AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mme (elle) **CHOUHAYEBI Hajar**
Soutiendra : **le Samedi 21/09/2024** à **15H00**
Lieu : ***FSDM – Centre Visioconférence***

Une thèse intitulée :

« Human Emotion Recognition Based on Spation-Temporal Facial Features Using Deep Learning methods and textures Features »

En vue d'obtenir le ***Doctorat***

FD : ***Sciences et Technologies de l'Information et de la Communication***
Spécialité : ***Informatique***

Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
Pr TAIRI Hamid	Faculté des Sciences Dhar El Mahraz, Fès	PES	Président
Pr TAIME Abderazzak	Ecole Supérieure de Technologie, Khénifra	MCH	Rapporteur & Examinateur
Pr SABBANE Mohamed	Faculté des Sciences, Meknès	PES	Rapporteur & Examinateur
Pr EL FAZAZY Khalid	Faculté des Sciences Dhar El Mahraz, Fès	PES	Rapporteur & Examinateur
Pr RIFFI Jamal	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Examinateur
Pr BENNANI Mohamed taj	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Examinateur
Pr YAHYAYOUY Ali	Faculté des Sciences Dhar El Mahraz, Fès	PES	Examinateur
Pr MAHRAZ Mohamed Adnane	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Directeur de thèse

Résumé :

La reconnaissance des émotions se profile comme un domaine central de la recherche scientifique, embrassant diverses modalités telles que les expressions faciales, le langage corporel et les indices audio. L'objectif de doter les machines de la capacité à discerner et à interpréter les émotions humaines pour des interactions plus subtiles a engendré le développement de multiples méthodologies. Malgré des progrès notables, ce défi demeure de taille, rassemblant des domaines variés tels que l'interaction homme-machine, le traitement d'images, l'intelligence artificielle et la robotique.

Au fil des années, la recherche en analyse des émotions s'est principalement concentrée sur des approches spécifiques, avec une attention particulière portée à l'analyse des expressions faciales en raison de son importance dans la communication non verbale, contribuant jusqu'à 55% à la compréhension des émotions humaines. Bien que des avancées aient été réalisées en termes de précision, les premières méthodologies ont largement reposé sur des techniques de classification de caractéristiques essentielles.

Les récents efforts se sont tournés vers l'exploitation des techniques d'apprentissage profond pour extraire automatiquement des caractéristiques informatives à partir des données, dans le but de les intégrer et de les classifier efficacement. Cependant, des défis persistent dans la création de systèmes d'apprentissage profond et robustes capables de prédire avec précision les émotions humaines, notamment dans la modélisation des interactions spatio-temporelles dans les données vidéo et dans l'identification des caractéristiques essentielles pour améliorer la précision.Dans ce contexte, cette thèse doctorale aborde ces défis de front, notamment dans le domaine de la reconnaissance des expressions faciales dans des contextes spatio-temporels. Elle propose une fusion de méthodes d'apprentissage profond, de techniques d'apprentissage automatique et d'approches basées sur la texture pour affiner la précision de la reconnaissance des émotions.

La thèse présente deux architectures novatrices pour la reconnaissance des expressions faciales :

La première architecture combine les représentations des caractéristiques d'apprentissage profond et des caractéristiques de texture dynamique. En utilisant le modèle Groupe de Géométrie Visuelle (VGG19) pour l'apprentissage profond, les traits faciaux sont extraits et utilisés pour alimenter des cellules de mémoire à court terme (LSTM) afin de capturer les nuances spatio-temporelles entre les images. Simultanément, le descripteur HOG-TOP est employé pour capturer les textures dynamiques à partir de séquences vidéo, caractérisant efficacement les changements d'apparence faciale. Ces modèles sont ensuite combinés à l'aide de l'algorithme Multimodal Bilinéaire Compact (MCB), aboutissant à un vecteur de descripteur robuste.La deuxième architecture présente une fusion innovante d'algorithmes d'apprentissage profond et de méthodes de texture dynamique. Dans la phase initiale, les traits faciaux sont extraits en utilisant le modèle VGG19 et introduits dans des cellules LSTM pour capturer des informations spatio-temporelles. De plus, le descripteur HOG-HOF est utilisé pour capturer les caractéristiques dynamiques à partir de séquences vidéo, encapsulant les changements d'apparence faciale au fil du temps. La fusion de ces modèles à l'aide de l'algorithme MCB produit un vecteur de descripteur efficace.

Les résultats expérimentaux des deux méthodologies démontrent une performance supérieure par rapport aux approches de pointe existantes, validées à l'aide du jeu de données Enterface'05. Alors que la première méthode présente une précision exceptionnelle, la seconde méthode surpasse les approches contemporaines, soulignant ainsi l'efficacité des approches proposées.

Mots clés :
La reconnaissance des émotions humaines ; la reconnaissance des expressions faciales ; l'apprentissage profond ; le groupe de géométrie visuelle ; la mémoire à court et long terme ; la machine à vecteurs de support ; les histogrammes de gradients orientés ; l'histogramme du flux optique ; l'histogramme des gradients orientés à partir de trois plans orthogonaux.

HUMAN EMOTION RECOGNITION BASED ON SPATIO-TEMPORAL FACIAL FEATURES USING DEEP LEARNING METHODS AND TEXTURES FEATURES

Abstract:

Emotion recognition stands out as a focal point in scientific inquiry, encompassing various modalities such as facial expressions, body language, and audio cues. The quest to endow machines with the ability to discern and interpret human emotions for more nuanced interactions has spurred the development of multiple methodologies. Despite significant advancements, this endeavor remains a formidable challenge, drawing together disparate disciplines such as human-computer interaction, image processing, artificial intelligence, and robotics.

Over the years, research in emotion analysis has primarily focused on specific approaches, with particular attention given to facial expression analysis due to its significance in nonverbal communication, contributing up to 55% to the understanding of human emotions. While advances have been made in terms of accuracy, early methodologies largely relied on essential feature classification techniques.

Recent efforts have turned towards leveraging deep learning techniques to automatically extract informative features from data, aiming to integrate and classify them effectively. However, challenges persist in creating deep and robust learning systems capable of accurately predicting human emotions, particularly in modeling spatio-temporal interactions in video data and identifying pivotal features to enhance accuracy.In this context, this doctoral thesis addresses these challenges head-on, particularly in the domain of facial expression recognition within spatio-temporal contexts. It proposes an amalgamation of deep learning methods, machine learning techniques, and texture-based approaches to refine emotion recognition accuracy.

The thesis introduces two novel architectures for facial expression recognition:

The first architecture amalgamates representations of deep learning features and dynamic texture features. Leveraging the VGG19 model for deep learning, facial features are extracted and fed into Long Short Term Memory (LSTM) cells to capture spatio-temporal nuances between frames. Simultaneously, the (Histogram of Oriented Gradients from Three Orthogonal Planes) HOG-TOP descriptor is employed to capture dynamic textures from video sequences, effectively characterizing changes in facial appearance. These models are then combined using the Multimodal Compact Bilinear (MCB) algorithm, resulting in a robust descriptor vector.The second architecture presents an innovative fusion of deep learning algorithms and dynamic texture methods. In the initial phase, facial features are extracted using the Visual-Geometry-Group (VGG19) model and input into LSTM cells to capture spatio-temporal information. Additionally, the HOG-HOF descriptor is utilized to capture dynamic features from video sequences, encapsulating changes in facial appearance over time. The fusion of these models using the Multimodal Compact Bilinear (MCB) model yields an effective descriptor vector.

Experimental results from both methodologies showcase superior performance compared to existing state-of-the-art approaches, validated using the Enterface'05 dataset. While the first method demonstrates exceptional accuracy, the second method surpasses contemporary methodologies, underscoring the efficacy of the proposed approaches.

Key Words :
human emotion recognition; facial expression recognition; deep learning; visual geometry group; long short term memory; support vector machine; histograms of oriented gradients; histogram of optical flow, Histogram of Oriented Gradients from Three Orthogonal Planes