AGORA International Journal of Economical Sciences, http://univagora.ro/jour/index.php/aijes
ISSN 2067-3310, E-ISSN 2067-7669
Vol. 18, No. 1 (2024), pp. 240-249

A COMPREHENSIVE EVALUATION OF FACE RECOGNITION SOFTWARE: BALANCING TOTAL COST OF OWNERSHIP, ACCURACY AND SPEED

K. TIRANA

Kledia Tirana
University Metropolitan Tirana, Albania
https://orcid.org/0009-0008-0772-4243, E-mail: ktirana@umt.edu.al

Abstract: *As face recognition technology continues to play a pivotal role in various domains, selecting an optimal software solution becomes imperative. This paper thoroughly analyzes face recognition algorithms, emphasizing a holistic assessment of implementing them that includes the overall cost of ownership, accuracy, and speed. The research aims to guide decision-makers in choosing a solution that balances these three critical factors. The study employs a rigorous methodology and evaluates various intrusion detection solutions across multiple industries. It examines the cost of ownership comprehensively, including initial investment, maintenance expenses, and potential hidden costs. A cost-benefit analysis is conducted to unveil the true economic implications of each implementation. The accuracy of the face recognition algorithms which is the core of intrusion detection systems is assessed through the learnings from academia and real-world feedback from the security industry. Furthermore, the paper examines the critical aspect of speed, recognizing its important role in real-time applications. The evaluation considers the processing speed of each software solution, considering its responsiveness to various environments and working with large-scale datasets. This research provides a comprehensive overview of face recognition algorithms, offering small business decision-makers valuable insights to make informed choices for their intrusion detection solutions. By combining the total cost of ownership, accuracy, and speed, organizations can select a solution that aligns with their specific needs and maximizes the return on investment in intrusion detection technology.*

Keywords: *Face recognition, Deep Learning, Object Detection, total cost of ownership.*

INTRODUCTION

The overall security of small businesses is very important. The enterprises are vulnerable to a range of threats that can severely impact their daily operations, including financial stability and reputation. Small businesses face a set of unique challenges in terms of security due to their limited resources, size, and specific operational characteristics. The most important challenges are related to limited budgets, resource constraints, lack of in-house expertise, and dependency on third-party services. Recognizing the challenges allows small businesses to tailor their security plans accordingly, emphasizing cost-effective solutions.

While it is impossible to address all the challenges, we will focus on physical security by providing a free intrusion detection solution. For our paper, we will use data that we have collected in the USA market. However, the adoption of sophisticated surveillance systems, leveraging cutting-edge technologies such as AI-powered analytics and cloud-based solutions, is not exclusive to the United States. In today's interconnected world, the USA's video security landscape reflects a global paradigm applicable to other countries

Globalization and the knowledge driven technological revolution are important instruments for national economic progress. As a result, increasing innovation has emerged as a critical requirement for an organization's competitiveness and sustainability in both domestic and global markets. Small sized enterprises are crucial to the creation of each nation's gross domestic product in the modern economy. Due to the numerous amounts of small businesses, changes in the business environment have mitigated the structural disadvantages of small-sized businesses (Živanović, Abramović, Živanović, & Smolović, 2023).

This paper aims to examine the current landscape of commercial offerings for intrusion detection provided to small businesses and compare the total cost of ownership against an intrusion detection solution we will build by choosing the proper face recognition algorithm.

The solution is free of charge to small businesses. In this paper, we have a close look at the most popular face recognition methods and compare them against each other. We will choose two algorithms to implement our solution.

In future work, we will conduct a series of tests on the chosen algorithms to check their accuracy and validate that this is the right choice to address the video surveillance security of small businesses.

CURRENT VIDEO SECURITY SOLUTIONS CATERING TO SMALL BUSINESSES

We have conducted market research and selected a list of the most predominant security alarm companies in the USA that cater to small businesses:

- ADT: They provide custom-built security systems with features like 24/7 professional monitoring, intrusion detection, fire protection, and even cybersecurity options. While on the costlier side, they offer a comprehensive security solution. (ADT, 2024)
- SimpliSafe: Known for affordability and DIY installation, SimpliSafe offers business-friendly packages with features like entry sensors, security cameras, and environmental monitoring. (SimpliSafe, 2024)
- Vivint specializes in high-end security systems with features like smart home integration, access control, and video surveillance. It offers professional installation and monitoring.
  (Vivint, 2024)
- Frontpoint: Another DIY option, Frontpoint provides scalable security systems for small businesses. Their features include entry sensors, security cameras, and medical alert options. (Frontpoint, 2024)
- Guardian Alarm: Offers monitored security systems with features like 24/7 video surveillance, intrusion detection, and fire protection. They cater to businesses of all sizes. (Guardian Security, 2024)

Since the paper's objective is to evaluate the total cost of ownership and find out if there is scope for a tailored solution for small businesses, we also looked at the costs associated with each of the solutions mentioned above.

These are estimated costs based on publicly available information and may vary depending on chosen packages, equipment needs, and additional services. For each individual small business, the best way is to get quotes directly from each company for the most accurate pricing for their specific business needs. However, this does not change the validity of our analysis since we will be looking at the starting prices offered to small businesses and not at the more expensive custom-made prices. The following table is a summary based on the market research; to get some of the quotes, we had to use different ZIP codes; otherwise, the information is not possible to obtain from public sources.

Table 1
Summary based on the Market Research

Company	Equipment Cost	Monitoring Cost (Monthly)	Total Yearly Cost
ADT	Starts at ~$600	Starts at ~$37	Starts at ~$1,044
SimpliSafe	Starts at ~$500	Starts at ~$25	Starts at ~$800
Vivint	Starts at ~$1,000	Starts at ~$40	Starts at ~$1,480
Frontpoint	Starts at ~$400	Starts at ~$35	Starts at ~$820
Guardian Alarm	Starts at ~$400	Starts at ~$30	Starts at ~$760

Source: Author own elaboration. (2024)

For the sake of completeness, there are some additional factors to consider when choosing a small business security system, such as:

- Security Needs: Is it physical security of the premises only, or shall it include fire, environmental hazards, etc.?
- Business Size and Layout are important factors in choosing a system that can effectively cover the entire building.
- Scalability, the security needs might grow in the future, and a system that can adapt.
- Integration: whether the system integrates with other business tools you use (e.g., access control systems or productivity tools such as a work calendar).

FACE RECOGNITION ALGORITHMS

Face recognition approaches have evolved over the years. There are two broad categories: traditional methods and deep learning methods. Deep learning methods are specialized versions of the most generic object detection algorithms.

We will first describe the fundamentals of the most popular object detection and face recognition algorithms and then focus on their differences. We will use three criteria for the evaluation: computation power (directly related to cost), accuracy, and speed.

Fundamentals

Before we dive in, let's start with the main concept of face recognition. The steps involved in a typical face recognition system are as follows:

- Data Collection:

  - This step is fundamental since it is the algorithm's baseline. The idea is to acquire a dataset of facial images. This dataset should include images of various individuals under various conditions, such as different lighting, poses, and expressions.
  - The more diverse and representative the dataset, the better chance for the system to generalize to new faces.
- Preprocessing:
  - Normalize the images to ensure consistency in terms of lighting conditions, resolution, and facial expressions. In simple terms, we need to compare apples to apples.
  - Perform face detection and alignment to locate facial features accurately. This step is very important for standardizing the position and orientation of faces in the images.
- Feature Extraction:
  - Extract relevant features from the preprocessed facial images. Different techniques are used, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Convolutional Neural Networks (CNNs).
  - Generate a feature vector that represents the unique characteristics of each face in the dataset.
- Training the Model:
  - Use the extracted features to train a face recognition model. This step can use traditional machine learning algorithms or deep learning neural networks.
  - The model learns to differentiate between different individuals using the unique feature vectors that were generated during the previous phase.
- Dimensionality Reduction:
  - This step is optional and aims to make the algorithm more efficient and less prone to overfitting. The goal is to reduce the complexity of the feature space.
- Face Encoding:
  - Face encoding serves as a unique identifier for each person. For each individual in the dataset, a compact representation of their facial features is generated.
- Testing Phase:
  - There are similarities between the training phase and the testing phase. The preprocessing, feature extraction and dimensionality reduction in the testing phase are accomplished in the same way as in the training phase.
  - The trained model is used to compare the new face encoding with the stored encodings of known individuals.
- Decision Making:
  - This is decision time, so we apply a rule to determine the individual's identity. We look at the similarity between the new face encoding and the stored encodings.
  - In addition to the rule, a similarity score threshold must be used to decide whether the face belongs to a known individual or is an unknown face.
- Verification / Identification:
  - There is a slight difference between verification and identification.

    - In verification scenarios, the decision is binary; the presented face matches a specific individual's encoding.
    - In identification scenarios, the presented face is compared against the dataset of all known encodings to find the closest match.
- Post-Processing:
    - This is the time when we need to close the feedback loop. The goal is to refine the recognition results and/or update the model with new data.

As with any machine learning algorithm, the effectiveness of a face recognition system depends on the quality and diversity of the training data, the choice of features, and the robustness of the algorithm (Osowski & Siwek, 2020).

Eigenfaces

The primary goal of Eigenfaces is to represent facial features in a lower-dimensional space by capturing the principal components of the face images. It aims to preserve as much variance as possible, making it suitable for general facial recognition applications. The Eigenfaces algorithm uses Principal Component Analysis (PCA) to represent facial features in a lower-dimensional space. PCA recognizes the directions of maximum variance in the data, and the corresponding eigenvectors form the basis for the reduced-dimensional space (Peng, Portugal, Alencar, & Cowan, 2021). The representation obtained by Eigenfaces is based on the principal components of the face images. "These components are linear combinations of the original pixel values, capturing the overall variance in the dataset (Omer, 2024). The drawback of PCA is the sensitivity to variations in lighting conditions, as it tends to capture overall variance, which can include variations caused by illumination.

Fisherfaces

Fisherfaces emphasizes the differences between individuals rather than capturing the overall variance in the dataset. It employs Linear Discriminant Analysis (LDA) to find a projection that maximizes individual separation while minimizing variations within the same individual. Fisherfaces use Linear Discriminant Analysis (LDA) for dimensionality reduction. "LDA aims to find a projection that maximizes the ratio of the between-class scatter to the within-class scatter, focusing on the discriminative power of the features". The representation obtained by Fisherfaces is designed to maximize the differences between individuals. The features selected by LDA are most effective in discriminating between different classes (individuals). Fisherfaces can be sensitive to class variations, especially when dealing with small datasets or when the within-class scatter is high. In situations where there is significant variability in the images of the same individual (such as different facial expressions, poses, or lighting conditions), Fisherfaces may not effectively capture the essential features that distinguish between individuals.

Local Binary Patterns (LBP)

Local Binary Patterns (LBP) are commonly used in face recognition as a texture descriptor to capture discriminative features from facial images. The technique is particularly robust to variations in lighting conditions, making it suitable for handling challenges often encountered in face recognition applications.

The basis of LBP is on Local Neighborhood Definition (define a local neighborhood by specifying a radius and the number of sampling points on the circumference of a circle) and LBP Calculation (Create a binary pattern by concatenating the 0/1 values obtained from the comparisons in a specific order around the center pixel).

While LBP is effective in certain scenarios, it may not capture complex spatial relationships or be robust to other variations, such as pose changes. In practice, LBP is often combined with other feature extraction techniques or used as part of a more comprehensive face recognition system to improve overall performance.

Support Vector Machines (SVM)

It's a supervised machine learning algorithm renowned for effectively categorizing data elements into two groups. SVMs come in two primary categories: Linear and Non-Linear SVMs. Linear SVM is applied when data is perfectly separable by a straight line, while Non-Linear SVM handles scenarios where data points aren't linearly separable. In such cases, advanced techniques like kernel tricks are employed to facilitate classification. Since linear separability is rare in real-world applications, kernel tricks are commonly utilized to address these complexities.

SVM works faster and more accurately when the data is linearly separable, and when we employ the kernel trick any complex problem can be solved. However, some issues in face recognition must be considered, such as the difficulty of choosing a good kernel, the inaccuracy for big datasets, and the difficulty of tuning the hyper-parameters.

Deep Learning (CNN-based)

Deep Learning proves immensely valuable in object classification tasks, notably in face recognition algorithms. The initial approach employed was the Convolutional Neural Network (CNN). The essence of this algorithm involved segmenting the image into multiple regions and subsequently classifying each region into distinct categories. However, a challenge arises due to the considerable number of regions necessitating accurate prediction, resulting in prolonged computation times.

Subsequently, the Region-Based Convolutional Neural Network (RCNN) emerged as an advancement over CNN. Built upon the foundation of CNN, RCNN integrates selective search to generate regions, thereby reducing their number to around 2000 per image. Although this marks an improvement over CNN, the computational requirement remains high. Each region is individually passed to the CNN, resulting in processing times of 40 to 50 seconds per image—rendering it unsuitable for real-time applications.

The next algorithm was Fast RCNN. It resolves the main issue of the RCC so that each image (1) is passed only once to the CNN, (2) feature maps are extracted, and (3) selective search is used on these maps to generate predictions (Rajeshkumar et al., 2023). It reduces the time to 2 seconds per image, but the computation time is still high (Jiang et al., 2021).

The next algorithm was Faster RCNN. The main difference to the Fast RCNN is that “the selective search method is replaced with a region proposal network (Faulkner, 2021). This had a dramatic positive impact on the algorithm speed, going down to 0.2 seconds per image.

You Look Only Once (YOLO)

Deep Learning methods such as RCNN, Fast RCNN, and Faster RCNN are categorized as two-stage object detectors. The main concept is that they need several passes to the image. Two additional algorithms, YOLO (You Only Look Once) and SSD (Single Shot Detectors), are categorized as one-stage object detectors, where the image is taken all at once. The other object detection algorithms work on a classification problem, while YOLO works on a single regression problem (Redmon, Divvala, Girshick & Farhadi, 2016). The system only looks at the image once to detect (a) what objects are present and (b) where they are. The system divides the image into a grid of cells. Each grid cell predicts B bounding boxes and confidence scores for these boxes. The confidence score considers the probability that the box contains an object and the accuracy of what box it is predicting (Cortes & Jose, 2021).

Each grid cell predicts C conditional class probabilities, producing a single set of class probabilities per cell irrespective of the number of boxes B. In the testing phase, these conditional class probabilities are combined with individual box confidence predictions. This multiplication yields class-specific confidence scores for each box, indicating both the class's probability and the box's fit to the object (Zhang & Cloutier, 2022). The algorithm uses the concept of IoU to consider a bounding box. By default, the value is set at 0.5, and setting it to higher values will reduce the false positive while increasing the false negatives. The bounding box is not considered when the corresponding IoU is less than the predefined threshold (Pham, Courtrai, Friguet, Lefèvre & Baussard, 2020). Another pillar of the algorithm is non-max suppression. Depending on the specific circumstances, the algorithm may find multiple detections of the same object; however, using non-max suppression, the object will be detected only once. YOLO has been in several iterations, from YOLOv1 to YOLOv7, and each iteration has brought major advancements in features and overall performance. Several key features, such as anchor boxes, intersection over union, and non-max suppression, offer great help in detecting objects of various sizes and classes.

Single Shot Detectors (SSD)

SSD and YOLO exhibit numerous parallels. Both partition the input image into grids, assessing the presence of objects within each cell. However, their primary difference lies in how they tackle bounding box regression. SSD treats each bounding box prediction as a regression problem, commencing with the anchor box possessing the highest IoU and progressively adjusting towards the ground truth bounding box while calculating loss. Conversely, YOLO predicts multiple bounding boxes for each recognized object, employing non-max suppression to eliminate redundant boxes while preserving the final box coordinates.

We have kept the same criteria for evaluating the object detection and face recognition algorithms: total cost of ownership, speed, and accuracy. If we have to compare SSD and YOLO against each other, SSD is somewhat more accurate, most of which occurs because of its ability to recognize things of different sizes. However, YOLO speed is faster than SSD (Sun, 2023). As part of our future work, we will implement both these algorithms in our solution and compare them against each other. We will use available Python libraries for both algorithms and check out both CPU and GPU options (Kumari Sirivarshitha, Sravani, Priya, & Bhavani, 2023).

COMPARISON BETWEEN ALGORITHMS

Deep learning algorithms (CNN-based) generally offer the best accuracy but at the cost of higher computational power (additional steps) and slower processing times. In general, simpler algorithms might be a better choice for speed-critical applications or those with limited resources, but with a trade-off in accuracy. The optimal choice depends on the specific application and its requirements.

The table below summarizes the comparison between today's algorithms used for face recognition (to be fair, there are other dimensions to add to the comparisons, such as data set size and complexity, so we are considering an "average" size dataset):

Table 2

Summarizes the comparison between today's algorithms used for face recognition

Category	Algorithm	Computational Power	Accuracy	Speed
Traditional	Eigenfaces	Low	Moderate	Fast
	Fisherfaces	Low	Moderate	Fast
	Local Binary Patterns (LBP)	Low	Moderate	Moderate
	Support Vector Machines (SVM)	Low	Moderate	Moderate
Deep Learning two stage detection	Deep Learning (CNN-based) ● CNN ● RCNN ● Fast RCNN ● Faster RCNN	High	High	Slow
Deep Learning one stage detection	Single Shot Detectors (SSD)	Moderate	High	Fast
	You Look Only Once (YOLO)	Moderate	High	Fast

Source: Author own elaboration. (2024)

Based on the research results and industry feedback, we will consider using one of the Deep Learning algorithms with one-stage detection since they check all our evaluation criteria. If we stack them against each other (SSD vs YOLO), the SSD algorithm provides slightly better accuracy but at the cost of the speed of detection. Both these algorithms are implemented in Python and provide a practical approach to being utilized in real-life scenarios.

COMPARISON BETWEEN SOLUTIONS (COMMERCIAL OFFERINGS VS OURS)

Now that we have selected the best-suited algorithm to build our solution, we need to compare it against commercial offerings for small businesses. In the “Current Video Security Solutions Catering Small Businesses” chapter, we examined the estimated costs based on publicly available information on the most popular security companies catering to small businesses.

As we mentioned our solution will be free of charge so we need to consider only the initial costs which include the cost of IP cameras and the computer where the software will be running. There are no specific requirements for IP cameras, so generic ones will be considered adequate for the job. The cost of such cameras is about 30 USD per unit. Based on the layout of the building, we believe that a maximum of eight (8) cameras will cover all the spots. Buying the cameras in bulk changes the unit economics to 25 USD per unit. The overall cost of cameras is about 200 USD. The computer running the software does not have any stringent requirements. Furthermore, this computer is not dedicated to intrusion detection only and can be used for other tasks. For better performance, we can use GPU in addition to CPU, but it is not mandatory. We consider an estimated cost of 800 USD for the computer to cover the high-end demand.

The following table can be used to compare the commercial offerings versus the solution we are building:

Table 3

Compare the commercial offerings versus the solution we are building

Company	Equipment Cost	Monitoring Cost (Monthly)	Total Yearly Cost
ADT	Starts at ~$600	Starts at ~$37	Starts at ~$1,044
SimpliSafe	Starts at ~$500	Starts at ~$25	Starts at ~$800
Vivint	Starts at ~$1,000	Starts at ~$40	Starts at ~$1,480
Frontpoint	Starts at ~$400	Starts at ~$35	Starts at ~$820
Guardian Alarm	Starts at ~$400	Starts at ~$30	Starts at ~$760
Our solution	Starts at ~$1,000	Free of charge	Free of charge

Source: Author own elaboration. (2024)

While there is an initial investment in hardware (IP cameras and the computer), the solution, which is free of charge, can help small businesses address physical security via intrusion detection.

CONCLUSIONS

Our initial goal was to find two face recognition algorithms best suited for intrusion detection applications. We evaluated three criteria: total cost of ownership, accuracy, and speed. Based on the research results and industry feedback, we will consider using one of the Deep Learning algorithms for our next research step. Having selected our algorithms (Deep Learning with one-stage detection), we critically examined whether this solution outperformed existing commercial offerings, thereby justifying its use in our research.

We believe that our solution addresses the need for intrusion detection. Offering it free of charge and having small businesses only purchase the hardware is a valid option for small businesses to address physical security.From an implementation perspective, we will use standard IP cameras (inexpensive) connected to a computer (either CPU or GPU), several Linux utilities to convert the video stream to images (we will configure the sampling rate), and then analyze the images. In future work, we will conduct a series of tests on the chosen algorithms (SSD and YOLO) to check their accuracy and validate that they are their accuracy and validate that they are the right choice to address the video surveillance security of small businesses.

The goal is to exceed 99% accuracy if the person is captured on camera for more than one minute (non-contiguous). We will tune the system to find the ideal sampling video rate and the number of cameras covering an indoor space. Based on that, we will calculate the platform's cost and compare it to the current offerings for smaller businesses.

REFERENCES

1. Cortes, V., & Jose, C. (2021). Joint Neural Networks for One-shot Object Recognition and Detection.
2. Faulkner, H. (2021). General and Fine-Grained Video Understanding using Machine Learning & Standardised Neural Network Architectures.
3. Jiang, L., Chen, J., Todo, H., Tang, Z., Liu, S., & Li, Y. (2021). *Application of a Fast RCNN Based on Upper and Lower Layers in Face Recognition* (Vol. 2021).
4. Kumari Sirivarshitha, A., Sravani, K., Priya, K., & Bhavani, V. (2023). An approach for Face Detection and Face Recognition using OpenCV and Face Recognition Libraries in Python. *2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS 2023*.
5. Luca, A.-V., Simon-Várhelyi, M., Mihály, N.-B., & Cristea, V.-M. (2023). *Fault type diagnosis of the wwtp dissolved oxygen sensor based on fisher discriminant analysis and assessment of associated environmental and economic impact* (Vol. 13). MDPI.
6. Omer, F. M. (2024). Application of Principal Component Analysis to Vehicle Sales: Case Study at IBM Company. . *Journal of Humanity Sciences.* .
7. Osowski, S., & Siwek, K. (2020). *CNN application in face recognition* (Vol. 96).
8. Peng, P., Portugal, I., Alencar, P., & Cowan, D. (2021). A face recognition software framework based on principal component analysis. *PLoS ONE, 16*(7 July).
9. Pham, M., Courtrai, L., Friguet, C., Lefèvre, S., & Baussard, A. (2020). *YOLO-fine: One-stage detector of small objects under various backgrounds in remote sensing images* (Vol. 12).
10. Rajeshkumar, G., Braveen, M., Venkatesh, R., Josephin Shermila, P., Ganesh Prabu, B., Veerasamy, B., Jeyam, A. (2023). *Smart office automation via faster R-CNN based face recognition and internet of things* (Vol. 27).
11. Rajeshkumar, G., Braveen, M., Venkatesh, R., Shermila, P., Prabu, B., Veerasamy, B., . . . Jeyam, A. (2023). *Smart office automation via faster R-CNN based face recognition and internet of things* (Vol. 27). Elsevier.
12. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection.*
13. Sun, F. (2023). *Face Recognition Analysis Based on the YOLO Algorithm* (Vol. 2).
14. Zhang, H., & Cloutier, R. (2022). *Review on One-Stage Object Detection Based on Deep Learning* (Vol. 7).
15. Živanović, S., Abramović, N., Živanović, M., & Smolović, S. (2023, 4). *INNOVATION MANAGEMENT ON THE WAY TO BUSINESS EXCELLENCE* (Vol. 17).
16. Business protection from a leader in business security, https://www.adt.com/business accessed 04/04/2024
17. The SimpliSafe Difference, https://www.adt.com/business accessed 18/03/2024
18. Security cameras designed to deter real threats, https://www.vivint.com/ppc/cameras accessed 26/02/2024
19. Video surveillance, https://www.myshield.org/video-surveillance accessed 20/02/2024