Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

An Intelligent System to Encode Arabic Alphabet Based on Eyelids Movement

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Doctor of Philosophy in
Computer Science**

By

Meaad Mohammed Salih Sultan

Supervised by

Prof. Dr. Khalil Ibrahim Al Saif

2023 A.D. **1445 A.H.**

Abstract

The use of gestures as a way for communication between computers and humans has received increasing attention from researchers in recent years, especially those researches that employed computer vision methods to recognize gestures. Eyelid movement (or blinking of the eye) is one of the most important eye gestures that have been used in many vision-based computer applications for a variety of purposes. In this thesis, two methods have been developed to generate codes utilizing eye gestures: the first method is by adopting the movement of the eyelids (eye blinking), and the second method is by adopting five Gaze Directions.

The proposed system to generate codes using the eye blink is based on blink detection process which requires classification of the eye state (closed or open) during a number of consecutive video frames. Three approaches of eye state classification were proposed and tested on two standard video datasets (Talking Face, Eyeblink8) and a video acquired through a webcam.

Approach 1(using EAR): classifying the eye state using the geometric features of the eye by calculating the EAR (Eye Aspect Ratio) value, a polynomial curve fitting was used to estimate six new eye landmarks, three on the upper eyelid and the others on the lower eyelid, thus calculating the EAR value based on five vertical distances instead of two to increasing the accuracy of the EAR value calculation. The resulting EAR value is compared to the $EAR_{threshold}$ to determine whether the eye is closed or open. The test results showed that using five vertical distances to calculate the EAR instead of two caused an increase in the AUC value (area under the Roc curve) by 6% from 80.4 to 85.26, the same was true for the accuracy from 77.89 to 70.39, i.e. an increase of 10.6%.

Approach2 (Using deep learning) : where a convolutional neural network model (EyeCNN) was proposed, trained and tested on the standard image dataset(MrlEye).

Approach3 (Using EAR and Deep Learning): This approach was based on the results of the classification of the two previous approaches. The final decision for the eye state is obtained by combining previous classifications' results based on the majority voting. The experimental

showed that Approach3 achieved a preference over the last two approaches with an accuracy of 97.7%, and the AUC equals 88.12%, i.e. an increase of 3.35 and 6.03 over its value in Approach1 and Approach2 which were 85.26 and 83.1, respectively. The three Approaches achieved higher accuracy in classifying the eye state than the previous works.

Code Generation based on eye blinking included proposing new alphabetical codes for Arabic letters (**Blink Code)** based on the duration of the blink of an eye (long/short blinks). The generated blink code was applied as Arabic text entry approach in addition to upgraded to a binary code by representing the long blink with (1) and the short blink with (0) and used in the proposed Coverless Data Hiding technique. Experimental results showed that the proposed technique is more robust against geometric attacks than previous works.

Code generation based on gaze direction classification involved **designing an Encoding Triple Tree (ETT)** to encode Arabic letters and numbers using gaze directions. A Convolutional Neural Network (CNN) model called (GD_CNN) was proposed to classify five eye directions (right, left, up, down, and center). The standard database (Chimera) was used for the purpose of training and testing, and GD_CNN had an accuracy of 95.6% on the test samples and 99% on the training samples. The coding methods (Blink Code, ETT) proposed in this thesis were tested as approaches for entering Arabic text through eye blinking or changing gaze directions.

The experimental results showed that the text input speed using the ETT method reached 2.07 WPM (word per minute) with an error rate of 0.015 and 2.35 WPM using Blink Code with an error rate of 0.06.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

نظام ذكائي لترميز الابجدية العربية بالاعتماد على حركة اجفان العيون

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

ميعاد محمد صالح سلطان

بإشراف
أ.د. خليل إبراهيم السيف

1445هـ 2023 م

المستخلص

لاقى استخدام الايماءات كأداة للتواصل بين الكمبيوتر والإنسان اهتماما متزايدا من قبل الباحثين في السنوات الأخيرة. حركة الاجفان (او غمز العين/رمشة العين) هي من اهم ايماءات العين التي استخدمت في العديد من التطبيقات الحاسوبية القائمة على الرؤية ولأغراض متنوعة. في هذه الاطروحة تم توظيف اسلوبين لتوليد الرموز باعتماد ايماءات العين: الطريقة الأولى باعتماد حركة الاجفان (ومضات العين Eye Blinking), والطريقة الثانية باعتماد أربعة اتجاهات لقزحية العين Gaze Directions.

النظام المقترح لتوليد الرموز باستخدام ومضة العين يعتمد على تحديد ومضات العين في الفيديو Blink Detection والتي تتطلب تصنيف حالة العين (مغلقة ام مفتوحة) خلال عدد من إطارات الفيديو(frames) المتتالية. تم اقتراح ثلاثة اساليب لتصنيف حالة العين واختبارها على قاعدتي بيانات فيديوية قياسية (Talking Face, Eyeblink8) بالإضافة الى فيديوهات تم اكتسابها من خلال كاميرا الحاسوب.

الأسلوب الأول : تصنيف حالة العين باستخدام السمات الهندسية للعين وذلك من خلال حساب قيمة EAR. لغرض زيادة دقة حساب قيمة EAR تم استخدام موائمة منحنى متعدد الحدود لتخمين ستة معالم جديدة للعين, ثلاثة على الجفن العلوي والثلاثة الاخرى على الجفن السفلي وبالتالي حساب قيمة EAR اعتمادا على خمسة مسافات راسية بدلا من اثنتين. قيمة EAR الناتجة يتم مقارنتها بحد العتبة $EAR_{threshold}$ لتحديد هل العين مغلقة ام مفتوحة. أظهرت نتائج الاختبار ان استخدام خمسة مسافات راسية لحساب EAR بدلا من اثنتين احدث زيادة في قيمة AUC (Area Under The Roc Curve) بنسبة 6% من 80.4 الى 85.26 وكذلك بالنسبة للدقة اذ تحسنت قيمتها من 77.89 الى 70.39 اي بنسبة 10.6% .

الأسلوب الثاني: استخدام التعلم العميق, حيث تم اقتراح نموذج شبكة عصبية التفافية (EyeCNN) وتدريبه واختباره على قاعدة البيانات القياسية (MrlEye).

الأسلوب الثالث: اعتمد هذا الأسلوب على نتائج تصنيف الاسلوبين السابقين. اذ يتم الحصول على القرار النهائي لحالة العين من دمج نتائج التصانيف السابقة بالاعتماد على التصويت بالأغلبية. أظهرت النتائج ان هذا الاسلوب حقق افضلية عن الاسلوبين السابقين بدقة تصل الى 97.7%, و AUCيساوي 88.12% اي زيادة بنسبة 3.35 و 6.03 عن قيمته في الأسلوبين الأول والثاني

والتي كانت تساوي 85.26 و 83.1 على التوالي. الأساليب الثلاثة حققت دقة اعلى بتصنيف حالة العين مقارنة بالأعمال السابقة.

توليد الرموز باستخدام رمشة العين تضمن اقتراح رموز لهجائية جديدة للحروف العربية سميت (**Blink Code**) باعتماد الفترة الزمنية لغمزة العين (غمزات طويلة/ قصيرة)، تم تطبيق رمز الوميض الناتج كطريقة لإدخال النص العربي بالإضافة إلى ترقيته إلى رمز ثنائي من خلال تمثيل الوميض الطويل بـ (1) والومض القصير بـ (0) واستخدامه في تقنية إخفاء البيانات غير المغطاة المقترحة. أظهرت النتائج التجريبية أن التقنية المقترحة أكثر قوة ضد الهجمات الهندسية من الأعمال السابقة.

اما توليد الرموز باستخدام اتجاهات النظر فتضمنت تصميم **شجرة ترميز ثلاثية ETT باستخدام اتجاهات نظرة العين** لترميز الحروف والأرقام العربية. تم اقتراح نموذج شبكة عصبية التفافية (GD_CNN) لتصنيف خمسة اتجاهات للعين (يمين، يسار, اعلى، أسفل، وسط). استخدمت قاعدة البيانات القياسية (Chimera Dataset) لغرض التدريب والاختبار وكانت GD_CNN ذات دقة وصلت الى 95.6% على بيانات الاختبار و99% على بيانات التدريب. تم اختبار طرق الترميز (Blink Code, ETT) التي تم اقتراحها في هذه الاطروحة كأساليب لإدخال النصوص العربية من خلال ومض العين او تغيير اتجاه نظرة العين. أظهرت النتائج ان سرعة ادخال النص باستخدام أسلوب ETT وصلت 2.07 WPM ونسبة خطأ 0.015 و 2.35 WPM باستخدام Blink Code وبنسبة خطأ 0.06.