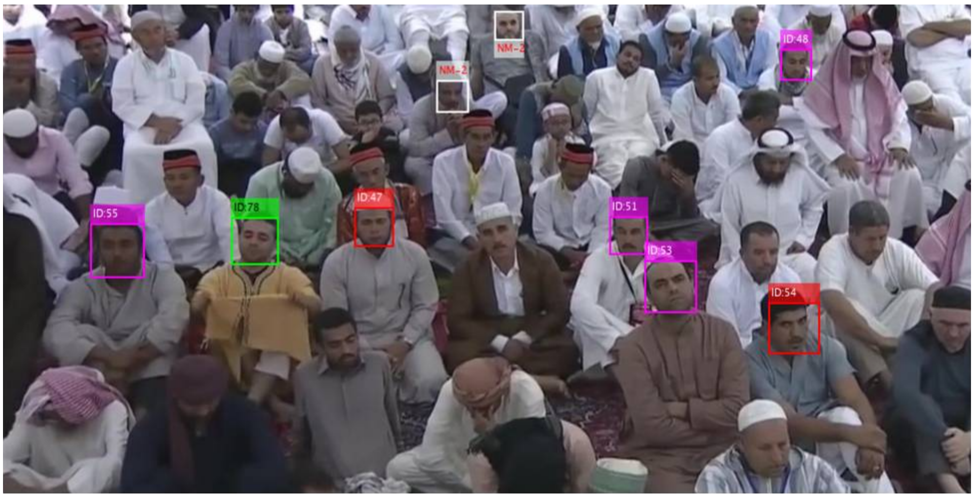
UNIVERSITAT DE BARCELONA

Artificial Vision Miniproject
Dynamic facial recognition in crowded environments



Adrià Alfonso, Noah Márquez, Kamil Natan Malysa, Jiaqi Li

December 2023

CONTENTS

1 Abstract

This miniproject covers a possible application of artificial vision algorithms to help locate and track missing people through video surveillance across large areas with dense crowds [1]. We will study how applying facial detection and recognition algorithms to video footage of connected cameras can be used to detect lost individuals within large crowds.

2 Context

In large gatherings, people travelling together are always at risk of being separated while attempting to traverse dense crowds. In groups such as families this is of particular concern as locating a child or a person with a disability separated from the group can be difficult with no method of communicating with them.

This miniproject covers a possible solution to this problem, tested in the context of pilgrim gatherings in the Al-Nabawi Mosque in Madinah. The Al-Nabawi is the second-largest mosque in Islam and is large enough for 1 million pilgrims to visit at a time; the sheer density of pilgrims makes getting separated from a group a real problem.

3 Goal

The goal of the application is to facilitate a service where groups of people can report missing peers for them through a mobile phone app or web service to be searched and located across a large network of cameras automatically, with the intention of reuniting them.

To accomplish this, the cameras will employ face detection and recognition algorithms, reporting the presence as well as approximate position of any person reported lost.

This speeds up the search for lost people by narrowing down where they could be within the area covered by the cameras, as well as tracking their movement across the area, thus providing both spatial and temporal information on their whereabouts.

4 Implementation & solution

The solution consists of using multiple facial detection and recognition algorithms on video surveillance footage. Locating a lost person requires being able to search a wide area; for this, the mosque the research was conducted on, was divided into 25 regions ("geofences") with multiple IP cameras (network-attached cameras) installed in each.

Given the high amount of pilgrims in each location, to be able to detect faces in real-time, the cameras capture relatively low-resolution shots to speed up processing of the facial detection algorithms. These algorithms as such must be robust even if faces consist of a few pixels.

The facial detection algorithm employed consists of combining multiple detectors based on Viola Jones's classifier cascades to achieve greater accuracy than using individual detectors.

Aside from the Haar-based facial detector we've seen in class, two additional detectors are used:

- Local Binary Patterns (LBP) face detector
- Classification and Regression Tree (CART)

1 LBP Face Detector

The LBP-based face detector uses Local Binary Patterns as a texture descriptor for faces.

Creating an LBP-based descriptor for an image window consists of dividing it into various blocks of equal size. For each pixel within the blocks, an 8-bit LBP value is computed from the neighbour pixels. Finally, a histogram of the LBP values is created for each block, and the concatenation of the histograms forms the feature descriptor.

The feature descriptor is then classified as a face or non-face using a cascade of Adaboost classifiers, just like in the Viola-Jones detector. The classifiers in this case are created from the characteristics of the LBP descriptors of sample images - in this case, the people entering the mosque who have agreed to have their photo taken for this purpose.

Recognition is afterwards done by comparing the distance of the descriptors of detected faces to the ones in the database.

2 CART Face Detector

Classification and Regression Trees (CART) are a machine learning technique that can be used to build classifiers - in this case to classify windows of the image as containing a face or not, as well as for facial recognition afterwards.

The CART algorithm is a top-down approach that consists of running the input through a binary decision tree, where the leaf nodes of the tree correspond to classifications of the input.

Each node corresponds to a feature being checked. Nodes are created based on the homogeneity of the features used. For example, if 90% of the samples of a class have a similar value for a feature, then a node is created in the decision tree that splits the classification based on that value.

The metric for deciding the features to use for the nodes of the decision trees is usually the Gini index, which measures the probability for a random sample being misclassified at each node. A lower Gini index indicates that a particular feature performs an accurate split of the data for classification.

CART makes classifying new samples very quick and performs feature selection automatically, however by itself creates models that are either simplistic or overfit. Since the decision tree is made with individual features, possible interactions between features are disregarded.

It's possible to also use a classifier cascade with CART, creating multiple decision trees of increasing complexity to incrementally discard non-faces in this case.

3 Detection and recognition

All 3 facial detectors are run every 10th frame of the video feed of each camera and their outputs are merged, detecting more faces than methods that rely on just one detector. Overlapping face regions are merged if the overlap of the bounding boxes is high enough.

The detected faces are then compared against the faces of the database to determine if the video footage contains a person detected as missing. This requires a face of the missing person to already be in the database, which could be attained by asking visitors of the mosque to have their photo taken and stored temporarily for this purpose during their visit.

To reduce the possibility of false-positives, a temporal factor is used: a detected face is only considered a true positive if it is consistently detected in the next few video frames analyzed.

After the faces are detected, they are run through multiple recognition algorithms: Principal Component Analysis (Eigenfaces), Discrete Cosine Transform (DCT), LBP descriptor comparison, and Adaptive Sparse Representations (ASR+).

5 Results

The method discussed was tested on 2000 video frames from the mosque, and the faces detected per-algorithm were recorded. The Haar and CART detectors performed better than LBP, however the combination of detected faces of all 3 yielded ~40% more faces detected than the use of Haar or CART by themselves. The use of classifier cascades for all three kept false positive rates and computational requirements low.

The concept of combining multiple algorithms was also applied to the recognition process, achieving accuracies of ~95% with as few as 3 training images per-person, greatly out-performing individual recognition algorithms. This was done by classifying faces based on the "majority vote" of the various recognition algorithms.

The comprehensive evaluation of these face detection algorithms demonstrates that while individual algorithms exhibit strengths, there are large benefits to collaborative integration of multiple of them. The combined approach leads to more robust and accurate face detection systems in real-world applications.

6 Relationship with the Artificial Vision Course

This project shows a direct application of several key concepts and algorithms discussed in the Artificial Vision course.

1 Feature Detection and Recognition

The Viola-Jones detector, a central focus of this project, is an embodiment of the edge and feature detection methodologies covered in the course. It leverages integral images for rapid computation, demonstrating the practical utility of discrete approximations in real-world applications.

2 Classifier Cascades

The use of classifier cascades in this project popularized by the Viola-Jones detector is a direct application of a key concept discussed in our Artificial Vision course. Classifier cascades, particularly in the realm of face detection, are an effective method for quickly discarding the many non-face regions in an image, significantly reducing the computational burden.

- In the Viola-Jones face detection framework, a cascade of increasingly complex classifiers is employed. This approach aligns with the course teachings on the structure of image processing and object detection.
- The LBP-based detector uses texture analysis in its cascades, demonstrating an alternative approach to face detection.
- The CART-based detector illustrates the application of machine learning techniques, highlighting the use of decision trees in facial detection and yet another application of the classifier cascade concept.

The cascading approach significantly improves processing efficiency and enhances accuracy, which is critical for real-time applications in crowded environments.

3 Texture Analysis and Shape Recognition

The LBP-based face detector is an application of texture analysis, a topic extensively covered in the course. By using texture descriptors for facial detection, we see a direct correlation between theoretical concepts and their application in distinguishing facial features in a dense crowd.

4 Eigenfaces and Dimensionality Reduction

The concept of eigenfaces and dimensionality reduction, as discussed in the course, finds practical application in enhancing the speed and accuracy of the facial recognition process. By reducing the redundancy in facial data, we can achieve faster processing times, crucial for real-time applications.

5 Challenges in Recognition and Detection

The project also touches upon the difficulties in facial recognition, such as occlusion and varying lighting conditions. These challenges resonate with the course content on problems and difficulties in biometrics and facial analysis. In the context of the study, this problem is alleviated by sampling faces every few video frames, and employing multiple cameras for each location - thus reducing the likelihood of faces being obscured.

7 Possible Extensions

This project opens several avenues for further exploration and development.

1 Crowd Dynamics and Safety

One significant extension is the use of facial detection algorithms for monitoring crowd dynamics. By analyzing the density and flow of crowds, we can predict and prevent dangerous situations, such as stampedes. This application is particularly relevant in the context of large gatherings, where safety is a paramount concern.

Managing large crowds presents a significant challenge, particularly in preventing overcrowding in open areas that can lead to dangerous stampedes, like the incident in Korea last Halloween [5].

To address this concern, we could track the congregation density in each zone. When a threshold is passed, immediate warnings could be issued to the crowd or real-time reports could be transmitted to the police in order to handle the problem before it's too late. We could also publicly report this information, so other visitors can know in advance how crowded the area will be. This is particularly useful in contexts where the amount of visitors present in the area is not already tracked through other means.

2 Integration with Other Biometric Technologies

Another extension could be the integration of facial recognition with other biometric technologies, such as gait analysis or voice recognition. This multi-modal approach could potentially increase the accuracy and reliability of identifying individuals in complex environments.

3 Real-Time Data Analytics for Public Safety

The system could be extended to provide real-time analytics on crowd density and movement patterns, offering valuable insights for emergency response and urban planning. This could be particularly beneficial in managing large events or responding to emergency situations.

4 Ethical Implications and Privacy Concerns

As we extend the application of these technologies, it becomes increasingly important to consider the ethical implications and privacy concerns associated with widespread surveillance. Developing guidelines and policies for the ethical use of AI in public spaces will be a critical part of this work.

5 Advanced Algorithms for Challenging Environments

Finally, the development of more advanced algorithms capable of handling highly dynamic and challenging environments (e.g., low light conditions, high crowd density) will be a key area of future research.

8 Problems and Challenges

- **Occlusion and Environmental Factors:**
  In crowded environments, occlusion poses a major challenge. Partially visible faces and varying environmental conditions like lighting and weather can significantly impact the effectiveness of facial detection algorithms.
- **Data Quality and Privacy Issues:**
  The quality of data, especially in low-resolution images, can hinder the accuracy of face recognition. Additionally, there are significant privacy concerns related to collecting and storing facial data, which need to be addressed with strict data governance policies.
- **Real-Time Processing Constraints:**
  The requirement for real-time processing imposes constraints on the complexity of the algorithms that can be used. Balancing accuracy with computational efficiency remains a key challenge.
- **Diverse and Changing Facial Features:**
  Human faces can vary significantly due to factors like age, facial hair, or cosmetics. Adapting algorithms to reliably recognize faces despite these variations is a complex task.
- **System Scalability and Reliability:**
  Ensuring the system's scalability to handle large volumes of data and maintaining consistent reliability across different scenarios are ongoing challenges in this domain.

9 Conclusions

This mini-project demonstrated the effective application of artificial vision techniques for the challenging task of locating missing people in crowded environments, specifically in the context of the Al-Nabawi Mosque.

The integration of multiple facial detection algorithms, namely the Haar-based detector, LBP, and CART, showcased how combining different methodologies can enhance overall detection accuracy. The use of classifier cascades was particularly effective in reducing computational load while maintaining high detection rates, crucial for real-time processing in densely populated areas.

Future work can explore integrating more advanced AI techniques, addressing privacy and ethical considerations, and expanding the system's application to other crowded environments for public safety and crowd management.

References

[1] Adnan Nadeem et al. *Tracking Missing Person in Large Crowd Gathering Using Intelligent Video Surveillance.* [Online]. 2022. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9323688/.

[2] Adnan Nadeem et al. *A Novel Integration of Face-Recognition Algorithms with a Soft Voting Scheme for Efficiently Tracking Missing Person in Challenging Large-Gathering Scenarios.* [Online]. 2022. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8838238/.

[3] Parth Singh. *Understanding Face Recognition using LBPH algorithm.* [Online]. 2021. URL: https://www.analyticsvidhya.com/blog/2021/07/understanding-face-recognition-using-lbph-algorithm/.

[4] Shailey Dash. *Decision Trees Explained — Entropy, Information Gain, Gini Index, CCP Pruning.* [Online]. 2022. URL: https://towardsdatascience.com/decision-trees-explained-entropy-information-gain-gini-index-ccp-pruning-4d78070db36c.

[5] *Estampida de Halloween en Seúl.* [Online]. 2023. URL: https://es.wikipedia.org/wiki/Estampida_de_Halloween_de_Se%C3%BAl.