الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي 08 ماي 1945. بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Engineer**
Field: **Computer Science**
Specialty: **Ingénierie des Systèmes Informatiques (ISI)**

Theme

Development and Evaluation of an Intelligent Multimodal Human-Computer Interaction System Integrating Gaze, Gesture, and Speech Recognition

---

Presented by:
Tebabna Ahmed Rami

Submission Date: **XX, 2024**
In front of the jury composed of:

Mr. Someone	President
Ms. Bedjaoui Mohamed	Supervisor
Mr. Someone	Examiner

Academic Year : 20023/2024

Abstract

In the context of enhancing human-computer interaction, this thesis presents the development and evaluation of an intelligent multimodal HCI system incorporating gaze estimation, gesture recognition, and speech recognition. The problem addressed is the need for a more intuitive and seamless user interaction experience. The objective was to integrate advanced computer vision techniques using MediaPipe for real-time gaze and gesture tracking, and Vosk for robust speech recognition, aiming to improve interaction efficiency and accessibility. Comprehensive experiments were conducted to assess the system's accuracy and responsiveness under various conditions. The results demonstrated significant improvements in interaction fluidity and user satisfaction, highlighting the potential of multimodal HCI systems in diverse applications. In conclusion, the integrated system not only enhances user experience but also sets a precedent for future developments in creating more intuitive and efficient HCI systems.

Keywords— Human-computer interaction, distributed systems, deep learning, Natural language processing, gestures recognition, voice user interface

الملخص

في سياق تحسين التفاعل بين الإنسان والحاسوب، يقدم هذا البحث تطوير وتقييم نظام تفاعل بين الإنسان والحاسوب (HCI) متعدد الوسائط الذكي الذي يتضمن تقدير النظرة، التعرف على الإيماءات، والتعرف على الكلام. تتمثل المشكلة التي تم معالجتها في الحاجة إلى تجربة تفاعل مستخدم أكثر سلاسة وبديهية. كان الهدف هو دمج تقنيات الرؤية الحاسوبية المتقدمة باستخدام MediaPipe لتتبع النظرة والإيماءات في الوقت الفعلي و Vosk للتعرف على الكلام بدقة، بهدف تحسين كفاءة التفاعل وسهولة الوصول. تم إجراء تجارب شاملة لتقييم دقة النظام واستجابته في ظل ظروف مختلفة. أظهرت النتائج تحسينات كبيرة في سلاسة التفاعل ورضا المستخدم، مما يبرز إمكانيات أنظمة HCI متعددة الوسائط في تطبيقات متنوعة. في الختام، لا يعزز النظام المتكامل تجربة المستخدم فحسب، بل يضع أيضًا سابقة للتطورات المستقبلية في إنشاء أنظمة HCI أكثر بديهية وكفاءة.

الكلمات المفتاحية: التفاعل بين الإنسان والحاسوب، الأنظمة الموزعة، التعلم العميق، معالجة اللغة الطبيعية، التعرف على الإيماءات، واجهة المستخدم الصوتية.

Résumé

Dans le contexte de l'amélioration de l'interaction homme-machine, cet thèse présente le développement et l'évaluation d'un système intelligent d'interaction homme-machine (HCI) multimodal intégrant l'estimation du regard, la reconnaissance des gestes et la reconnaissance vocale. Le problème abordé est la nécessité d'une expérience d'interaction utilisateur plus intuitive et fluide. L'objectif était d'intégrer des techniques avancées de vision par ordinateur utilisant MediaPipe pour le suivi en temps réel du regard et des gestes, et Vosk pour une reconnaissance vocale robuste, visant à améliorer l'efficacité et l'accessibilité de l'interaction. Des expériences complètes ont été menées pour évaluer la précision et la réactivité du système dans diverses conditions. Les résultats ont démontré des améliorations significatives de la fluidité de l'interaction et de la satisfaction des utilisateurs, soulignant le potentiel des systèmes HCI multimodaux dans diverses applications. En conclusion, le système intégré améliore non seulement l'expérience utilisateur, mais établit également un précédent pour les développements futurs dans la création de systèmes HCI plus intuitifs et efficaces.

Mots clés: Interaction homme-machine, systèmes distribués, apprentissage profond, traitement du langage naturel, reconnaissance des gestes, interface utilisateur vocale