

Nom et prénom	Etablissement	Grade	Qualité
YAHYAOUY Ali	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
BADRADDINE Aghoutane	Faculté des sciences, Meknès	PES	Rapporteur & Examineur
TAIME Abderazzak	Ecole Supérieure de Technologie, USMS, Khénifra	MCH	Rapporteur & Examineur
MAHRAZ Mohamed Adnane	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Rapporteur & Examineur
RIFFI Jamal	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Examineur
EL FAZAZY Khalid	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Examineur
AZOUGH Ahmed	Ecole Supérieure d'ingénieurs Léonard de Vinci, Paris	MCH	Expert
TAIRI Hamid	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Directeur de thèse



## Résumé :

Chaque jour, le monde humain évolue et les zones habitées se transforment en villes intelligentes. Cette évolution suit le développement technologique imposé par l'émergence de l'intelligence artificielle, qui a envahi tous les domaines, y compris les systèmes de vidéosurveillance.

Le domaine de la vidéosurveillance s'est énormément développé ces dernières années. Aujourd'hui, on trouve des applications de vidéosurveillance partout : dans les hôpitaux, les centres commerciaux, les parkings, les gares et même dans la rue. La vidéosurveillance nous aide à observer plusieurs positions différentes en même temps et à reconnaître rapidement les événements anormaux qui se produisent dans une scène.

Cependant, la prolifération des caméras installées dans les espaces publics et privés rend de plus en plus difficile l'exploitation par des opérateurs humains des masses de données produites par ces systèmes. L'analyse de dizaines d'écrans de contrôle et d'enregistrements vidéo par un être humain est une tâche chronophage avec un taux d'erreur très élevé. De nombreuses techniques d'analyse vidéo automatique ont été étudiées en recherche et commencent même à être commercialisées pour des solutions industrielles.

L'une des méthodes les plus efficaces est celle des algorithmes d'apprentissage profond (deep learning). Ces algorithmes et méthodes ont été développés pour devenir des produits entièrement déployables avec des interfaces conviviales. Par exemple, l'apprentissage profond appliqué à la reconnaissance faciale est maintenant disponible dans des applications de moteur de recherche conçues pour trouver des personnes disparues à partir de séquences vidéo. Cependant, la plupart de ces algorithmes considèrent les caméras indépendamment les unes des autres.

L'objectif de cette thèse est de rendre possible la surveillance de grandes zones couvertes par de multiples caméras intelligentes légères et peu coûteuses, et de permettre aux différentes caméras intelligentes de communiquer au sein d'une plateforme distribuée. L'un des problèmes auxquels nous nous intéressons est la détection de la violence par l'analyse des séquences vidéo enregistrées par les caméras et l'identification des comportements malveillants sur la base de signatures directement liées à certains types d'attaques ou de mouvements bien connus. La méthode que nous proposons, basée sur un modèle CNN combiné à un réseau de mémoire bidirectionnelle à long terme (BiLSTM), doit être installée sur des caméras intelligentes légères, connectées les unes aux autres dans une architecture distribuée. Notre approche, mise en œuvre à l'aide de cartes Raspberry Pi, a donné de bons résultats. Des expériences ont été menées sur deux ensembles de données de référence : RLVS et RWF. La précision globale de la classification était respectivement de 92,47 % et de 91,40 %. Comparé aux méthodes de pointe, notre modèle montre de bonnes performances tout en réduisant le nombre de paramètres et le coût de calcul.

## Mots clés :

Système de vidéosurveillance, Apprentissage profond, Camera intelligente, Détection de violence, Réidentification des personnes, Requête NoSQL ;



## DISTRIBUTED VIDEO SURVEILLANCE SYSTEMS AND EVENT DETECTION THROUGH DEEP LEARNING AND NOSQL DATABASES

### Abstract :

Every day, the human world evolves, and populated areas develop towards intelligent cities. This evolution follows the technological development imposed by the emergence of artificial intelligence, which has invaded every field, including video surveillance systems.

The field of video surveillance has expanded enormously in recent years. Today, video surveillance applications can be found anywhere: in hospitals, shopping centers, car parks, railway stations and even in the street. Video surveillance helps us to observe many different positions at the same time and to quickly recognize abnormal events taking place in a scene.

However, the proliferation of cameras installed in public and private spaces is making it increasingly difficult for human operators to exploit the masses of data produced by these systems. The analysis of dozens of control screens and video recordings by a human being is a time-consuming task with a very high error rate. Numerous automatic video analysis techniques have been studied in research and are even beginning to be commercialized for industrial solutions.

One of the most effective of these methods is deep learning algorithms. These algorithms and methods have been developed into fully deployable products with user-friendly interfaces. For example, deep learning applied to facial recognition is now available in search engine applications designed to find missing persons from video footage. However, most of these algorithms consider cameras independently of each other.

The aim of this thesis is to make it possible to monitor large areas covered by multiple lightweight and low-cost smart cameras, and to enable the various intelligent cameras to communicate in a distributed platform. One of the problems we are interested in is violence detection by analysing video footage recorded by cameras and identifying malicious behaviour based on signatures that are directly related to certain types of well-known attacks or movements. Our proposed method, based on a CNN model combined with Bidirectional Long Short-Term Memory network (BiLSTM), is to be installed on lightweight smart cameras, connected to each other in a distributed architecture. Our approach, implemented using Raspberry Pi boards, has shown good results. Experiments were carried out on two reference datasets: RLVS and RWF. The overall classification accuracy was 92.47\% and 91.40\% respectively. Compared to the state-of-the-art methods, our model shows strong performance while reducing the number of parameters and computational cost.

### Key Words :

Video surveillance system , Deep learning, Smart camera, Violence detection, Person re-identification, NoSQL Database;