```

Then we exploited two libraries for keyword extraction for a further cleaning of the text; we did this because with a preliminary keyword extraction we can select the words that are the most redundant among the patents, and so they are not needed for our analysis because they wouldn’t give us much valuable info.

We exploited these two Keyword extraction libraries: PKE and KeyBERT (they will be used also later).

With PKE, we selected the best 10 candidates of the text and removed them; we also removed the 5 keywords obtained with KeyBERT.

PKE Workflow:

```
import pke

# initialize keyphrase extraction model, here TopicRank
print("Initializing extractor...")
extractor = pke.unsupervised.TopicRank()

# load the content of the document, here document is expected to be in raw
# format (i.e. a simple text file) and preprocessing is carried out using spacy
print("Loading text...");
extractor.load_document(input=lowera_text, language='en')

# keyphrase candidate selection, in the case of TopicRank: sequences of nouns
# and adjectives (i.e. `(Noun|Adj)*` )
print("Candidate Selection...")
extractor.candidate_selection()

# candidate weighting, in the case of TopicRank: using a random walk algorithm
print("Weighting...")
extractor.candidate_weighting()

# N-best selection, keyphrases contains the 10 highest scored candidates as
# (keyphrase, score) tuples
print("Selecting 10 best candidates...")
keyphrases = extractor.get_n_best(n=10)
#delete the keywords from the text
for tuple in keyphrases:
   print(tuple[0])
   lowera_text = lowera_text.replace(tuple[0],'')
print(keyphrases)
```

KeyBERT workflow:

```
from keybert import KeyBERT
kw_model = KeyBERT()
keywords = kw_model.extract_keywords(lowera_text)
for tuple in keywords:
   print(tuple[0])
   lowera_text = lowera_text.replace(tuple[0],'')

print(kw_model.extract_keywords(lowera_text, keyphrase_ngram_range=(1, 1),
stop_words=None))
```

The next step is **tokenization**; we use nltk to tokenize the text obtained so far:

```
import nltk
# the output is a list, where each element is a sentence of the original text
nltk.sent_tokenize(lowera_text)
tokenized_text_a = nltk.word_tokenize(lowera_text)
```

Then we performed the **removal of the stop words**; we used the English stop words since the patent are written in English:

```
from nltk.corpus import stopwords
stopwords_en = stopwords.words('english')
# we prepare a empty list, which will contain the words after the stopwords
removal
tokenized_vector_a = []
# we iterate into the list of tokens obtained through the tokenization
for token in tokenized_text_a:
        # if a token is not a stopword, we insert it in the list
        if token not in stopwords_en:
            tokenized_vector_a.append(token)
# the output is a list of all the tokens of the original text excluding the stopwords
```

After this we did the **POS tagging** to assign each word to its part-of-speech tag and then we cleaned again and simplified the POS text:

```
pos_tagging_a = nltk.pos_tag(tokenized_vector_a)
cleaned_POS_text_a = []

for tuple in pos_tagging_a:
  # POS tagged text is a list of tuples, where the first element tuple[0] is a token and the
second one tuple[1] is
  # the Part of Speech. If the POS has length == 1, the token is punctuation, otherwise it is not,
and we insert it
   # in the list cleaned_POS_text
  if len(tuple[1]) > 1:
    cleaned_POS_text_a.append(tuple)

def simpler_pos_tag(nltk_tag):
  if nltk_tag.startswith('J'):
    return "a"
  elif nltk_tag.startswith('V'):
    return "v"
  elif nltk_tag.startswith('N'):
    return "n"
  elif nltk_tag.startswith('R'):
    return "r"
  else:
    return None

simpler_POS_text_a = []

# for each tuple of the list, we create a new tuple: the first element is the token, the second is
# the simplified pos tag, obtained calling the function simpler_pos_tag()
# then we append the new created tuple to a new list, which will be the output
for tuple in cleaned_POS_text_a:
  if tuple[1] == 'NNP':   #this is because there is some text in japanese categorized as 'NNP';
 #no other relevant words are categorized in such a way
     continue;
  POS_tuple = (tuple[0], simpler_pos_tag(tuple[1]))
  simpler_POS_text_a.append(POS_tuple)
```

The next thing we did is **lemmatization**, so that we can obtain the vector with all the relevant lemmas on which we can then finally perform the keyword extraction.

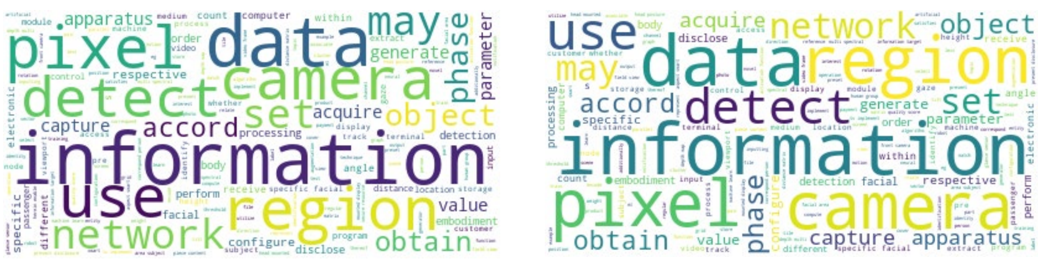
```
from nltk.stem.wordnet import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
lemmatized_text_a = []

for tuple in simpler_POS_text_a:
  if (tuple[1] == None):
    lemmatized_text_a.append(lemmatizer.lemmatize(tuple[0]))
  else:
    lemmatized_text_a.append(lemmatizer.lemmatize(tuple[0], pos=tuple[1]))
lem_text_a = ""
for abstract_text in lemmatized_text_a:
  lem_text_a += abstract_text + " "
#we pick each word and add to a variable, which will contain all the text
```

The final step of patent's text mining is the **KeyWord Extraction**, for which we used again the PKE library, setting for a high, but not too much, number of candidates in order to find the most interesting topics and argument described in the patents. The code is the same as before except for the number of candidates, the input text and the keyword deletion part (obviously here we don't delete any keyword).

As result of this whole text mining pipeline, we obtained several keywords, consisting of the main terms and topics regarding face recognition technologies; we also exploited a word cloud library to try to show them, but the results were more generic and less relevant than the ones obtained with the keyword extraction library.

```
from wordcloud import WordCloud
# Create a WordCloud object
wordcloud = WordCloud(background_color="white", max_words=5000,
contour_width=3, contour_color='steelblue')
# Generate a word cloud
wordcloud.generate(lem_text_a)
# Visualize the word cloud
wordcloud.to_image()
```



Word clouds of the abstract (left) and claims (right) of the patents.

This is a list of some of the keywords we obtained in this way (we considered the most interesting ones among the top keywords given in output by the algorithm):

- fusion method
- vergence distance
- location
- compute system
- position
- video frame
- body information
- facial feature classifier
- salient data indicative salient event
- posture
- face embodiments
- field view head-mounted display
- face cluster
- keypoint heatmap
- specific facial neural aggregation network

Some of the keywords obtained are not much self-explanatory, since they regard specific algorithms and techniques used in the face recognition field; we selected these words in order to find some papers which could help with our analysis and for treating all the key aspects of the technology for a better understanding of its role in business.

The topic we selected are:

- Fusion Methods
- Keypoint Heatmap
- Neural Aggregation Networks
- Face Clustering

3. Main themes of the technology

We still need to understand how this technology can be implemented for applications that could be sold to the market; for this reason, we are going to explain in some details (avoiding going too deep into the AI algorithms) the workflow and themes of the technology and to analyze more specific topics that we obtained through keyword extraction, based on the available papers regarding these topics.

3.1 Face Recognition Systems

A complete face recognition system includes two steps, face detection and face recognition. A great number of successful face recognition systems all adopted the linear discriminant analysis (LDA) approach to enhance class separability of all sample images for recognition purposes.

There are two main categories of face recognition applications, for instance it has recently seen a lot of success in a family of less-demanding applications such as online image search and family photo album organization (e.g., Google Photo, Microsoft Photo Gallery, and Apple iPhoto, and, at the other hand, there are the terrorist watchlist and mass surveillance applications that have dominated the field of face recognition research. However, there are many face recognition applications that fall roughly between these extremes, where very high recognition performance is desired, but the users in the gallery are still allies of the system rather than adversaries. These applications include access control for secure facilities (e.g., prisons and office buildings), computer systems, automobiles, or automatic teller machines, where controlled gallery images can be obtained in advance. These applications are very interesting due to their potential sociological impact.

Another application field in which face recognition has caught more attention is the society of network multimedia information access, for instance network security, content indexing and retrieval, and video compression, which all benefit from face recognition technology because "people" are the center of attention in a lot of videos. Network access control via face recognition not only makes hackers virtually impossible to steal one's "password", but also increases the user-friendliness in human-computer interaction. Indexing and/or retrieving video data based on the appearances of specific people it's useful for users such as news reporters, political scientists, and moviegoers. For the applications of videophone and teleconferencing, the assistance of face recognition also provides a more efficient coding scheme.

Face recognition technology is also an important research project in the field of computer vision and pattern recognition, it can identify the identities and other information according to the visual features of face image, having a very broad prospects for development. It is widely used in authentication, criminal investigation, video surveillance, robot intelligence and medical science and so on. It has wide application value and commercial value. As a biological feature, facial features have the characteristics of good, direct and convenient compared with other biological features. Therefore, face recognition is more acceptable for users.

Face recognition has the potential of working with test data that are much less controlled than other biometrics such as fingerprints and iris recognition, allowing the access control system to be made less intrusive to the users of the system. Very few recognition systems specifically target applications where many well-controlled training images are available.

In most cases, a face recognition algorithm can be divided into the following functional modules: a face image detector finds the locations of human faces from a normal picture against simple or complex background, and a face recognizer determines who this person is. Both the face detector and the face recognizer follow the same framework; they both have a feature extractor that transforms the pixels of the facial image into a useful vector representation, and a pattern recognizer that searches the database to find the best match to the incoming face image. The difference between the two is the following: in the face detection scenario, the pattern recognizer categorizes the incoming feature vector to one of the two image classes: “face” images and “non-face images. In the face recognition scenario, assuming that the feature vector is in the form of face images, the recognizer classifies it as “Smith’s face”, “Jane’s face”, or some other person’s face that is already registered in the database.

The variations in facial images could be categorized as follows:

• Camera distortion and noise

• Complex background

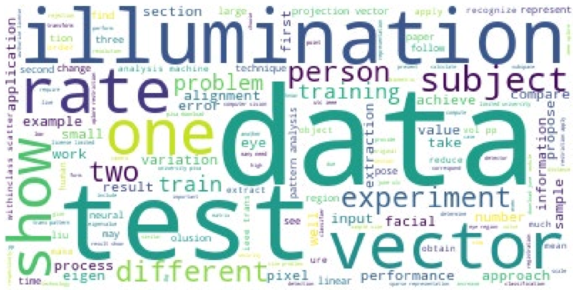
• Illumination

• Translation, rotation, scaling, and occlusion

• Facial expression

• Makeup and hair style

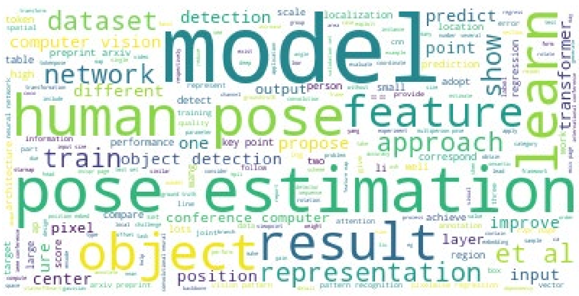
Camera distortion and noise are standard variations in image recognition problems. Previous researchers have developed numerous tools to increase the signal-to-noise ratio. To deal with complex image background, the recognizer requires a good face detector to isolate the real faces from other parts of the image. Illumination is often a major factor in the obstruction of the recognition process.



Word cloud of the face recognition systems papers.

3.2 Keypoint Heatmap

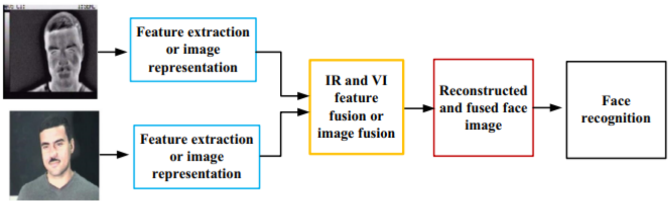
The keypoint heatmap is crucial for estimating the human pose, which is an important requirement for face recognition applications. One way to perform this task is through a typical bottom-up human pose estimation framework, which includes two stages, keypoint detection and grouping; applying these stages together with the tradeoff heatmap estimation loss for balancing the background and keypoint pixels and thus improving heatmap estimation quality, we get the state-of-the-art bottom-up human pose estimation result. Other approaches that improve the performance of the bottom-up framework have also been proposed.



Word cloud of the Keypoint Heatmap papers.

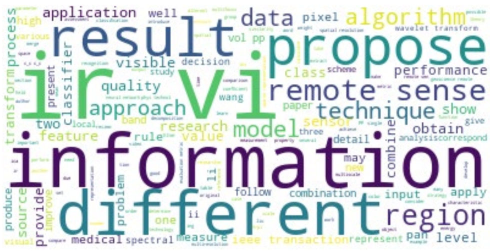
3.3 Fusion Methods

Fusion methods are a family of algorithms which are used to get improved performance in the face recognition task. Even though face recognition techniques of visual images have reached a significant level of maturity with many practical successes, the recognition rate of visual face-based method may degrade under poor illumination conditions or for subjects of various skin colors. Infrared (IR) and visual (VI) image fusion is designed to fuse multiple source images into a comprehensive image to boost imaging quality and reduce redundancy information, which is widely used in various imaging equipment to improve the visual ability of human and robot.



The idea of the image fusion was used for periocular region-based person identification. VI image-based face recognition techniques have been well studied, but there still some problems needed to be solved in practical applications. In real scenario, illumination, shooting angle, facial expression variation, decorations, background and so on, which will seriously impact the recognition effect. While, IR image will provide complementary information for face recognition, this information will be not revealed in VI image. Due to the characteristics of IR image, the IR and VI image fusion-based face recognition techniques will have a good performance when there is no control over illumination or for detecting disguised faces.

Illumination plays an important role in the efficiency of face recognition on visible images. Infrared image is independent of the ambient illumination, but it is sensitive to temperatures. Face recognition algorithms applied to the fusion of IR and visible images consistently demonstrated better performance than when applied to either visible or IR imagery alone.



Word cloud of the fusion methods papers

3.4 Neural aggregation networks

Neural aggregation networks are a family of methods which is nowadays very commonly used for trying to solve the Facial Expression Recognition (FER) task. Analyzing this kind of approaches, we can have a more complete view of the things we can do with face recognition technologies, and these methods expose new challenges to overcome to develop and deploy a face recognition application.

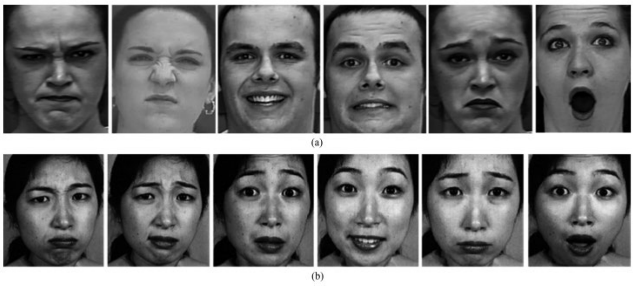
Video frames may suffer severe motion blur and out-of-focus blur due to camera jitter and small oscillation in the scene. One way to address the large variations in face quality is to select key frames and eliminate poor quality images; the recognition performance is undermined by removing low quality images.

There are some available solutions for solving some typical issues of the FER task:

Multi-mode Aggregation Recurrent Network (MARN): based on the fact that face recognition performance deteriorates when face images are of very low quality, but for low quality video sequences, more discriminative features can be obtained by aggregating the information in video frames.

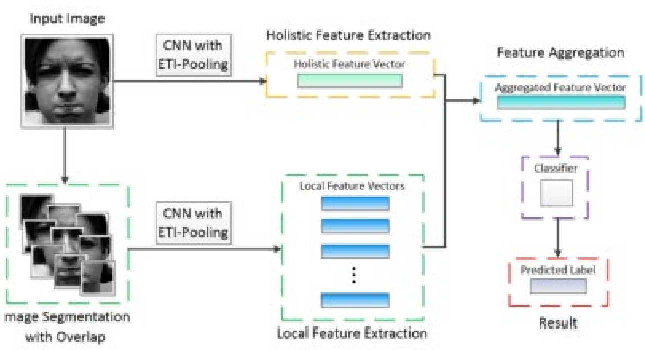
Deep Neural Networks (DNNs): it has been introduced since approaches to face recognition for high quality still images are not able to deal with challenges in face recognition in unconstrained videos; for instance, video sequences from CCTV are generally of lower resolution and may contain noisy frames with poor quality and unfavorable viewing angle, which will undermine the performance of the video face recognition if we directly apply the algorithms developed for recognizing high quality still images, even though they share some challenges with the video surveillance recognition.

Facial expressions are important carriers for human to convey emotions in communications. Study on nonverbal communication reveals that 55% of a person’s emotional or intentional information is conveyed through facial expressions. The main target of facial expression analysis is to establish a system that can automatically classify different expressions. In general, facial expressions can be categorized into six basic expressions, which include anger, disgust, fear, happiness, sadness and surprise. Therefore, the primary task of current expressional analysis is to classify these six basic expressions.



Example of the six prototypic expressions. (a) The CK+ dataset. (b) The JAFFE database. (left to right: anger, disgust, fear, happiness, sadness, surprise).

A method which has gained a lot of popularity and success in the last few years is Convolutional Neural Network (CNN): in FER problem, previous studies have revealed that expressional changes usually occur on some salient facial regions such as neighborhood of mouth, eyes and nose, implying that details of local facial regions can be discriminative for expressional recognition; most current CNN-based methods on FER only extract features from the whole expressional image. These methods emphasize the integrality of a facial expression but ignore the information of local details.



A challenging task is also unconstrained set-based face recognition, which defines the minimal facial representation unit as a set of images and videos instead of a single medium. This task raises the necessity to build subject-specific face models for each subject individually. The most significant challenge in the unconstrained set-based face recognition task is how to learn good representations for the media set, even in presence of large intra-set variance of real-world subject faces caused by varying conditions in illumination, sensor, compression, etc., and subject attributes such as facial pose, expression and occlusion.

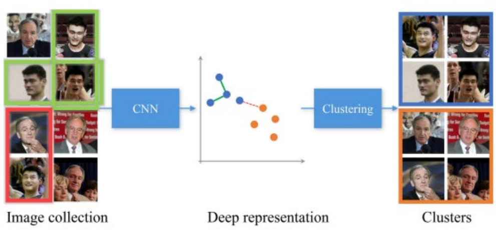
Word cloud of the Neural Aggregation Network papers.

3.5 Face Clustering

Face clustering is a key component either in image managements or video analysis. It tries to overcome the challenges of the possible variations of wild human faces: poses, expressions, and illumination changes, and noises, like block occlusions, random pixel corruptions, and various disguises.

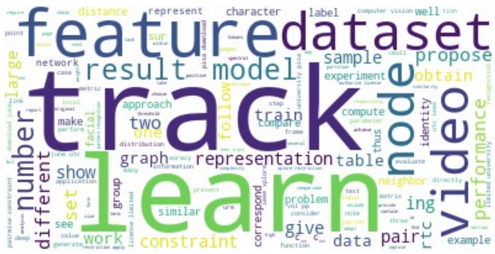
Face clustering aims to cluster the given facial images into their underlying groups. Example: to manage the human-centered photos, faces referring to the same people are usually first clustered together, and then face tagging techniques are applied to associate these faces with people names. Such systems have been successfully exploited in some commercial cases like Google Picasa, Apple iPhoto, and Microsoft Easy Album.

The main application area of face clustering is videos: for instance, it is an important matter for surveillance cameras, which are a popular security mechanism employed by government agencies and businesses alike, resulting in the capture of suspects based on their facial images in high profile cases (i.e., the 2013 Boston Marathon Bombing). Locating suspects’ facial images typically requires manual processing of large volumes of images and videos of an event, leading to the need for automatic processing of still images and videos to assist in forensic investigations.



Face clustering in videos can also be found in video indexing and content analysis, preprocessing for face recognition, determining the cast of a feature-length film, etc. The general workflow is the following: faces are organized as face tracks and the frame index of each face is provided.

One possible application of face clustering is understanding videos such as TV series and movies, which requires analyzing who the characters are and what they are doing, meaning clustering face tracks based on their identity. The challenge here is that the number of characters is not known a priori and tracks belonging to background characters are not discarded.



Word cloud of the Face Clustering papers.

4. How to insert the technology in the market

4.1 Main topics and application fields

After showing the whole picture behind the development of a face recognition application we have a clear view of the main topics of the technology, as we could see from the keywords extracted through the NLP analysis of the patents. The key points of face recognition technologies are:

- Video face recognition
- Still-image face recognition
- Posture Analysis
- Location Analysis
- Face feature extraction
- Face Expression Recognition

All these characteristics allow us to define a face recognition algorithm, combined with the specific techniques we discussed in the previous chapter, in order to develop useful applications, like some already available in the market.

This leads us to what are the main topics of the technology in terms of possible applications:

- **Personal Authentication**: applications like mobile phones authentication mechanisms and facial authentication for monetary transactions need to correctly identify the user; this can be achieved through FER and face feature extraction for accurately recognizing the face features; issues due to low quality frames and illumination can be dealt with the previously described video recognition techniques. There are also other kinds of authentication applications that can be developed with this technique; for instance, a bank can exploit face recognition for employee’s authentication.

- **CCTV security cameras**: it's one of the most targeted products of face recognition technologies, with multiple possible usages, such as security facilities; as described previously, video face recognition based on CNN are applied to deal with illumination issues and blurred frames.

- **Criminal activities detection**: in this case there is an additional challenge, which is the task to identify a person in a crowd of people in different locations; for this aim face recognition technologies include posture analysis techniques, which exploit keypoint heatmap methods, to correctly identify the posture of a person even in a crowd; combining this to the other face recognition techniques, an application of this kind can be implemented.

- **Mobile Phones Photo Applications**: in these applications the people in the photos are recognized and identified correctly, whatever their location, posture and expression is; an example of this is Apple iPhoto: applications like this identify a person with their name and put together all their photos in the phone's gallery.

- **Emotion Recognition:** through FER, the face recognition technologies can be exploited for distinguish human emotions; this could be used for improving customer-service based on their feelings in many possible fields.

- **Movie casting:** as previously described, applying video face recognition through face clustering it's possible to identify a pattern in characters faces and use them to build a model for predicting the role for a movie or tv-series casting.

4.2 Main Users of the technology

Based on the applications we discussed, we can try to answer to the question: which are the users of the technology? After analyzing the topics, we concluded that some of the best candidates for being users of this technology are the following:

- **Banks:** this category may include different kinds of users, like the institution itself, which can exploit face recognition for employees' identification, or the users of the bank, which can identify themselves through a face authentication mechanism for being able to use the bank's services.
- **Mobile Phone Users:** as we previously described there are many photo applications and authentication mechanisms that can be implemented for smartphones.
- **Security Facilities:** the technology is used for monitoring people in this kind of facilities, like convicts or external people, so we can consider it as a direct user.
- **Newspapers reporters and analysts, Police, Investigation institutions:** these categories may exploit face recognition for criminal activities detection.

There are also many other possible users, since this field is continuously developing and it’s an important subject for researchers; it could be used for car users, therapies (in this case we refer specifically to the facial expression recognition for emotion analysis), video games players and so on.

4.3 Possible Strategies

For concluding this report, we wanted to analyze what are the possible strategy for designing, developing, and selling to the market a face recognition application, given all the information we gathered and the main topics of the technology. As we have seen so far there are already many available applications which exploit this technology and there are many other possibilities to make new ones. For creating a face recognition application which can be interesting in the market, we should take in consideration the two kinds of differentiation strategies we can apply:

- **Improve current solutions:** since most of the challenges of face recognition regard the environment illumination, the blurred video frames and other camera variations, the color of the skin or any kind of disguise on the face, the most straightforward approach is developing a software which improves the performance of the ones used for applications in the market; this could be achieved by trying different datasets for training or different feature extraction and selection algorithms; another way could be improving the tradeoffs between efficiency and costs, for instance in the case of surveillance cameras, which means to find an algorithm that has a better recognition rate in proportion to the cost of the devices used.

- **Propose new applications:** since there have been already reached good results for the currently available applications, researchers are testing face recognition in many different fields in order to eventually make possible to develop new applications. One of the most interesting “Blue Oceans” we can explore is the Emotion Recognition: as we saw this is a very important field in research, because it could be very important for medical applications in order to make therapies based on the facial expressional behavior of the patients.