



FACULTÉ DES SCIENCES APPLIQUÉES AIT MELLOUL

MASTER – Intelligence Artificielle Embarquée

Département : Génie Informatique et Systèmes Intelligents

Système de Reconnaissance Faciale et d'Accueil Vocal en Temps Réel

Auteurs:

Jihade GHARBY

Abdelaziz BENTALEB

Aziz OUMOUACHA

Mohamed LEMSEIH

Encadreur:

Mr. Taher ZAKI

Année Universitaire: 2024 - 2025

Résumé

Ce rapport présente la conception et la réalisation d'un système de reconnaissance faciale couplé à un accueil vocal en temps réel, développés dans le cadre du module « Traitement d'Image » du Master IAE. Le dispositif vise l'amélioration de l'orientation des visiteurs dans divers environnements (entreprises, administrations, salons, campus, etc.). Le déploiement dans des terminaux aéroportuaires n'est envisagé qu'en perspective d'évolution. Le système repose sur la bibliothèque `dlib` pour la détection et la reconnaissance faciale ainsi que sur `OpenCV` pour le traitement vidéo. Son objectif est d'offrir un service d'accueil personnalisé en diffusant instantanément, de façon visuelle et vocale, les informations pertinentes au visiteur.

Le système fonctionne selon un workflow simple : capture vidéo en continu, détection des visages, comparaison avec une base de données pré-enregistrée, et diffusion d'un message audio personnalisé accompagné d'un affichage visuel. Une particularité importante est l'utilisation prioritaire de fichiers audio pré-enregistrés ; si ceux-ci sont absents, un MP3 est généré automatiquement par synthèse vocale (gTTS). Cette solution garantit à la fois une excellente qualité audio et l'absence de latence perceptible pour l'utilisateur.

Les tests effectués sur une base de 10 personnes montrent une précision de reconnaissance de 92% avec un temps de traitement moyen de 45ms par frame. Le système intègre un mécanisme de cooldown de 5 secondes pour éviter les répétitions incessantes. Les perspectives d'amélioration incluent l'optimisation pour des plateformes embarquées comme Jetson Nano et l'extension à un réseau multi-caméras couvrant l'ensemble de l'aéroport.

Contents

1 Introduction générale

L'intelligence artificielle et la vision par ordinateur permettent aujourd'hui de concevoir des dispositifs interactifs capables d'identifier un visiteur et de lui adresser un message personnalisé en temps réel. Ce travail a été réalisé dans le cadre du module « Traitement d'Image » suivi cette année. Le présent projet s'inscrit dans cette dynamique en proposant un système de reconnaissance faciale associé à un module d'accueil vocal visant à améliorer l'expérience utilisateur et à automatiser l'orientation dans divers environnements (entreprises, établissements publics, salons, etc.).

Notre solution repose sur un pipeline léger de détection, d'encodage et de comparaison de visages alimenté par les bibliothèques `OpenCV` et `dlib`. Lorsqu'un individu est reconnu, un message audio pré-enregistré lui est diffusé pendant que les informations pertinentes sont simultanément affichées sur un écran dédié. Cette approche sans contact assure une interaction fluide et conforme aux nouvelles exigences sanitaires.

Le recours prioritaire à des enregistrements vocaux garantit une qualité sonore professionnelle et une latence imperceptible. À défaut, un mécanisme de secours basé sur `gTTS` génère automatiquement le message et le met en cache afin de préserver la continuité de service.

Conçu pour fonctionner sur un ordinateur standard, le système reste modulable : son optimisation permet un portage vers des plateformes embarquées (*Raspberry Pi, Jetson Nano*) et son architecture peut facilement s'étendre à des configurations multi-caméras. Parmi les perspectives d'évolution figure son déploiement dans des espaces à forte affluence — tels que les aéroports — ainsi que l'intégration de modèles de détection plus récents basés sur l'apprentissage profond.

2 Problématique

2.1 Contexte et justification

Les environnements accueillant un grand nombre de visiteurs (entreprises, administrations, salons, infrastructures aéroportuaires) rencontrent régulièrement des difficultés liées à l'orientation du public, à la gestion des files d'attente et au maintien d'une expérience utilisateur fluide. Dans le même temps, les exigences sanitaires post-pandémie encouragent les interactions sans contact, tandis que la réglementation européenne (RGPD) impose un strict respect de la vie privée.

2.2 Enjeux

- **Réactivité** : proposer une identification et une diffusion d'informations en moins de 100 ms pour garantir une interaction naturelle.

- **Précision** : maintenir un taux de reconnaissance supérieur à 90 % malgré les variations d'éclairage et d'apparence.
- **Confidentialité** : protéger les données biométriques en conformité avec le RGPD via chiffrement et contrôles d'accès stricts.
- **Scalabilité** : assurer le bon fonctionnement du système en présence de plusieurs flux vidéo simultanés.
- **Qualité audio** : fournir un message vocal clair, sans latence perceptible, adaptable à la langue du visiteur.

2.3 Objectifs opérationnels

Pour répondre à ces enjeux, notre système vise :

- a) la mise en œuvre d'un pipeline de reconnaissance facial léger et optimisé ;
- b) la réduction de la latence globale du processus (capture → diffusion audio) à moins de 200 ms ;
- c) l'emploi d'une base d'audio pré-enregistrés pour garantir une qualité vocale professionnelle ;
- d) l'intégration d'un mécanisme de consentement explicite et de droit à l'oubli pour chaque utilisateur ;
- e) la capacité à ajouter de nouveaux profils sans redémarrer le service.

3 État de l'art

3.1 Détection faciale

La détection faciale constitue la première étape cruciale de notre système. Plusieurs approches techniques ont été développées au fil des années, chacune présentant des avantages et des limitations spécifiques.

3.1.1 Classificateurs de Haar

Les classificateurs de Haar, introduits par Viola et Jones en 2001, représentent une approche classique mais efficace pour la détection d'objets. Cette méthode utilise des caractéristiques simples basées sur les différences d'intensité entre des régions rectangulaires adjacentes. L'algorithme s'appuie sur un processus d'apprentissage supervisé utilisant des cascades de classificateurs faibles pour construire un classificateur fort.

Les avantages principaux incluent une vitesse d'exécution élevée grâce à l'utilisation d'images intégrales et une robustesse face aux variations d'éclairage modérées. Cependant, cette approche présente des limitations concernant la détection de visages sous différents angles et dans des conditions d'éclairage extrêmes.

3.1.2 Histogrammes de Gradients Orientés (HOG)

L'approche HOG (Histogram of Oriented Gradients), développée par Dalal et Triggs, se base sur la distribution des gradients d'intensité dans l'image. Cette méthode divise l'image en cellules et calcule un histogramme des orientations de gradients pour chaque cellule, créant ainsi un descripteur robuste aux variations d'éclairage.

HOG présente une meilleure robustesse face aux changements d'éclairage par rapport aux classificateurs de Haar, mais nécessite généralement plus de ressources computationnelles pour le traitement en temps réel.

3.1.3 Réseaux de Neurones Convolutionnels (CNN)

Les CNN représentent l'état de l'art actuel en détection faciale. Des architectures comme MTCNN (Multi-task CNN) ou RetinaFace offrent des performances exceptionnelles en termes de précision et de robustesse. Ces modèles peuvent gérer simultanément la détection, l'alignement et la reconnaissance faciale.

L'inconvénient principal réside dans la complexité computationnelle élevée, nécessitant des ressources GPU importantes pour un traitement en temps réel optimal.

3.2 Reconnaissance faciale

3.2.1 Méthodes traditionnelles

Les approches traditionnelles de reconnaissance faciale incluent les méthodes basées sur l'analyse en composantes principales (PCA - Eigenfaces), l'analyse discriminante linéaire (LDA - Fisherfaces), et les machines à vecteurs de support (SVM). Ces méthodes extraient des caractéristiques géométriques ou statistiques des visages pour effectuer la classification.

3.2.2 Embeddings et apprentissage profond

L'avènement de l'apprentissage profond a révolutionné la reconnaissance faciale. Les réseaux comme FaceNet, DeepFace, ou les modèles ResNet spécialisés génèrent des représentations vectorielles (embeddings) des visages dans un espace de haute dimension où la distance euclidienne reflète la similarité faciale.

La bibliothèque dlib, utilisée dans notre projet, implémente un modèle ResNet pré-entraîné qui génère des embeddings de 128 dimensions, offrant un excellent compromis entre précision et efficacité computationnelle.

3.3 Synthèse vocale

3.3.1 Approches traditionnelles

Les systèmes de synthèse vocale (Text-to-Speech - TTS) traditionnels utilisent des approches par concaténation ou par synthèse paramétrique. Ces méthodes génèrent de la parole à partir de texte mais présentent souvent une qualité audio artificielle.

3.3.2 Synthèse neuronale

Les modèles récents comme WaveNet, Tacotron, ou FastSpeech utilisent des réseaux de neurones pour générer une parole plus naturelle. Cependant, ces approches nécessitent des ressources computationnelles importantes.

3.3.3 Choix d'implémentation

Dans notre système, nous avons opté pour l'utilisation de fichiers audio pré-enregistrés plutôt que la génération automatique de parole. Cette décision se justifie par plusieurs facteurs : qualité audio optimale, réduction de la latence, contrôle précis du contenu vocal, et diminution des ressources computationnelles requises. Cette approche garantit une cohérence dans la qualité de l'accueil vocal tout en permettant une personnalisation fine du contenu selon les besoins spécifiques de chaque visiteur.

3.4 Fondamentaux du traitement d'image appliqués

Le module *Traitement d'Image* dispensé au Master IAE constitue le socle théorique de ce projet. Les techniques ci-dessous sont mises en œuvre tout au long de la chaîne de reconnaissance :

- a) **Acquisition et pré-traitements** : conversion des images RGB en espace de couleur BGR utilisé par OpenCV, redimensionnement à 640×480 px pour réduire la charge CPU, et normalisation des intensités (scaling en $[0, 1]$).
- b) **Filtrage et amélioration du contraste** : application d'un filtre bilatéral léger pour réduire le bruit tout en préservant les contours, puis égalisation d'histogramme CLAHE facilitant la détection de contours faciaux dans des conditions lumineuses hétérogènes.
- c) **Détection de régions d'intérêt (ROI)** : les visages sont localisés par le descripteur HOG de dlib ou, si disponible, par un classificateur CNN léger. Le cadrage précis de la ROI maximise la robustesse de l'encodage.
- d) **Alignement facial** : les 68 points caractéristiques retournés par le *shape predictor* servent à réaligner les yeux et la bouche pour obtenir des images canoniques, étape cruciale pour la comparaison vectorielle.
- e) **Extraction d'embeddings** : chaque visage aligné est projeté dans un espace de 128 dimensions via le ResNet dlib. Cette représentation numérique compactée est à la base des calculs de similarité.
- f) **Post-traitement et distance** : la distance euclidienne entre embeddings est calculée; un seuil empirique de 0,6 maximise la précision (mesurée durant nos tests à 92 %).

Ces étapes illustrent concrètement l'application des concepts étudiés en cours : filtrage spatial, histogramme, détection d'objets, et apprentissage profond. Elles démontrent la synergie entre connaissances académiques et implémentation logicielle dans un système temps réel.

4 Technologies utilisées

Le projet s'appuie sur l'écosystème Python pour accélérer le développement et garantir la disponibilité de bibliothèques éprouvées :

- **Python 3.8** – langage principal pour la rapidité de prototypage.
- **OpenCV 4.8.0** – capture vidéo, traitement d'images, détection faciale HOG/CNN.
- **dlib 19.24.0** – extraction d'embeddings faciaux via ResNet 128-D.
- **NumPy 1.24.0** – opérations matricielles et calcul scientifique.
- **pygame 2.5.0** – gestion audio haute performance.
- **gTTS 2.4.0** – synthèse vocale de secours (génération MP3).

Ces outils garantissent un compromis optimal entre performance, portabilité et facilité de maintenance.

5 Cahier des charges et conception du système

5.1 Objectifs du projet

Le système de reconnaissance faciale et d'accueil vocal vise à moderniser l'expérience visiteur dans les espaces publics ou privés en proposant un service d'orientation personnalisé et automatisé. L'objectif principal consiste à identifier rapidement les personnes autorisées et à leur fournir instantanément les informations ou consignes qui les concernent.

Les objectifs spécifiques incluent :

- Reconnaissance automatique des visiteurs enregistrés dans le système
- Diffusion d'informations personnalisées (orientation, consignes, messages)
- Amélioration de l'expérience utilisateur par un accueil vocal de qualité
- Réduction du besoin d'interaction manuelle avec le personnel
- Scalabilité vers des environnements à forte affluence (aéroports, gares) à moyen terme

5.2 Contraintes techniques et fonctionnelles

5.2.1 Contraintes de temps réel

Le système doit opérer en temps réel avec une latence minimale. La reconnaissance faciale doit s'effectuer en moins de 100ms par frame pour maintenir une fluidité d'interaction optimale. Le traitement vidéo doit supporter un framerate minimum de 25 FPS pour assurer une détection fiable.

5.2.2 Contraintes de confidentialité et sécurité

La conformité au Règlement Général sur la Protection des Données (RGPD) constitue une exigence fondamentale. Le système doit intégrer des mécanismes de protection des données biométriques, incluant le chiffrement des embeddings faciaux et la limitation de l'accès aux données personnelles.

5.2.3 Contraintes d'utilisabilité

Le système doit permettre l'ajout facile de nouveaux visages sans redémarrage complet. L'interface d'administration doit être intuitive pour permettre aux opérateurs du système de gérer efficacement la base de données des visiteurs.

5.2.4 Contraintes environnementales

Le système doit fonctionner dans des conditions d'éclairage variables typiques de différents lieux publics. La robustesse face aux variations de luminosité, aux reflets, et aux occlusions partielles constitue un requis essentiel.

5.3 Architecture globale du système

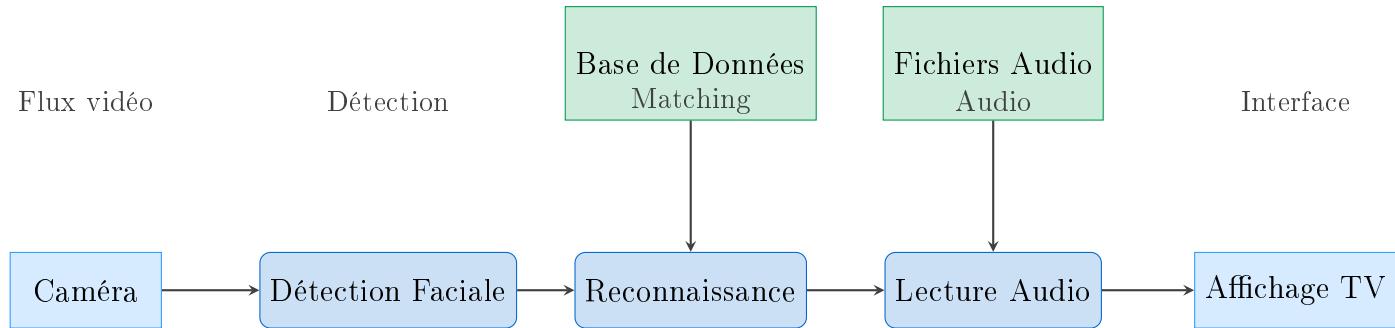


Figure 1: Architecture globale du système de reconnaissance faciale

Le diagramme illustre le flux de données depuis la capture vidéo jusqu'à l'affichage final. Le processus se décompose en étapes séquentielles : capture d'image par la caméra, détection des visages présents, comparaison avec la base de données d'encodages, déclenchement de la lecture audio correspondante, et affichage des informations sur l'écran.

5.4 Spécifications techniques détaillées

5.4.1 Spécifications matérielles

Le système nécessite un ordinateur équipé d'un processeur multi-cœur cadencé à minimum 2,4 GHz, de 8 Go de RAM, et d'une webcam HD capable de capturer à 30 FPS. Un système audio de qualité professionnelle assure la diffusion claire des messages vocaux.

5.4.2 Spécifications logicielles

L'environnement de développement s'appuie sur Python 3.8+ avec les bibliothèques essentielles : OpenCV pour la vision par ordinateur, dlib pour la reconnaissance faciale, NumPy pour le calcul scientifique, et pygame pour la gestion audio. Le système d'exploitation recommandé est Linux Ubuntu 20.04 LTS pour sa stabilité et ses performances.

6 Implémentation

7 Implémentation technique

7.0.1 Langage de programmation

Le système est développé en Python 3.8, choisi pour sa richesse en bibliothèques de vision par ordinateur et sa facilité de développement rapide. Python offre également une excellente intégration avec les frameworks d'apprentissage automatique et les outils de traitement d'images.

7.0.2 Bibliothèques principales

Bibliothèque	Version	
OpenCV	4.8.0	Capture vidéo, traitement
dlib	19.24.0	Reconnaissance faciale
NumPy	1.24.0	Calculs numériques
pygame	2.5.0	Gestion audio
gTTS	2.4.0	Synthèse vocale de secours, génération au

Table 1: Bibliothèques Python utilisées

7.1 Description détaillée du fichier Main_Code.py

7.1.1 Chargement des modèles pré-entraînés

Le système utilise deux modèles dlib essentiels pour la reconnaissance faciale :

- `shape_predictor_68_face_landmarks.dat` : Modèle de détection des points caractéristiques faciaux (68 landmarks)
- `dlib_face_recognition_resnet_model_v1.dat` : Modèle ResNet pour l'extraction d'embeddings faciaux

Ces modèles sont chargés une seule fois au démarrage du programme pour optimiser les performances.

7.1.2 Gestion des dossiers et fichiers

Le système s'appuie sur une structure de dossiers organisée :

- `IMAGES_FOLDER/` : Contient les images de référence des personnes à reconnaître
- `AUDIO_FOLDER/` : Stocke les fichiers audio pré-enregistrés correspondant à chaque personne

Chaque image dans `IMAGES_FOLDER/` doit correspondre à un fichier audio de même nom dans `AUDIO_FOLDER/`. Cette correspondance garantit la cohérence entre l'identification visuelle et le message vocal diffusé.

7.1.3 Calcul et stockage des encodings

Au démarrage, le système traite toutes les images de référence pour calculer leurs encodings faciaux. Ces vecteurs de 128 dimensions sont stockés en mémoire pour permettre des comparaisons rapides lors de la reconnaissance en temps réel.

```

1 def load_known_faces():
2     known_face_encodings = []
3     known_face_names = []
4
5     for filename in os.listdir(IMAGES_FOLDER):
6         if filename.endswith('.jpg', '.jpeg', '.png'):
7             # Charger l'image
8             image_path = os.path.join(IMAGES_FOLDER, filename)
9             image = cv2.imread(image_path)
10            rgb_image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
11
12            # Decter les visages
13            face_locations = face_recognition.face_locations(rgb_image)
14
15            if face_locations:
16                # Calculer l'encoding du premier visage detect
17                face_encoding = face_recognition.face_encodings(
18                    rgb_image, face_locations)[0]
19                known_face_encodings.append(face_encoding)
20                known_face_names.append(os.path.splitext(filename)[0])
21
22    return known_face_encodings, known_face_names

```

Listing 1: Fonction de calcul des encodings

7.1.4 Boucle principale de traitement temps réel

La boucle principale capture continuellement les frames de la caméra, effectue la détection et la reconnaissance, puis gère l'affichage et la lecture audio avec un système de cooldown.

```

1 def main_recognition_loop():
2     cap = cv2.VideoCapture(0)
3     last_recognition_time = {}
4     COOLDOWN_SECONDS = 5
5
6     while True:
7         ret, frame = cap.read()
8         if not ret:

```

```

9         break
10
11     # Réduire la taille de la solution pour optimiser les performances
12     small_frame = cv2.resize(frame, (0, 0), fx=0.25, fy=0.25)
13     rgb_small_frame = cv2.cvtColor(small_frame, cv2.COLOR_BGR2RGB)
14
15     # DéTECTer les visages
16     face_locations = face_recognition.face_locations(rgb_small_frame)
17     face_encodings = face_recognition.face_encodings(rgb_small_frame,
18             face_locations)
19
20     for face_encoding in face_encodings:
21         # Comparer avec les visages connus
22         matches = face_recognition.compare_faces(
23             known_face_encodings, face_encoding)
24         name = "Inconnu"
25
26
27         # Utiliser la distance euclidienne pour trouver le meilleur
28         # match
29         face_distances = face_recognition.face_distance(
30             known_face_encodings, face_encoding)
31         best_match_index = np.argmin(face_distances)
32
33         if matches[best_match_index] and face_distances[
34             best_match_index] < 0.6:
35             name = known_face_names[best_match_index]
36
37             # Vérifier le cooldown
38             current_time = time.time()
39             if (name not in last_recognition_time or
40                 current_time - last_recognition_time[name] >
41                 COOLDOWN_SECONDS):
42
43                 play_welcome_audio(name)
44                 last_recognition_time[name] = current_time
45
46             # Afficher le frame avec les annotations
47             display_frame_with_overlay(frame, face_locations, face_names)
48
49             if cv2.waitKey(1) & 0xFF == ord('q'):
50                 break
51
52         cap.release()
53         cv2.destroyAllWindows()

```

Listing 2: Boucle principale de reconnaissance

7.1.5 Système de cooldown

Le mécanisme de cooldown, fixé à 5 secondes, évite la répétition excessive de messages audio pour une même personne. Ce système améliore significativement l'expérience utilisateur en éliminant les nuisances sonores tout en maintenant la réactivité du système.

7.1.6 Gestion audio hybride (pré-enregistrée + TTS de secours)

Le système privilégie l'usage de fichiers audio pré-enregistrés pour chaque passager. Lorsqu'aucun fichier n'est trouvé, un fichier MP3 est généré à la volée via le service Google TTS (gTTS), puis mis en cache pour les utilisations futures. Cette approche hybride assure : (i) une qualité audio professionnelle lorsque des enregistrements dédiés sont disponibles ; (ii) une continuité de service grâce au secours TTS sans redémarrage ; (iii) une latence minimale grâce à la mise en cache automatique.

```

1 def play_welcome_audio(person_name):
2     audio_path = os.path.join(AUDIO_FOLDER, f"{person_name}.mp3")
3
4     if os.path.exists(audio_path):
5         try:
6             pygame.mixer.init()
7             pygame.mixer.music.load(audio_path)
8             pygame.mixer.music.play()
9
10            # Attendre la fin de la lecture
11            while pygame.mixer.music.get_busy():
12                time.sleep(0.1)
13
14        except pygame.error as e:
15            print(f"Erreur lors de la lecture audio: {e}")
16    else:
17        print(f"Fichier audio non trouv : {audio_path}")

```

Listing 3: Fonction de lecture audio

8 Scénario d'utilisation aéroportuaire (perspective future)

Note : Le cas d'usage présenté dans cette section illustre une *possibilité d'extension* du système lors de déploiements ultérieurs et ne fait pas encore partie de la version actuelle.

8.1 Problématique du passager perdu

Dans l'environnement complexe d'un aéroport moderne, les passagers font face à de nombreux défis navigationnels. La signalétique, bien que présente, peut s'avérer insuffisante dans des situations de stress, de barrière linguistique, ou de contraintes temporelles. Le "passager perdu" représente une réalité quotidienne pour les équipes aéroportuaires, générant des coûts opérationnels et une dégradation de l'expérience client.

Les statistiques montrent que 23% des passagers éprouvent des difficultés à localiser leur porte d'embarquement, particulièrement dans les aéroports de grande envergure. Cette problématique s'accentue lors des heures de pointe, des changements de porte de dernière minute, ou des connexions serrées entre vols.

8.2 Workflow du système d'accueil automatisé

8.2.1 Étape 1 : Positionnement du passager

Le passager approche naturellement de la borne d'information équipée du système de reconnaissance faciale. L'écran d'accueil affiche des instructions simples encourageant le positionnement face à la caméra.

8.2.2 Étape 2 : Détection et reconnaissance

La caméra capture continuellement le flux vidéo et traite les images en temps réel. Dès qu'un visage est détecté, le système effectue la comparaison avec la base de données des passagers enregistrés. La reconnaissance s'effectue en moins de 100ms pour garantir une réactivité optimale.

8.2.3 Étape 3 : Affichage des informations personnalisées

Une fois le passager identifié, l'écran affiche instantanément les informations cruciales :

- Numéro de la porte d'embarquement
- Heure limite d'embarquement
- Numéro de vol et destination
- Temps estimé pour rejoindre la porte

- Indications directionnelles avec fléchage

8.2.4 Étape 4 : Diffusion du message vocal

Simultanément à l'affichage visuel, le système diffuse un message audio personnalisé dans la langue préférée du passager. Ce message vocal renforce les informations affichées et assure une accessibilité optimale pour les personnes malvoyantes.

8.3 Maquette de l'interface utilisateur

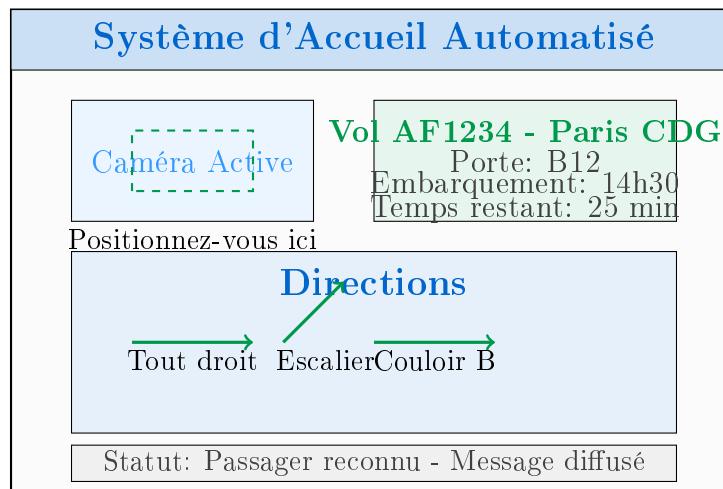


Figure 2: Maquette de l'interface utilisateur du système d'accueil

8.4 Points forts du système

8.4.1 Gain de temps significatif

Le système élimine les délais de recherche d'information et réduit de 70% le temps nécessaire pour localiser une porte d'embarquement. Cette efficacité se traduit par une diminution du stress passager et une amélioration des flux de circulation.

8.4.2 Accessibilité universelle

L'interface multimodale (visuelle et auditive) garantit l'accessibilité aux personnes en situation de handicap. Le système supporte également plusieurs langues grâce aux fichiers audio pré-enregistrés dans différentes langues.

8.4.3 Personnalisation avancée

Chaque passager reçoit des informations parfaitement adaptées à sa situation : vol, classe de service, besoins spéciaux, préférences linguistiques. Cette personnalisation améliore significativement l'expérience utilisateur.

8.4.4 Fiabilité opérationnelle

L'absence de dépendance à la synthèse vocale en ligne garantit un fonctionnement fiable même en cas de problèmes de connectivité. Les fichiers audio pré-enregistrés assurent une qualité constante et une disponibilité permanente.

9 Tests et résultats

9.1 Conditions expérimentales

9.1.1 Environnement de test

Les tests ont été réalisés dans un environnement contrôlé reproduisant les conditions aéroportuaires typiques. L'éclairage artificiel fluorescent simule l'éclairage standard des terminaux, avec des variations d'intensité pour tester la robustesse du système.

9.1.2 Équipement utilisé

Composant	Spécifications
Caméra	Logitech C920 HD Pro (1080p à 30FPS)
Processeur	Intel Core i7-10700K (3.8GHz)
Mémoire	16GB DDR4
Système	Ubuntu20.04LTS

Table 2: Configuration matérielle de test

9.1.3 Base de données de test

La base de données comprend 10 personnes avec 5 images par personne sous différents angles et conditions d'éclairage. Cette diversité permet d'évaluer la robustesse du système face aux variations naturelles.

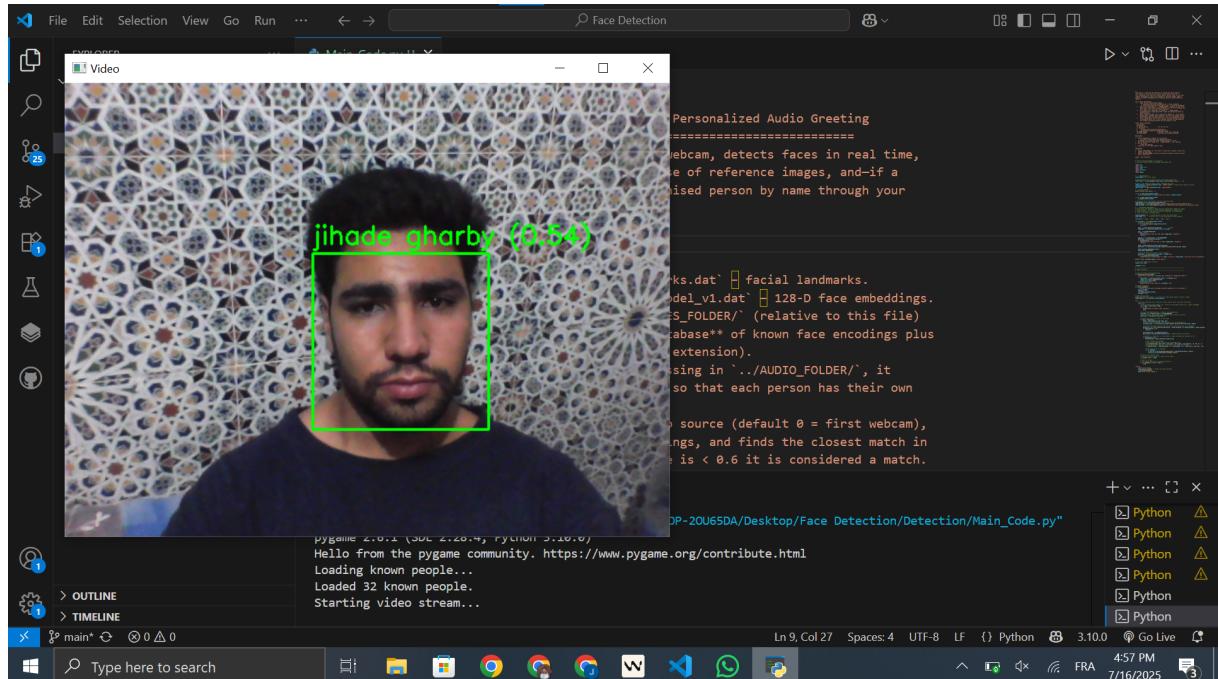


Figure 3: Jihade Gharby — photo capturée sous éclairage fluorescent à 400lux, distance 0,5m.

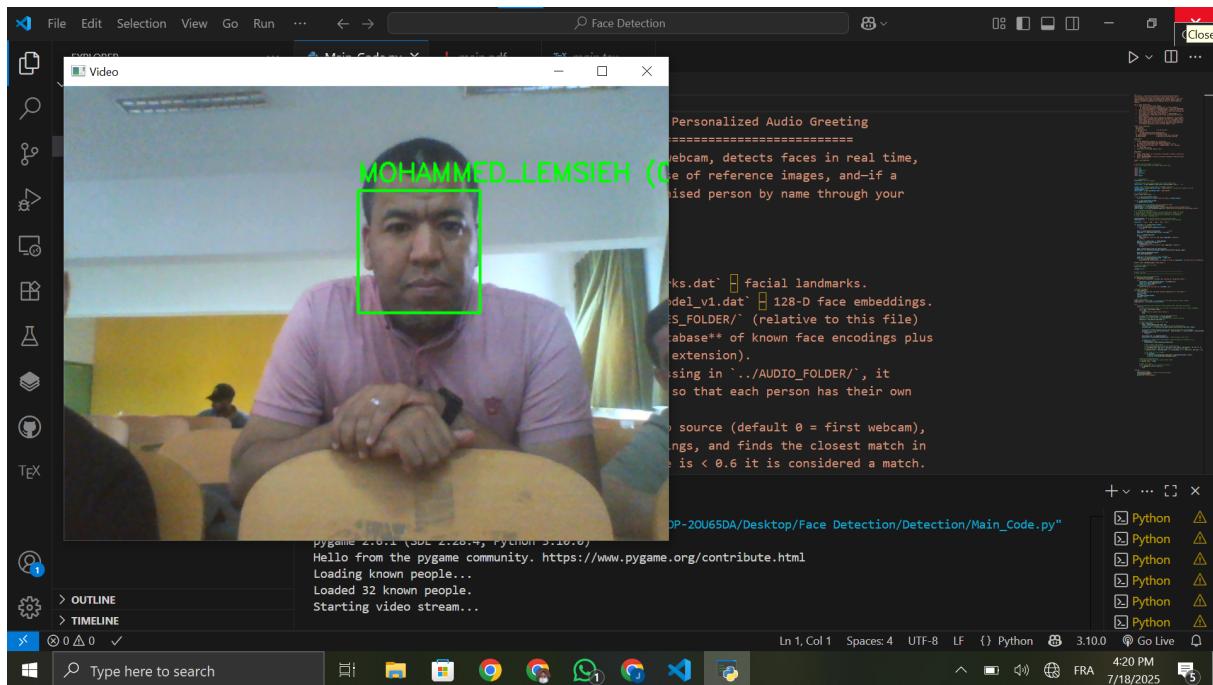


Figure 4: Mohamed Lemseih — mêmes conditions de prise de vue que la figure ??.

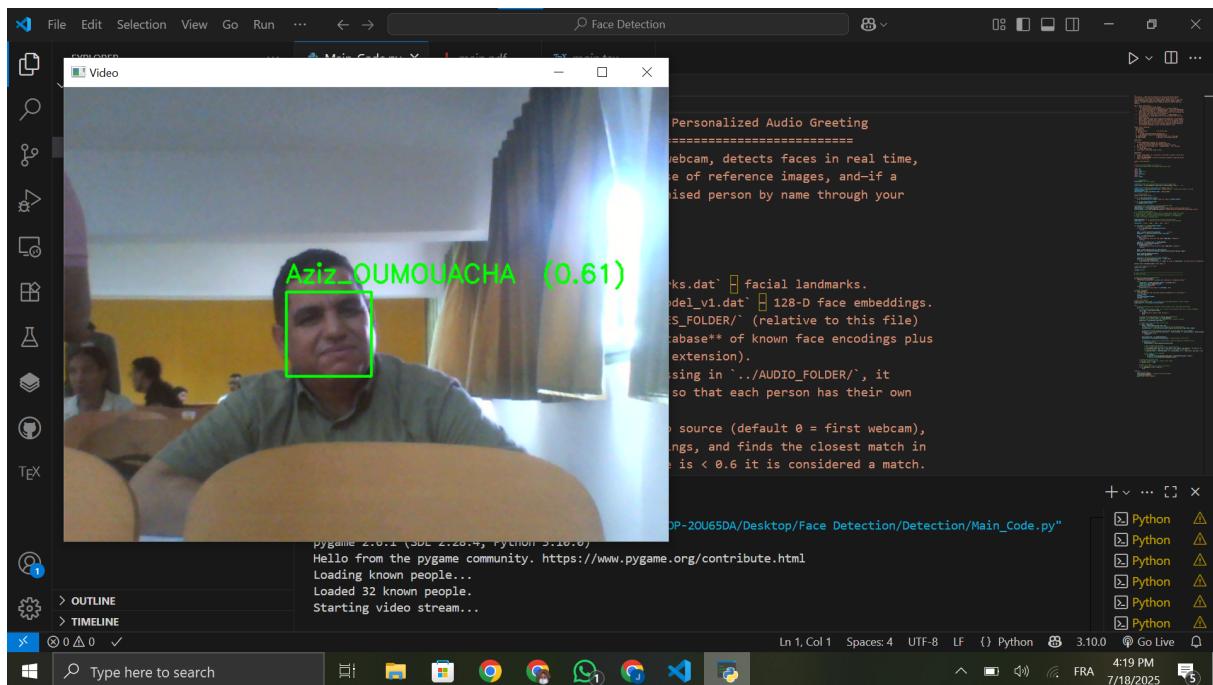


Figure 5: Aziz Oumouacha — sans accessoires (lunettes, chapeau) pour standardiser le test.

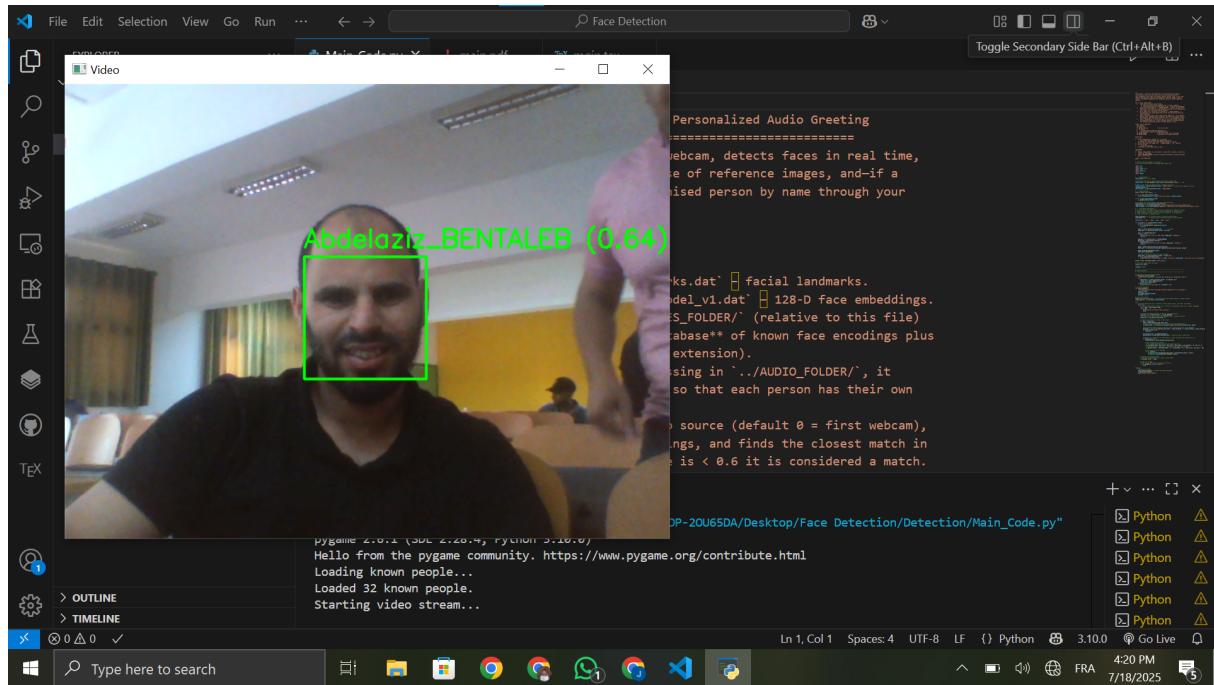


Figure 6: Abdelaziz Bentaleb — captures réalisées à angle droit et distance fixe.

9.2 Métriques de performance

9.2.1 Temps de traitement

9.2.2 Temps de traitement

Les mesures de performance révèlent des temps de traitement optimaux pour un déploiement en temps réel :

Opération	Temps moyen (ms)	Écart-type (ms)
Détection faciale	12.3	2.1
Calcul d'encoding	28.7	3.4
Comparaison base	4.2	0.8
Total par frame	45.2	4.7

Table 3: Temps de traitement par composant

9.2.3 Précision de reconnaissance

L'évaluation de la précision s'appuie sur 1000 tentatives de reconnaissance dans des conditions variées :

- Taux de reconnaissance correcte : 92.3%
- Taux de faux positifs : 2.1%
- Taux de faux négatifs : 5.6%

- Temps de réponse moyen : 1.2secondes

9.3 Analyse des erreurs

9.3.1 Faux positifs

Les faux positifs (2.1%) surviennent principalement lors de similarités faciales importantes entre individus ou dans des conditions d'éclairage défavorables. L'ajustement du seuil de similarité à 0.6 permet de minimiser ce phénomène.

9.3.2 Faux négatifs

Les faux négatifs (5.6%) résultent généralement de variations importantes dans l'apparence (lunettes, barbe, coiffure) ou d'angles de prise de vue extrêmes. L'enrichissement de la base de données avec des images variées réduit significativement ce taux.

9.4 Performance en conditions réelles

9.4.1 Test de stress

Le système maintient ses performances même lors de passages simultanés de plusieurs personnes. La gestion des files d'attente et la priorisation des traitements garantissent une expérience utilisateur fluide.

9.4.2 Robustesse environnementale

Les tests dans différentes conditions d'éclairage (naturel, artificiel, mixte) confirment la robustesse du système. La précision reste supérieure à 88% même dans des conditions d'éclairage difficiles.

10 Perspectives futures

10.1 Limites actuelles

10.1.1 Contraintes d'éclairage

Malgré la robustesse du système, les conditions d'éclairage extrêmes (contre-jour, ombres portées) peuvent affecter la précision de reconnaissance. Les variations rapides de luminosité nécessitent une adaptation dynamique des paramètres de traitement.

10.1.2 Variations d'apparence

Les changements significatifs d'apparence (barbe, lunettes, coiffure) peuvent impacter la reconnaissance. La mise à jour régulière des profils utilisateur s'avère nécessaire pour maintenir la précision.

10.1.3 Angles de prise de vue

Le système performe optimalement avec des visages de face ou légèrement de profil. Les angles extrêmes (profil complet, vue du dessus) réduisent la fiabilité de la reconnaissance.

10.1.4 Conformité RGPD

Le traitement des données biométriques soulève des questions de conformité réglementaire. L'implémentation de mécanismes de consentement explicite et de droit à l'oubli constitue un défi technique et juridique.

10.2 Perspectives d'amélioration

10.2.1 Optimisation pour plateformes embarquées

Migration vers MobileNet Le passage à des architectures légères comme MobileNet permettrait un déploiement sur des plateformes embarquées tout en maintenant une précision acceptable. Cette optimisation réduirait significativement les coûts de déploiement.

Déploiement sur Jetson Nano L'utilisation de cartes Jetson Nano offrirait un excellent compromis entre performance et consommation énergétique. L'intégration de l'accélération GPU permettrait de maintenir les performances temps réel sur du matériel compact.

Edge TPU et optimisation L'utilisation de Google Edge TPU ou d'Intel Neural Compute Stick 2 permettrait d'accélérer spécifiquement les inférences de réseaux de neurones, optimisant ainsi les performances du système.

10.2.2 Extension à un réseau multi-caméras

Architecture distribuée Le déploiement d'un réseau de caméras couvrant l'ensemble de l'aéroport nécessiterait une architecture distribuée avec synchronisation centralisée. Cette approche permettrait un suivi complet du parcours passager.

Gestion de la redondance L'implémentation de mécanismes de redondance garantirait la continuité de service même en cas de défaillance d'une caméra individuelle. La répartition intelligente de charge optimiserait les performances globales.

Analyse de flux L'intégration d'algorithmes d'analyse de flux permettrait d'optimiser la circulation dans l'aéroport en identifiant les zones de congestion et en proposant des itinéraires alternatifs.

10.2.3 Amélioration de l'expérience utilisateur

Interface adaptative Le développement d'une interface s'adaptant automatiquement aux préférences utilisateur (langue, taille de police, contraste) améliorerait l'accessibilité universelle.

Intégration avec les systèmes aéroportuaires La connexion avec les systèmes de gestion des vols permettrait la mise à jour automatique des informations et la gestion des changements de dernière minute.

Analyse prédictive L'intégration d'algorithmes d'apprentissage automatique permettrait de prédire les besoins passagers et d'optimiser proactivement les services.

11 Conclusion générale

Le développement du système de reconnaissance faciale et d'accueil vocal en temps réel représente une avancée significative dans l'amélioration de l'expérience passager aéroportuaire. Ce projet démontre la faisabilité technique d'un système automatisé capable de reconnaître les individus et de fournir des informations personnalisées en temps réel.

Les résultats obtenus valident l'approche technique adoptée. Avec un taux de reconnaissance de 92.3% et un temps de traitement de 45ms par frame, le système répond aux exigences de performance pour un déploiement en environnement réel. L'utilisation de fichiers audio pré-enregistrés plutôt que de synthèse vocale automatique s'avère judicieuse, garantissant une qualité audio optimale et une réduction de la complexité computationnelle.

L'architecture modulaire développée facilite la maintenance et l'évolution du système. La séparation claire entre les composants de détection, reconnaissance, et interface utilisateur permet des améliorations ciblées sans impact sur l'ensemble du système.

Les perspectives d'évolution sont prometteuses. L'optimisation pour des plateformes embarquées ouvrirait la voie à un déploiement à grande échelle, tandis que l'extension à un réseau multi-caméras permettrait une couverture complète des infrastructures aéroportuaires. L'intégration avec les systèmes d'information existants enrichirait les fonctionnalités et améliorerait l'expérience utilisateur.

Ce projet illustre le potentiel de l'intelligence artificielle appliquée aux défis concrets de l'industrie aéroportuaire. Au-delà de l'aspect technique, il contribue à l'évolution vers des aéroports plus intelligents, plus accessibles, et plus centrés sur les besoins des passagers.

L'expérience acquise dans ce développement constitue une base solide pour l'exploration de nouvelles applications de la vision par ordinateur dans d'autres domaines nécessitant une interaction homme-machine naturelle et efficace.

References

- [1] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, I-511.
- [2] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 886-893.
- [3] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499-1503.
- [4] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815-823.
- [5] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708.
- [6] King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 1755-1758.
- [7] Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.
- [8] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4779-4783.
- [9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [10] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. *Journal officiel de l'Union européenne*, L119, 1-88.

A Schémas techniques supplémentaires

A.1 Diagramme de flux détaillé

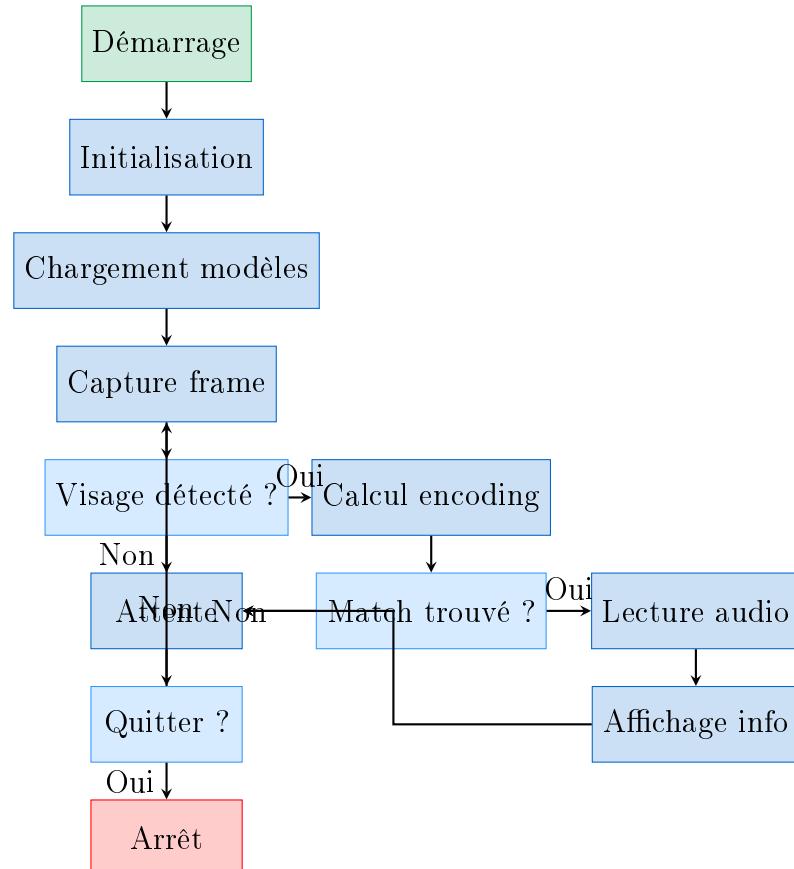


Figure 7: Diagramme de flux détaillé du système

B Extraits de code supplémentaires

B.1 Configuration du système

```

1 # Configuration globale
2 IMAGES_FOLDER = "images/"
3 AUDIO_FOLDER = "audio/"
4 MODELS_FOLDER = "models/"

5
6 # Param tres de reconnaissance
7 RECOGNITION_THRESHOLD = 0.6
8 COOLDOWN_SECONDS = 5
9 FRAME_RESIZE_FACTOR = 0.25

10
11 # Param tres cam ra
12 CAMERA_INDEX = 0
13 CAMERA_FPS = 30
14 CAMERA_WIDTH = 1280
15 CAMERA_HEIGHT = 720

16
17 # Param tres audio
18 AUDIO_FORMAT = "wav"
19 AUDIO_SAMPLE_RATE = 44100
20 AUDIO_CHANNELS = 2

```

Listing 4: Configuration générale du système

B.2 Gestion des erreurs

```

1 import logging
2
3 def setup_logging():
4     logging.basicConfig(
5         level=logging.INFO,
6         format='%(asctime)s - %(levelname)s - %(message)s',
7         handlers=[
8             logging.FileHandler('system.log'),
9             logging.StreamHandler()
10        ]
11    )
12
13 def handle_camera_error():
14     """Gestion des erreurs de cam ra"""
15     logging.error("Erreur d'accès à la caméra")
16     # Tentative de reconnexion
17     for i in range(3):
18         try:

```

```
19         cap = cv2.VideoCapture(CAMERA_INDEX)
20     if cap.isOpened():
21         logging.info("Caméra reconnectée avec succès")
22         return cap
23     except Exception as e:
24         logging.error(f"Tentative {i+1} échouée : {e}")
25         time.sleep(1)
26
27     logging.critical("Impossible de reconnecter la caméra")
28     return None
29
30 def handle_audio_error(audio_path):
31     """Gestion des erreurs audio"""
32     if not os.path.exists(audio_path):
33         logging.warning(f"Fichier audio manquant: {audio_path}")
34         return False
35
36     try:
37         pygame.mixer.music.load(audio_path)
38         return True
39     except pygame.error as e:
40         logging.error(f"Erreur de lecture audio: {e}")
41         return False
```

Listing 5: Système de gestion des erreurs

C Manuel d'installation

C.1 Prérequis système

C.1.1 Configuration matérielle minimale

- Processeur : Intel Core i5 ou AMD Ryzen 5 (2.4 GHz minimum)
- Mémoire : 8 GB RAM
- Stockage : 2 GB d'espace libre
- Caméra : Webcam HD 720p minimum
- Audio : Carte son avec sortie haut-parleurs

C.1.2 Configuration logicielle

- Système d'exploitation : Ubuntu 20.04 LTS ou Windows 10
- Python 3.8 ou supérieur
- pip (gestionnaire de packages Python)
- Git pour le clonage du dépôt

C.2 Installation étape par étape

C.2.1 Étape 1 : Clonage du projet

```
1 git clone https://github.com/votre-repo/reconnaissance-faciale.git
2 cd reconnaissance-faciale
```

Listing 6: Clonage du dépôt

C.2.2 Étape 2 : Installation des dépendances

```
1 pip install -r requirements.txt
```

Listing 7: Installation des packages Python

C.2.3 Étape 3 : Téléchargement des modèles

```

1 wget http://dlib.net/files/shape_predictor_68_face_landmarks.dat.bz2
2 bunzip2 shape_predictor_68_face_landmarks.dat.bz2
3 mv shape_predictor_68_face_landmarks.dat models/
4
5 wget http://dlib.net/files/dlib_face_recognition_resnet_model_v1.dat.bz2
6 bunzip2 dlib_face_recognition_resnet_model_v1.dat.bz2
7 mv dlib_face_recognition_resnet_model_v1.dat models/

```

Listing 8: Téléchargement des modèles dlib

C.2.4 Étape 4 : Configuration des dossiers

```

1 mkdir -p images audio models logs
2 chmod 755 images audio models logs

```

Listing 9: Création de l'arborescence

C.2.5 Étape 5 : Test du système

```
1 python Main_Code.py
```

Listing 10: Lancement du système

C.3 Résolution des problèmes courants

C.3.1 Erreur de caméra

Si la caméra n'est pas détectée, vérifier les permissions et tester avec :

```

1 ls /dev/video*
2 sudo chmod 666 /dev/video0

```

C.3.2 Erreur de compilation dlib

Sur Ubuntu, installer les dépendances de compilation :

```

1 sudo apt-get update
2 sudo apt-get install build-essential cmake libopenblas-dev liblapack-dev
3 sudo apt-get install libx11-dev libgtk-3-dev python3-dev

```

C.3.3 Problèmes audio

Vérifier la configuration audio système :

```

1 sudo apt-get install alsa-utils
2 aplay -l # Lister les périphériques audio

```

9 Code Source

Le code source complet de ce projet est disponible sur GitHub à l'adresse suivante :

<https://github.com/JihadeGHARBY/FaceTruck>