العنوان:	تمييز حروف المسند باستخدام الخوارزميات الجينية
المصدر:	المجلة العراقية للعلوم الإحصائية
الناشر:	جامعة الموصل - كلية علوم الحاسوب والرياضيات
المؤلف الرئيسي:	مجيد، جمال صلاح الدين سيد
مؤلفين آخرين:	طه، طارق حازم، على، سيل وليد(م. مشارك)
المجلد/العدد:	ع 20
محكمة:	نعم
التاريخ الميلادي:	2011
الصفحات:	761 - 785
رقم MD:	422079
نوع المحتوى:	بحوث ومقالات
قواعد المعلومات:	EcoLink
مواضيع:	البرامج الاحصائية ، الخوارزمية الجينية ، الاساليب الاحصائية ، تكنولوجيا المعلومات ، الحاسبات الالكترونية ، الحروف السبئية ، البرامج الالكترونية
رابط:	http://search.mandumah.com/Record/422079

المجلة العراقية للعلوم الإحصائية (20) 2011
عدد خاص بوقائع المؤتمر العلمي الرابع لكلية علوم الحاسوب والرياضيات
ص ص [761–785]

تمييز حروف المسند باستخدام الخوارزميات الجينية

د.جمال صلاح الدين سيد مجيد * طارق حازم طه ** سيل وليد علي***

الخلاصة

في هذا البحث، تم بناء نظام ذكائي لتمييز حروف المسند (السبئية) الداخلة بواسطة الماسح الضوئي وخزنه تحت امتداد (bmp.) وذلك للمحافظة على تفاصيل الحرف ودقته. وتم بناء النظام ألذكائي بالاعتماد على تقنيات الخوارزمية الجينية اذ تمت عملية قراءة وتحليل الصورة الداخلة من الماسح الضوئي وتقطيعها (Segmentation) إلى سطور وأحرف واستخلاص خواصها بطريقة ايجاد العزوم السبعة اللاعتمادية (Seven Invariant Moment) وبناء الجيل الابتدائي في الخوارزمية الجينية لجميع الأحرف السبئية من (أ) الى (ي) فضلا عن الخط العمودي الذي يفصل بين الكلمات التي يبلغ عددها 33 كروموسوماً. اذ أن كل كروموسوم يكون مكوناً من سبعة جينات (Seven genes) والتي تمثل العزوم الثابتة أللاعتمادية (Seven Invariant Moments) لحرف معين .

يهدف البحث الى تحويل صورة الحرف السبئي الى نص الكتروني للحرف السبئي بهيئة text مقروء داخل الحاسبة، اذ تتم الاستفادة منه لاغراض التحويل السريع لمحتوى الورقة للنص المكتشف بغية ارشفته حاسوبياً ومن ثم ترجمته الى اللغة العربية او الانكليزية.

Abstract

In this research, an intelligent system has been designed for recognizing Almusnad (Saba's) letters that entered to the computer by using a scanner and saving them under (.bmp) extension to keep the litters' shape and accuracy. The intelligent system is constructed depending on Genetic algorithms technique, then some processes are over the entered

* استاذ مساعد/كلية علوم الحاسوب والرياضيات/جامعة الموصل.

** مدرس مساعد/كلية علوم الحاسوب والرياضيات/جامعة الموصل.

*** مدرس /كلية علوم الحاسوب والرياضيات/جامعة الموصل.

تاريخ التسلم :2011/10/1 ـــــــــــ تاريخ القبول : 21 /12 /2011

picture like reading, analyzing and segmented to lines and litters, after that an extraction of their features was done by finding the Seven Invariant Moments and constructs the initial (generation) in the Genetic algorithm for all Saba's letters from 'a' to 'z', in addition to the vertical which separates the words of 33 chromosome's length.

The aim of this research is to convert the image of Saba's letter to an electronic reading text in the computer, where they are used for the purpose of faster translation for the content of the discovered text paper and saving it in the computer then translate it to the Arabic or English language.

1- المقدمة

1-1 نظرة تاريخية في ابجدية كتابة المسند

تستقل كتابة عرب الجنوب (اليمنيين)عن كتابات عرب الشمال وكتابات الساميين الغربية الشمالية تماماً. ولهذه الكتابة قسمان [6]:

1. قسم شمالي موطنه شمال غرب الجزيرة العربية الى جنوبي سوريا وداخلها يشتمل على الكتابة اللحيانية والثمودية والصفوية.
2. قسم جنوبي يشتمل على الكتابة السبئية (المسند) والكتابة الاثيوبية.

تدعى الكتابة السبئية عند الإخباريين (مؤرخو العرب القدامى) بـ (الخط المسند) وعند المستشرقين (الكتابة العربية الجنوبية القديمة). وهي كتابة شعب حضاري سكن اليمن ومايجاوره من بقاع، وامتد نفوذه الى شمالي شبه الجزيرة العربية ووسطها والى الساحل الافريقي. ويلاحظ من خلال الاثار الكتابية ان تاريخها يمتد مابين القرن الثامن قبل الميلاد الى القرن السابع الميلادي. [15]

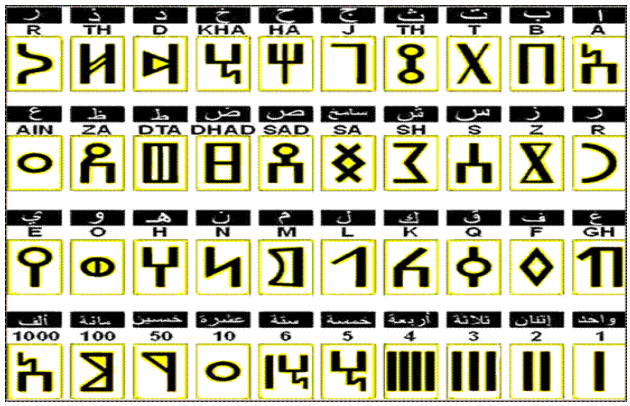
واعترض البعض على تاريخها معتبرا ان اقدم نماذج هذا الخط لاتتعدى القرن الخامس قبل الميلاد [8]. بيد ان هذا الاعتراض يتناقض مع اكتشاف نقش على تابوت تاجر معيني وجد في مصر يرجع الى عهد قمبيز بن كورش الاخمني وتاريخ نقشه سنة 525 ق.م اي في القرن السادس قبل الميلاد وليس الخامس [16].

بلــوغ الخط المسند من الانتشار ان عثر عليه في مدينة اور في العراق اذ عثر المنقب البريطاني ليوناردو وولي على كتابة مخطوطة يرجع تاريخها الى القرن السادس والسابع قبل الميلاد. وطبعت هذه الكتابات وحلت رموزها في المجلة الاسيوية الانكليزية [21].

وجاءتنا كتابة العرب الجنوبيين محفورة ومنقوشة على الحجر والخشب وانواع من المعادن مثل النحاس والقصدير والحديد والذهب والفضة وعلى القبور والمذابح والنقود والتماثيل.

2-1 سمات كتابة المسند:

1. تتألف ابجدية المسند من 29 حرفاً وهي حروف صامتة أي خالية من حروف الترميز الى حركات انظر الشكل(1) [3] .
2. يكون اتجاه الكتابة من اليمين الى الشمال وقد يكتب على الطريقة الثعبانية او ماتسمى دوران الثعبان وهي ان يكتب سطر من اليمين الى الشمال والذي بعده من الشمال الى اليمين وهكذا الى نهاية الكتابة[17].
3. ليس في حروف المسند شيء من النقط او الحركات او الاشارات [12].
4. لايوجد في المسند علامة لتشديد الحرف بل يكتب الحرف مرتين.
5. استعمال الحروف الصامتة جعل من الصعب البت في كيفية النطق بالكلمات والتعرف بسهولة على موضع الكلم في الاعراب [13].
6. لم يترك لنا كتبة المسند كتابة تشير الى ترتيب حروف الهجاء عندهم واسمائها التي كانت تعرف بها [13].
7. يوضع خط عمودي للفصل بين الكلمات وترتبط حروف الجر وادوات الوصل بالكلمة التالية ولاتفصل بهذا الخط [1] .
8. للفصل بين الارقام والكلمات يوضع الرمز | قبل الرقم وبعده [1].



الشكل(1) أبجدية الكتابة السبئية (المسند) وأرقامها

2- اهمية البحث

ان للبحث اهمية في دراسة الكتابات السبئية الورقية القديمة وذلك بتحويلها داخل الحاسبة مما يسهل معالجتها وتحويلها الى باقي اللغات الحية ليتسنى الاستفادة من هذه المخطوطات في مختلف ابواب العلوم .

3- التحليل

البيانات الداخلة سوف تكون عبارة عن صورة تحتوي على الاحرف السبئية والتي سوف يتم تحويلها الى صورة ثنائية لتسهيل عملية التعرف على النص فيتم تقطيعها الى سطور ومن ثم الى احرف لتدخل طور استخلاص صفات كل حرف وبناء فضاء الخواص لجميع الاحرف ليتسنى تمييز الاحرف الداخلة مستقبلا بالاعتماد على تقنية الخوارزمية الجينية.

4- تعريف الخوارزمية الجينية

هي تقنية بحث، تستخدم في مجال الذكاء الاصطناعي (artificial intelligence) وتحديداً في فرع البحث وحل المشاكل (Problem solving and search) اذ تقوم بإيجاد أفضل الحلول لمشاكل التحسين (optimization problems) بالاعتماد على العشوائية (stochastic) في البحث[2]. كما تستخدم الخوارزمية الجينية في التطبيقات المعلوماتية الإحيائية (bioinformatics) و علوم الحاسوب والهندسة و الاقتصاد و الكيمياء و الصناعات التحويلية (manufacturing) و الرياضيات والفيزياء وغيرها من الميادين [18] .

4-1 المصطلحات الأساسية في الخوارزمية الجينية

يوضح الجدول (1) قسما من المصطلحات الأساسية لعلم الوراثة وما يقابلها في الخوارزمية الجينية [5].

ألجين (Gene)	**القيمة(Value):تمثل الوحدة الأساسية في هيكلية بيانات الخوارزمية الجينية التي قد تكون قيمة عددية أو حرفية أو ثنائية.**
الكروموسوم (Chromosome)	**المقطع(String):هومجموعة من القيم(Gene)الأساسية التي تمتلك مجموعة من الحلول لتطبيق معين.**
الفرد (Individual)	**يمثل الكروموسوم.**
حجم المجتمع (population size)	**عدد الأفراد في الجيل(عدد الكروموسومات في الجيل).**

الجيل (Generation)	**يمثل مجموعة من الأفراد المتكونة في وقت معين.**
التعبير (phenotype)	**تمثل فضاء الحلول للأفراد، يترجم قيم الكروموسومات.**
الأليل (Genotype)	**يمثل التمثيل الجيني للأفراد.**

جدول(1) المصطلحات الأساسية في الخوارزمية الجينية

5- مقدمة في تمييز الأنماط

ان تمييز الأنماط أو التعرف على الأنماط Pattern recognition [23] هو أحد فروع علم معالجة الصور الرقمية والذكاء الإصطناعي.

تهدف البحوث والتقنيات الخاصة بهذا العلم إلى إيجاد أو تطوير تقنيات للتعرف على أنماط أو هياكل محددة في الصور الرقمية (اشارة رقمية ذات بعدين)، أو أشارات الرادار (اشارة رقمية ذات بعد واحد). وللتعرف على نمط او جسم في صورة يجب أولا الحصول على معلومات أولية a priori knowledge وإحصائية مباشرة عن الصورة التي يجب أن تكون رقمية لنتمكن من التعامل معها بالحاسوب. كذلك يجب تصنيف المعلومات الموجودة في الصورة لتسهيل عمليات لاحقة على الصورة مثل استخلاص المعلومات.

ان من التطبيقات المهمة والمفيدة؛ تمييز الحروف والذي يمكن إنجازه رقميا باستعمال الحاسب أو بصريا باستعمال أجهزة وتقنيات بصرية. ومعنى تمييز الأنماط موضح في الشكل(2) [11].

ان مراحل التعرف على أي نمط تمر بمرحلتين:

- مرحلة التعلم Learning.
- مرحلة التصنيف او التمييز Classification or recognition.

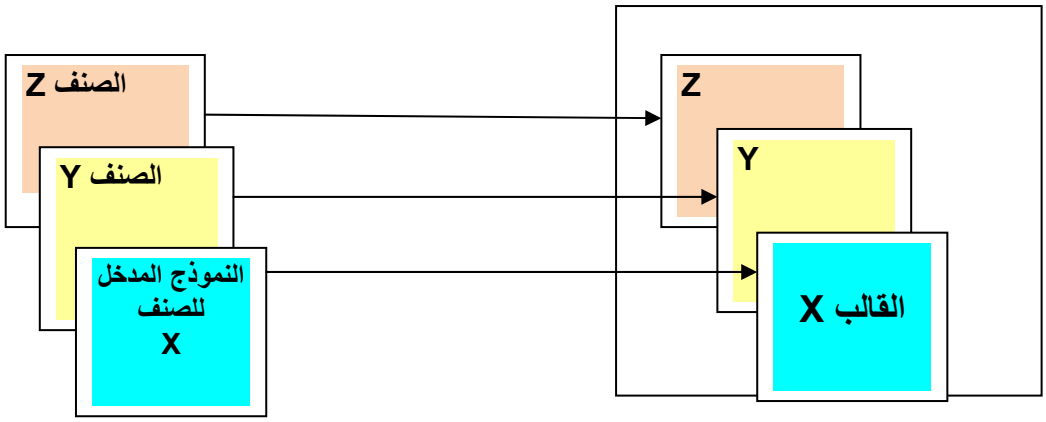
يوجد طرائق أساسية مستخدمة في تمييز الأنماط منها]11 [:

- طريقة قالب المطابقة.
- الطريقة الإحصائية.
- طريقة تمييز الانماط باستخدام الشبكات العصبية [14].
- طريقة تمييز الانماط باستخدام الخوارزمية الجينية [14].
- طريقة تمييز الانماط باستخدام النظام المناعي [10] .

الشكل(2) تمييز الانماط ومعالجة الصور

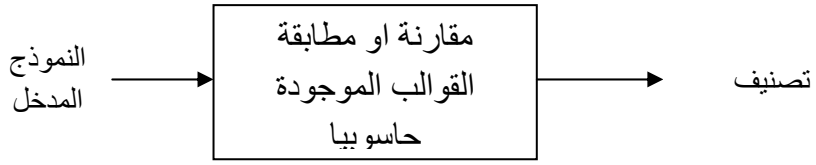
1-5 طريقة قالب المطابقة

مرحلة التعليم في هذه الطريقة تقوم على تخزين مجموعة من القوالب Templates أو النماذج Prototypes، قالب من كل صنف في الحاسوب [9] كما موضح في الشكل(3) [11] :



الشكل(3) مجموعة من القوالب

وفي مرحلة التصنيف تقارن الصورة الداخلة Input pattern مع كل صنف templates فإن كانت نتيجة مقارنتها مع الصنف x أكبر من نتيجة مقارنتها مع الصنف y فإنها تصنف ضمن الصنف x وهكذا أي انها تصنف ضمن القالب الاقرب اليه ويتم ذلك بعد مقارنتها بواسطة احدى الطرائق المقارنة (المسافة الاقليدية وغيرها) مع جميع القوالب. كما موضح في الشكل(4) :[11]



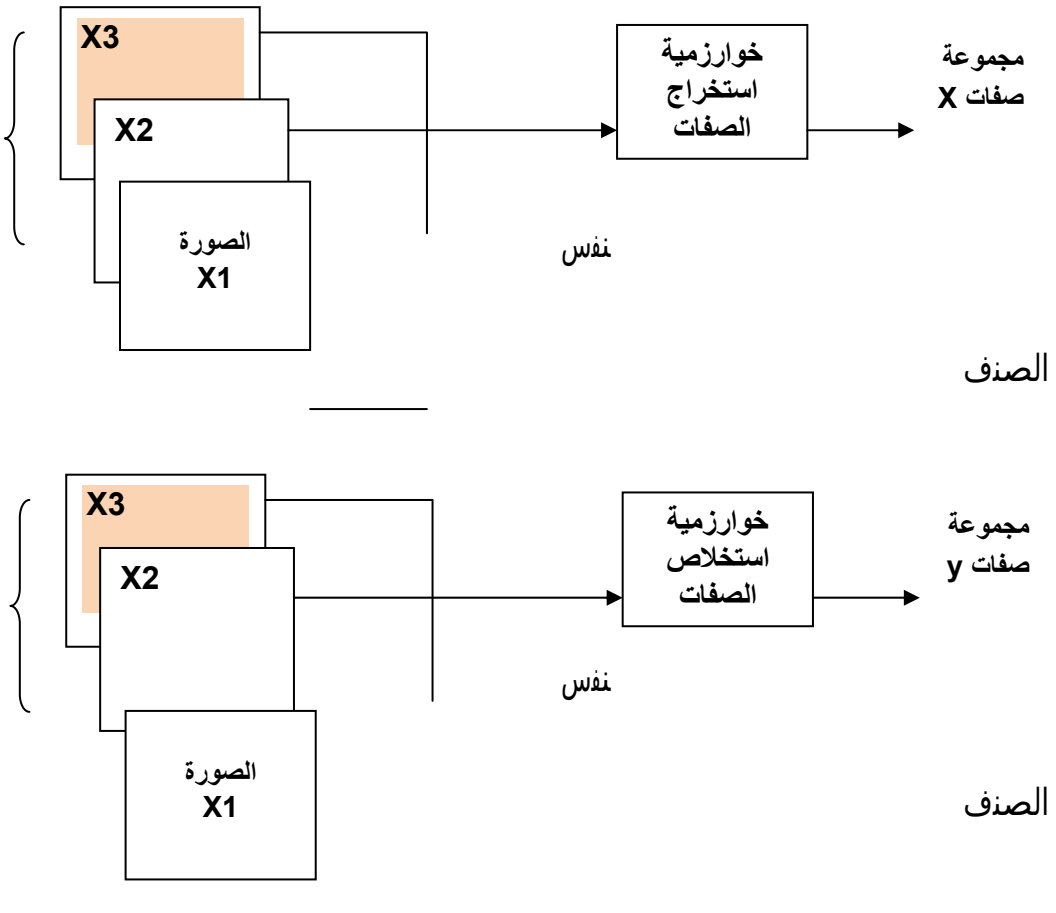
الشكل(4) مرحلة التصنيف

تتم عملية المطابقة بخزن الصورة الداخلة على شكل مصفوفة وتقارن مع القوالب الموجودة في الجهاز pixel by pixel وتعطي قيمة للمقارنة .

تعتبر هذه الطريقة طريقة سهلة، الصعوبة الوحيدة في هذه الطريقة هي الاختيار الجيد للقوالب من كل صنف فضلا عن تحديد معايير المقارنة وخصوصاً لو كانت الصورة الداخلة تحمل تشوهات [11].

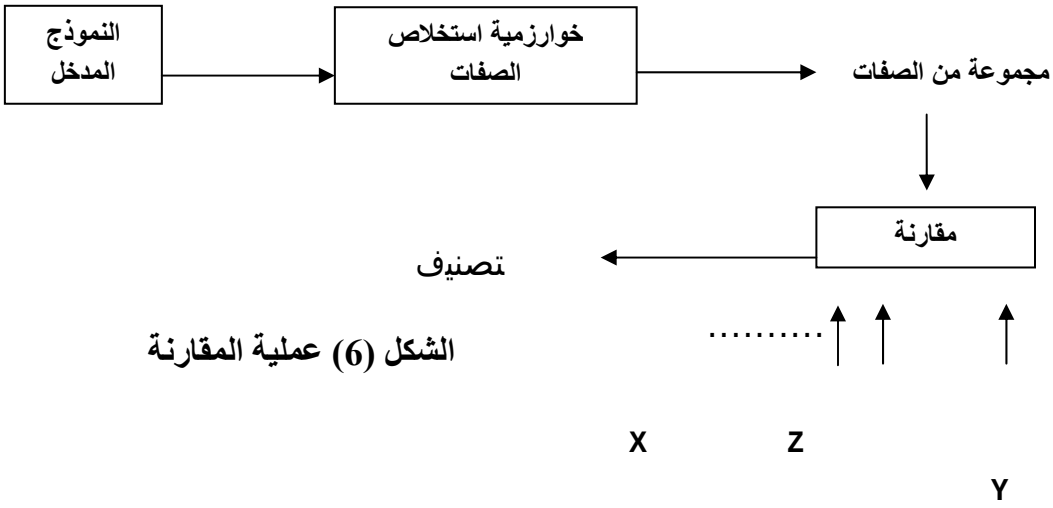
2-5 الطريقة الإحصائية

في هذه الطريقة، يوصّف كل pattern بواسطة مجموعة من الخصائص set of features التي من الممكن أن نعبر عنها بقيم حقيقية. في مرحلة التعلم: يقدّم كل نمط pattern كمتجه من الخصائصvector feature كما موضح في الشكل(5) [11] :



الشكل(5) يوضح كل نمط كمتجه من الخصائص

أما في مرحلة التعرف أو التمييز أو التصنيف، فهذه عادة تتم عن طريق تقسيم مساحة الصورة إلى مناطق مجزأة، كل منطقة تقارن مع صنف كما موضح في الشكل(6) [11] :



الشكل (6) عملية المقارنة

6- تمييز الحرف الضوئي (Optical Character Recognition)

OCR وهو تقانة إدخال النصوص إلى الحاسوب بواسطة قارئ الوثائق. في البداية، يعطي الماسحة الضوئية صورة رقمية للنص المراد تمييزه، ثمّ يستخدم برنامجاً يميز الحروف المعرفة المخزنة حول شكل الأحرف منفردة، ليحوّل الصورة الرقمية إلى مجموعة رموز داخلية يمكن تخزينها ومعالجتها في الحاسوب. تتطلب أنظمة تمييز الحروف ضوئياً OCR من حيث المبدأ أشكالاً خاصة للحروف.[7]

7- نظام تمييز الحروف السبئية باستخدام الخوارزمية الجينية

سيتم شرح الية تصميم وعمل نظام تمييز الحروف السبئية بالاعتماد على احد اساليب الذكاء الصناعي المتمثل بتقنية الخوارزمية الجينية.

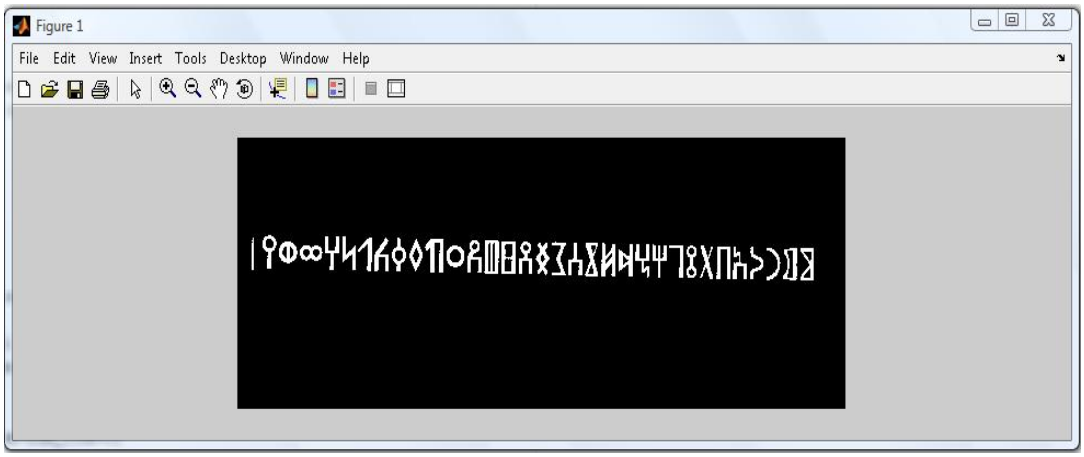
1-7 عملية تحضير بيانات الإدخال:

تعد عملية تجهيز البيانات من العمليات الاساسية في نظام تمييز الحروف وذلك لان النظام يعتمد على دقة البيانات المدخلة إليه في مرحلة تكوين قاعدة البيانات الخاصة به أو في مرحلة فحص النظام، اذ أن عملية تكوين قاعدة البيانات ومرحلة تكوين البيانات المستخدمة في مرحلة فحص النظام تكون متشابهة.

فضلا عن لذلك فان النظام يحتاج الى استخدام طريقة لادخال البيانات ويتم ذلك باستخدام الماسح الضوئي اذ يعتبر احد وسائل الإدخال في الحاسوب اذ يعمل على ادخال

الصورة أو الوثيقة أو المخططات التي تكون بشكل ورقة الى صورة رقمية تخزن بداخل الحاسبة ونستطيع التعامل معها والتعديل عليها. تم استخدام الماسح الضوئي لإدخال الصور المستعملة في تنفيذ برنامج المشروع وهما صورة الحروف السبئية وصورة النص المفحوص. وبعد مسح الصورة باستخدام جهاز الماسح الضوئي تم خزنها بصيغة ملف الصورة على شكل bmp. وذلك لكي يستطيع النظام التعامل معها واما ابعاد الصورة فأنها تكون غير محددة وذلك لان برنامج النظام قد تمت برمجته على استيعاب أي حجم من الأبعاد ولكن يجب الانتباه الى أن الصورة كلما كانت كبيرة الأبعاد فان ذلك يسبب بطئا في عملية تحضير البيانات فيجب مراعاة حجم الصورة في عملية المسح.

ان الصورة الثنائية التي يكون امتدادها (bmp.) تدعم الصورة التي تتكون من اللون الابيض والاسود اذ إن الصورة التي تم استخدامها تتكون فقط من الابيض والاسود. يجب عكس ألوان الصورة لان قيمة اللون الأسود تكون (0) واللون الابيض تكون قيمته (1) كما في الشكل (7):

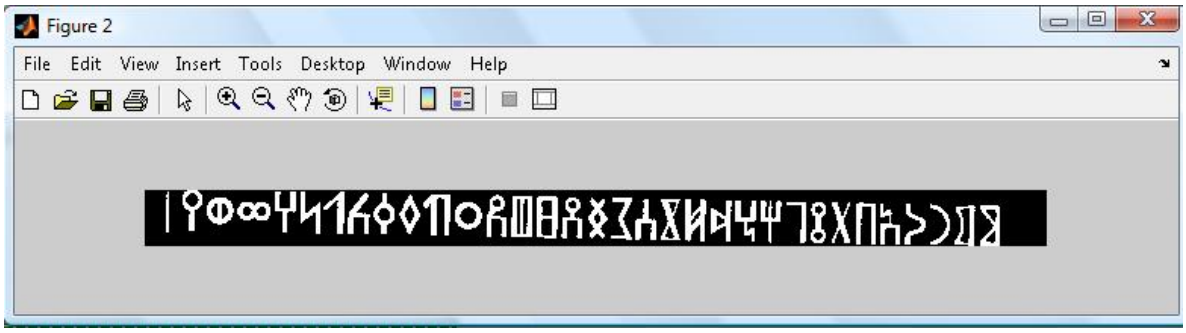


الشكل(7) عملية عكس الالوان

وتقسم عملية تحضير البيانات الى ثلاث مراحل اساسية قبل دخول البيانات الى مرحلة التمييز باستخدام الخوارزمية الجينية وهي كما يأتي:

1.1.7 مرحلة تقطيع السطور:

إن عملية تقطيع السطور يقصد بها عملية تقطيع صورة الوثيقة الى سطور وتحويلها الى مصفوفة رقمية لكل سطر على حدة .لقد تم ذلك باستخدام الدالة (im2bm) التي تعمل على تحويل الصورة من أي نوع من أنواع bmp الى النوع الثنائي أي تتكون بيانات الصورة من ('0','1') وذلك بعد تحديد قيمة الحد الفاصل وهي القيمة التي تستخدمها الدالة للفرز بين اللون الابيض واللون الأسود. ولقد قمنا ببناء الدالة (Segmentation) التي تعمل على استدعاء الصورة وإجراء عملية التقطيع الى سطور وخزنها في ملف (information.mat). والشكل التالي يوضح عملية تقطيع السطور.

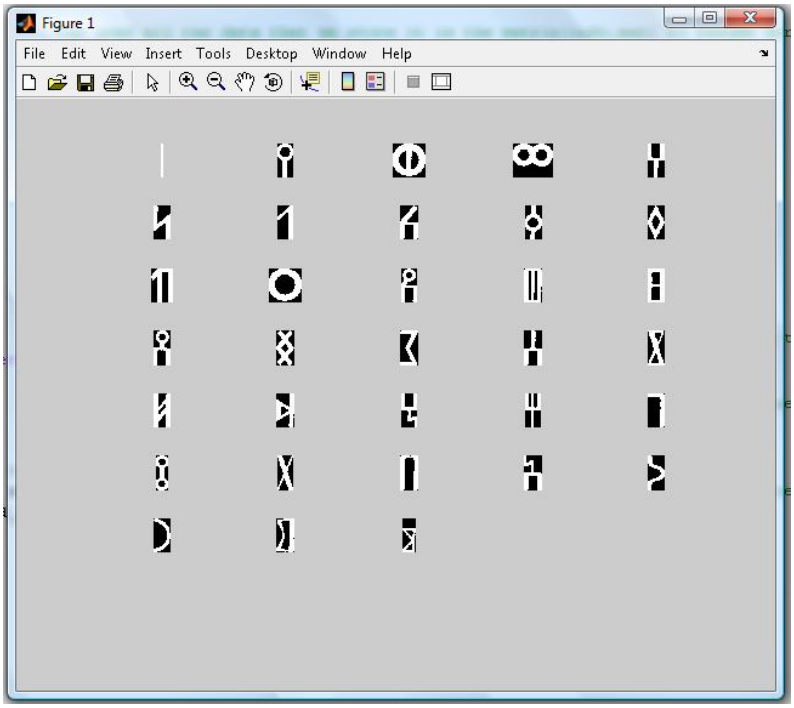


الشكل(8) عملية تقطيع السطور

2.1.7 مرحلة تقطيع الحروف:

في هذه المرحلة التي تلي عملية تقطيع السطور يتم تقطيع السطور التي تم الحصول عليها من المرحلة السابقة الى أحرف وبانتهاء هذه المرحلة تكون لدينا صورة مستقلة لكل حرف.

ان الدالة (Scan_hir) تعمل على مسح للصورة افقياً من اليسار الى اليمين وعند وجود اي نقطة مضيئة سوف تعتبر بداية اعلى السطر الذي يحوي الحرف ويتم خزن رقم السطر ثم تنتقل الى السطر الذي يليه وهكذا الى ان يأتي سطر فارغ فتخزن قيمته ويعتبر ماموجود بين القيمتين المخزونتين بمثابة سطر كامل فيتم نقله الى مصفوفة ثنائية. وبنفس الاسلوب يتم فحص الاحرف ليتم استخلاص الخواص لكل حرف بطرائق احصائية متنوعة. والشكل التالي يوضح عملية التقطيع الى احرف.



الشكل(9) عملية التقطيع الى احرف

3.1.7 تكوين صفات من صورة الحرف:

إن عملية تكوين صفات أو استخلاص الصفات من صورة الحرف قد تمت بالاعتماد على العزوم الثابتة والتي تعتمد كما هو الحال في الخاصية العشوائية على شدة إضاءة الصورة (المستويات الرمادية) وعلى الموقع الجغرافي للصورة،حيث يستفاد من هذه الخاصية في تمييز الحرف السبئي (المسند) لان الصورة المستخدمة هي ثنائية أي بيانات الصورة تكون ("0" و"1")، ألا أنها تعبر عن الأشكال التركيبية الموجودة في الصورة بشكل أفضل وذلك بسبب الطبيعة الرياضية المستخدمة لحسابها حيث أنها تمثل الترابط غير الخطي لعزوم المرتبة الثانية والثالثة التي تعتبر مقياساً لدرجة تباين الصورة (variance) ودرجة الالتواء (skewing) أي عدم تناظر توزيع الأشكال التركيبية او الهندسية في الصورة.وقد اعتمدت بشكل رئيسي في منظومات تمييز الحروف وكذلك في مجال تمييز صور الأشكال الهندسية المنتظمة. ومن مزايا هذه الخاصية انها تبقى مستقرة (ثابتة) عند تغيير المقياس أو دوران أو تزحيف الصورة بأية زاوية ولكن من مساوئ هذه الخاصية أنها تتأثر باختلاف ظروف الإضاءة أثناء التصوير ومن ثم تتأثر باختلاف شدة تباين الصورة ولكن يمكن الحد من هذه المشكلة بطريقة المعالجة الأولية بتطبيق تباين لصورتي المرجع و الاختبار قبل عملية التمييز باستخدام إحدى طرائق التباين(contrast matching method) ويمكن حساب هذه العزوم من دالة شدة الإضاءة

f(x,y) وإحداثيات وحدات الصورة ثنائية الأبعاد (x,y). فالدالة مستمرة ثنائية الأبعاد f(x,y) يمكن تمثيل العزم المنتظم (m) بأنه من المرتبة (p+q)[22] .

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \qquad \ldots\ldots\ldots\ldots(1)$$

إذ أن

P,q=0,1,2,……

تمثل العزم المنتظم، ومرتبة العزم المنتظم تساوي (p+q) .

اعتمد (Papoulis) على نظرية الوحدانية (uniqueness)[7] والمتضمنة على انه إذا كانت f(x,y) مستمرة كقطع (pieces wise continuous) ولها قيم غير الصفر فقط في الجزء المحدد في المستوي (x,y)، تكون العزوم الثابتة من جميع الدرجات للعزوم (العزم الاول، الثاني، ....الخ). ويمكن أن يعبر عن العزوم المنتظمة رياضياً كما يأتي[22] :

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y) dxdy \qquad \ldots\ldots\ldots\ldots(2)$$

$$\bar{x} = \frac{m_{10}}{m_{00}} \qquad \ldots\ldots\ldots\ldots(3)$$

$$\bar{y} = \frac{m_{10}}{m_{00}} \qquad \ldots\ldots\ldots\ldots(4)$$

حيث أن :

$\bar{x}$:تمثل مركز الصورة القياسي باتجاه محور x.

$\bar{y}$: تمثل مركز الصورة القياسي باتجاه محور y.

m_{00} : تمثل معدل قيم الوحدات الصورية .

ان النسبة بين قيم (x,y) عند تغير المقياس تكون ثابتة وكذلك النسبة بين كل من $(x-\bar{x})$ و$(y-\bar{y})$ قبل التدوير وبعد التدوير هي ثابتة لجميع وحدات الصورة ونفس الاسلوب يتبع عند الإزاحة.

ولدالة رقمية تصبح المعادلة كما يأتي[21] :

$$\mu_{pq} = \sum_{m}^{p} \sum_{n}^{q} \binom{p}{m} \binom{q}{n} (\bar{x})^{(p-m)} (-\bar{y})^{(q-n)} m_{mn} \qquad \ldots\ldots\ldots\ldots(5)$$

وبتطبيق المعادلة (5) يمكن حساب العزوم المركزية للمراتب الثلاثة (الأولى،الثانية،الثالثة) كما يأتي:[21]

$$\left.\begin{aligned}
\mu_{01} &= \sum_x \sum_y (x-\bar{x})^0 (y-\bar{y})^1 \ f(x,y) = m_{01} - \frac{m_{01}}{m_{00}} (m_{00}) = 0 \\
\mu_{10} &= \sum_x \sum_y (x-\bar{x})^1 (y-\bar{y})^0 \ f(x,y) = m_{10} - \frac{m_{10}}{m_{00}} (m_{00}) = 0 \\
\mu_{20} &= \sum_x \sum_y (x-\bar{x})^2 (y-\bar{y})^0 f(x,y) = m_{20} - \frac{2m_{01}^2}{m_{00}} + \frac{m_{10}^2}{m_{00}} = m_{20} - \frac{m_{10}^2}{m_{00}} \\
\mu_{02} &= \sum_x \sum_y (x-\bar{x})^0 (y-\bar{y})^2 \ f(x,y) = m_{02} - \frac{m_{01}^2}{m_{00}}, \\
\mu_{30} &= \sum_x \sum_y (x-\bar{x})^3 (y-\bar{y})^0 \ f(x,y) = m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2 m_{10}, \\
\mu_{12} &= \sum_x \sum_y (x-\bar{x})^1 (y-\bar{y})^2 \ f(x,y) = m_{12} - 2\bar{y}m_{11} + \bar{x}m_{02} + 2\bar{y}^2 m_{10}, \\
\mu_{21} &= \sum_x \sum_y (x-\bar{x})^2 (y-\bar{y})^1 \ f(x,y) = m_{21} - 2\bar{x}m_{11} + \bar{y}m_{02} + 2\bar{x}^2 m_{10}, \\
\mu_{03} &= \sum_x \sum_y (x-\bar{x})^0 (y-\bar{y})^3 \ f(x,y) = m_{03} - 3\bar{y}m_{20} + 2\bar{y}^2 m_{10},
\end{aligned}\right\} (6)$$

بايجاز :

$$\left.\begin{aligned}
\mu_{00} &= \mu_{00}, \\
\mu_{10} &= 0, \\
\mu_{01} &= 0, \\
\mu_{20} &= m_{20} - \bar{x}m_{10}, \\
\mu_{02} &= m_{02} - \bar{y}m_{01}, \\
\mu_{11} &= m_{11} - \bar{y}m_{10}, \\
\mu_{30} &= m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2 m_{10}, \\
\mu_{12} &= m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2 m_{10}, \\
\mu_{21} &= m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2 m_{01}, \\
\mu_{03} &= m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2 m_{01},
\end{aligned}\right\} (7)$$

ويمكن إيجاد العزوم المركزية القياسية كما يأتي:-

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{y}} \qquad \text{.....(8)}$$

اذ η تمثل قيمة العزم المركزي القياسي ذي المرتبة(p+q)

$$y = 0.5\times(p+q)+1 \qquad \text{.....(9)}$$

اذ أن

$p+q = 2,3,\ldots\ldots$

يمكن اشتقاق سبعة عزوم ثابتة (invariant moments) من العزمين الثاني والثالث [19]،والتي يعبر عنها بالمعادلات الآتية [22]:

$$\left.\begin{aligned}
&\phi_1 = \eta_{20}+\eta_{02},\\
&\phi_2 = (\eta_{20}+\eta_{02})^2+4\eta_{11}^2,\\
&\phi_3 = (\eta_{30}-3\eta_{12})^2+(3\eta_{21}-\eta_{03})^2,\\
&\phi_4 = (\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2,\\
&\phi_5 = (\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})\left[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2\right]+\\
&\qquad (3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})\left[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2\right]\\
&\phi_6 = (\eta_{20}-\eta_{02})\left[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2\right]+4\eta_{11}(\eta_{30}-\eta_{03})(\eta_{21}+\eta_{03}),\\
&\phi_7 = (3\eta_{21}-3\eta_{03})(\eta_{30}+\eta_{12})\left[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2\right]+\\
&\qquad (3\eta_{12}-\eta_{30})(\eta_{21}+\eta_{03})\left[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2\right]
\end{aligned}\right\} \quad (10)$$

ϕ تمثل قيمة العزم الثابت وتأخذ قيماً موجبة وسالبة وحسب طبيعة الصورة.

ونظراً لكون المديات الديناميكية لهذه العزوم كبيرة ،ففي التطبيقات العلمية تأخذ القيم اللوغاريتمية لها وكالاتي[22] :-

$$\psi(i) = \log/\phi(i)/ \qquad \text{.....(11)}$$

اذ أن i تشير إلى تسلسل دالة العزم (i=1,2,3,.....,7) وتؤخذ القيمة المطلقة($/\phi(i)/$) تحسباً للقيم السالبة التي ليس لها لوغاريتمية [19]. اذ تم تكوين العزوم الثابتة لكل حرف

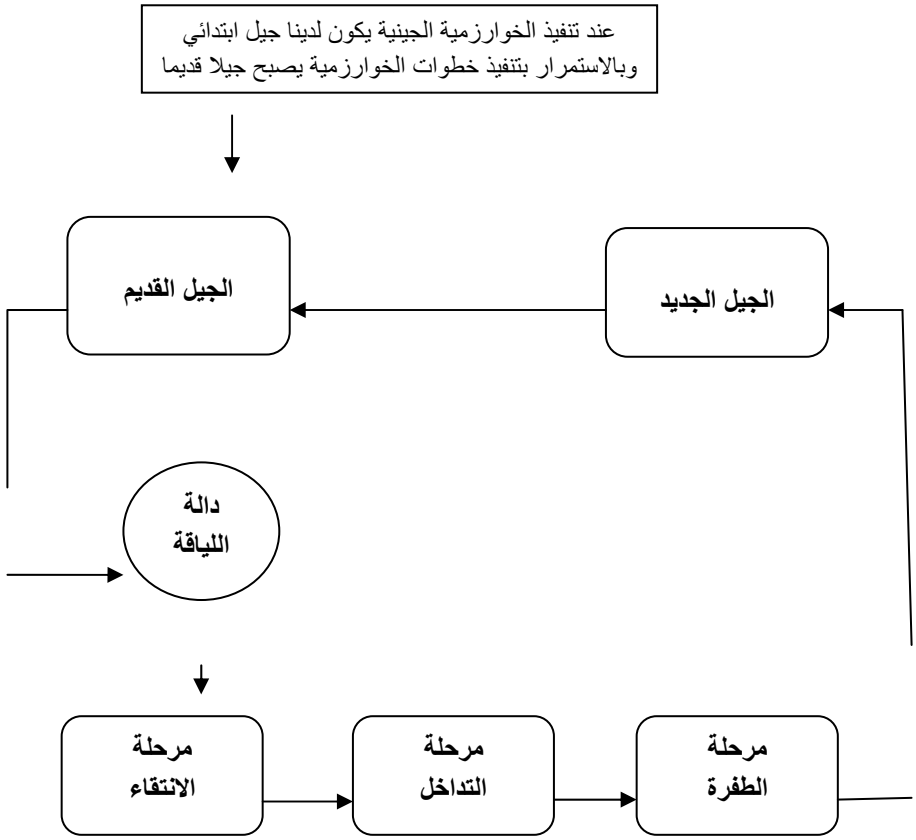
وخزنها في الملف (simo.mat) حيث تستخدم هذه العزوم كجيل ابتدائي في الخوارزمية الجينية.وكما في الجدول التالي.

الحرف	M1	M2	M3	M4	M5	M6	M7
ħ	-0.7646	-4.1118	-3.8661	-4.7934	-2.1163	-8.1948	-12.9410
∏	-0.7338	-10.5647	-7.7598	-8.7704	-3.5947	-9.0486	-17.1078
χ	-0.6780	-4.7106	-7.3025	-8.2787	-4.5298	-10.6160	-19.7774

الجدول(2)هيئة بيانات النظام المستخدمة في عملية التمييز

2-7 استخدام الخوارزمية الجينية في عملية التمييز:

بصورة عامة تتكون الية عمل الخوارزمية الجينية من ثلاث مراحل وهي مرحلة الانتقاء (Selection) ومرحلة التداخل (Crossover) ومرحلة الطفرة (Mutation) ولكن قبل الدخول الى هذه المراحل الثلاثة يجب تكوين الجيل الابتدائي (Initialize population) والشكل (10) يوضح الية عمل الخوارزمية الجينية[14] .



الشكل(10) الية عمل الخوارزمية الجينية

1.2.7 تكوين الجيل الابتدائي:

يعد تكوين الجيل الاول من المراحل المهمة في الخوارزمية الجينية حيث يتم توليد هذا الجيل بالاعتماد على نوع المسالة المراد حلها اذ ان بعض المسائل يتم توليد أرقام عشوائية وفي مسائل أخرى يتم تكوين الجيل الاول وهو إخراج عملية معينة وهكذا. في هذا البحث تم توليد الجيل الاول بأخذ العزوم الثابتة لجميع الاحرف السبئية . حيث لكل حرف مجهول يتم تكوين الجيل الاول بهذه الطريقة وثم يدخل على مراحل الخوارزمية الجينية.

2.2.7 مرحلة الانتقاء (Selection):

بعد عملية تكوين الجيل الابتدائي أو الجيل الاول تتم عملية اختيار احد الأبناء عن طريق استخدام دالة الهدف (Fitness Function) وان دالة اللياقة المستخدمة في البحث هي دالة الفرق الاقليدي لادنى مسافة (**Euclidean Minimum Distance**) كما في المعادلة الاتية [8] :

$$fitness=\sqrt{(|x1-x2|)^2+(|y1-y2|)^2}$$

اذ إن دالة الهدف تعمل على اختيار اكبر قيمة من الجيل الاول ومن ثم نختار قيمة عشوائية أخرى وذلك لتكوين الآباء (Parent Chromosome). وبعد انتهاء هذه العملية تبدأ عملية التداخل الابدالي.

3.2.7 مرحلة التداخل الابدالي (Crossover):

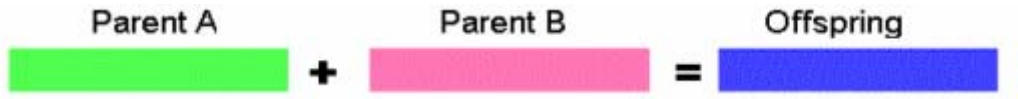
تتمثل هذه العملية بإجراء تبديل بين مقطعي الأبوين المنتخبين لغرض تكوين المقطع الجديد,وتحدد مواقع التبادل بين المقطعين إما بصورة عشوائية أو بالاعتماد على طريقة تتناسب مع صيغة المقطع وعلى خصوصية المسألة المراد تطبيقها ومن المحتمل إن يعطي المقطع الجديد حلا أفضل من المقاطع التي تكون منها وهناك عدة صيغ من عملية التداخل الابدالي وهو التداخل الحسابي. في هذا النوع من التداخل يتم ربط اثنين من الكروموسومات خطيا لإنتاج نوعين جديدين من الأبناء وحسب المعادلة الآتية:

Offspring1=a*parent1+ (1-a)*parent2

Offspring2=(1-a)*parent1+a*parent2

اذ ان:

a عدد عشوائي قيمته محصورة بين0 و1 ويتم اختياره قبل كل عملية تداخل.والشكل (11) يوضح التداخل الحسابي.[9]



الشكل(11) التداخل الحسابي[7]

4.2.7 مرحلة الطفرة (Mutation):

تتمثل هذه العملية بإجراء تغيير أو تبديل بين قيم محددة ضمن المقطع نفسه لتكوين مقاطع وإدخالها لتعطي حلولا جديدة (وراثيا تعني ادخال أفراد يمتلكون صفات جديدة الى الجيل الأتي أو الجيل اللاحق),وهي مقاطع لم يسبق تكوينها في الأجيال السابقة,وذلك بهدف توسيع

حيز الحلول الممكنة والمتمثل بتكوين اكبر عدد من المقاطع المختلفة ضمن الجيل إذ انه كلما كانت مقاطع الأجيال أكثر اختلافا,ساعد ذلك في إيجاد حلول إضافية مما يؤثر بشكل ايجابي في الحصول على حل أكثر اقترابا من الحل الأمثل. ولقد تم استخدام طفرة الاضافة أو الطرح (Adding or Subtracting) لإجراء الطفرة على الأبناء. وهي إضافة أو الطرح قيم معينة الى الأبناء المختارين.إذ تمت اضافة (0.11) الى الابن الاول وطرح (0.13) من الابن الثاني. اذ انه في كل عملية تمييز يتم تكوين 10 اجيال .والمثال الاتي يوضح عملية الطفرة (الاضافة والطرح) [9].

(5.55 **4.22 2.73** 5.68 1.29) → (5.55 **4.11 2.86** 5.68 1.29)

3-7 نظام التمييز باستخدام الخوارزمية الجينية:

تم استخدام الخوارزمية الجينية في عمل نظام التمييز وكما يأتي:

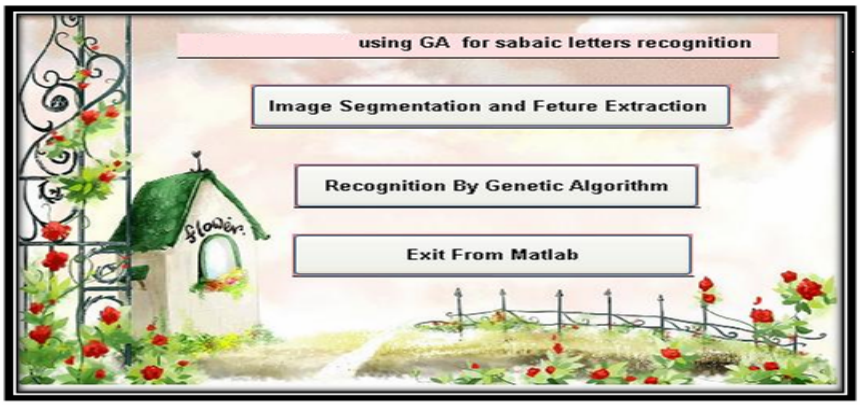
1- توليد جيل ابتدائي يحوي 33 كروموسوم (اي بعدد الاحرف السبئية) وكل كروموسوم يحوي على 7 جينات والتي تمثل قيم العزوم السبعة اللااعتمادية لكل حرف.
2- قراءة الملف الذي يحوي النص المجهول المراد تمييزه والذي يحوي كلمات سبئية، اذ يتم تقطيع الاحرف وحساب العزوم البعة لكل منها ووضعها بهيئة كروموسوم لكل حرف لغرض التمييز.
3- حساب دالة اللياقة لكل كروموسوم في الجيل الابتدائي باستخدام دالة المسافة الاقليدية مابين كل كروموسوم وباقي الكروموسومات واخذ المعدل للقيم.
4- اختيار كروموسومين بطريقة Tournament selection فيكون الكروموسوم الاول باعلى قيمة للياقة والاخير باقل قيمة للياقة.
5- عمل تزاوج بطريقة Arithmetic crossover وبنسبة احتمال 100%.
6- عمل طفرة بطريقة addsub mutation وبنسبة احتمال 0.001.
7- حساب دالة اللياقة للاجيال الناتجة عن العملية والتي تمثل المسافة الاقليدية بين كل جيل ناتج وبين الكروموسوم المجهول والمراد معرفته والذي تمت قرائته في الخطوة (2).
8- اجراء عملية استبدال حيث يستخدم الجيل المتولد مرة اخرى في الخوارزمية.

9- اجراء عملية الفحص اذا تحقق شرط التوقف والوصول الى الحل تتوقف الخوارزمية ويتم اقرار الحرف باقل قيمة لياقة للكروموسومات للجيل المتوقف عنده ويتم خزنه في ملف نصي ولاتعاد الخطوات4-9.

10- اختيار الحرف المجهول التالي وتعاد الخطوات من 3-9 وتستمر الى ان ينتهي الملف المراد تمييزه.

7-3-1 تنفيذ النظام بلغة Matlab:

يتكون برنامج نظام التمييز للحروف السبئية من واجهة مصممة باستخدام الواجهة الرسومية للمستخدم Graphical User Interface (GUI) التي تعمل على التحكم بمراحل تنفيذ البرنامج لبدء تنفيذ برامج النظام والشكل (12) يوضح واجهة تنفيذ نظام التمييز.

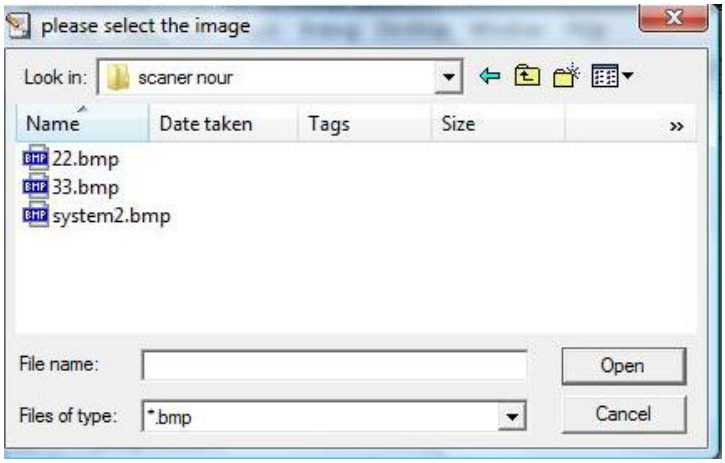


الشكل(12) واجهة تنفيذ نظام التمييز

كما موضح في الشكل أعلاه هناك ثلاثة اختيارات رئيسية تتم من خلالها تنفيذ البرنامج وهي كما يأتي:

7.3.1. 1 Segmentation and Feature Extraction:

عند الضغط على هذا (button) تظهر نافذة استدعاء صورة الوثيقة المراد تمييزها كما في الشكل(13).

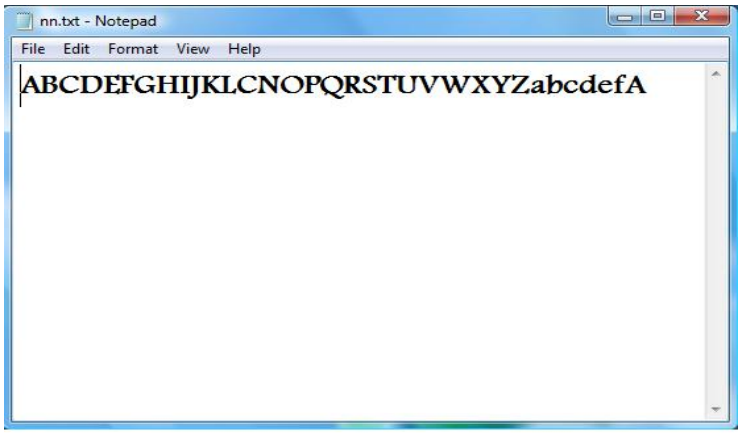


الشكل(13) استدعاء ملف صورة الوثيقة

بعد جراء عملية الاستدعاء يجري تقطيع الصورة وتكوين صفات لكل حرف موجود في الصورة حيث يتم حساب العزوم الثابتة لكل حرف وتخزن في الملف (simo.mat) لكي يتم استخدامها في مرحلة التمييز.

2.1.3.7 Recognition by using genetic algorithm:

عند الضغط على هذا(button) سوف تطبق تقنية الخوارزمية الجينية في تمييز الحروف السبئية وتولد مصفوفة بالحروف المميزة بشكل حروف انكليزية(Capital & Small English Letters) وذلك لعدم امكانية ماتلاب على عرض الحروف السبئية وقمنا بخزن النتائج على شكل ملف نصي(txt.) وذلك لغرض استخدامها في لغة (Visual Basic.Net) والشكل (14) يوضح الملف النصي .



الشكل(14) النتائج

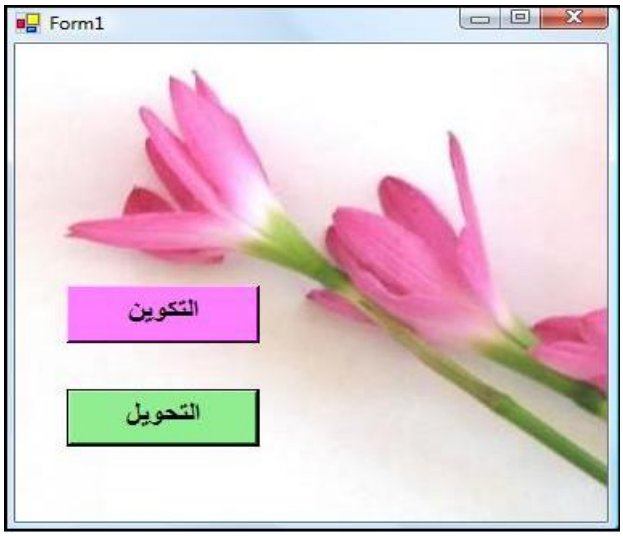
3.1.3.7 Exit from the system:

خروج من النظام.

7-3-2 اظهار الحروف السبئية المميزة على شاشة المستخدم:

كما ذكرنا سابقاً ان الحروف السبئية تم تمييزها باستخدام لغة Matlab وعندما تعذر اظهار الشكل النهائي للحرف باستخدام اللغة المذكورة فقد تمت كتابة البرامج الخاصة باظهار الحروف باستخدام لغة Visual Basic.Net.

يتكون برنامج نظام التمييز للحروف السبئية من واجهة مصممة باستخدام (Visual Basic.Net) والتي تحتوي على خيارين والشكل (15) يوضح الواجهة المستخدمة في البرنامج باستخدام (Visual Basic.Net).

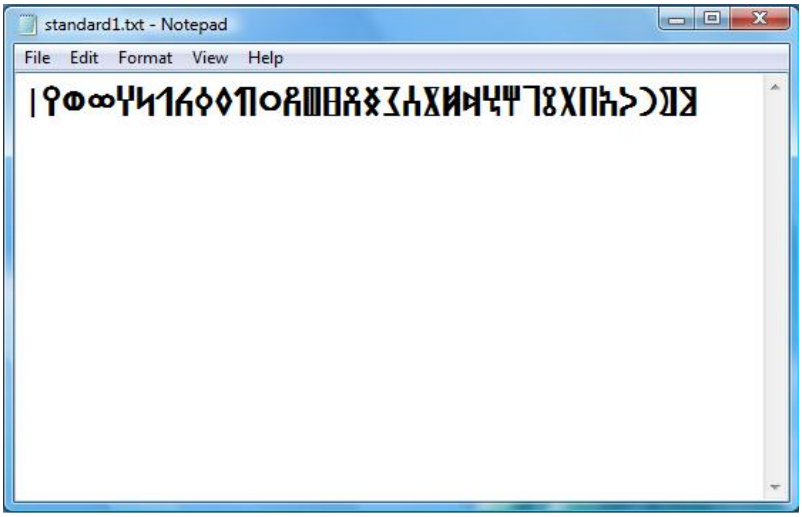


الشكل(15) واجهة (visual basic.net)

نلاحظ في الشكل (15) ان هناك خيارين رئيسيين تتم من خلالهما عملية التكوين والتحويل وهما كما يأتي:

7.3.2.1 التكوين:

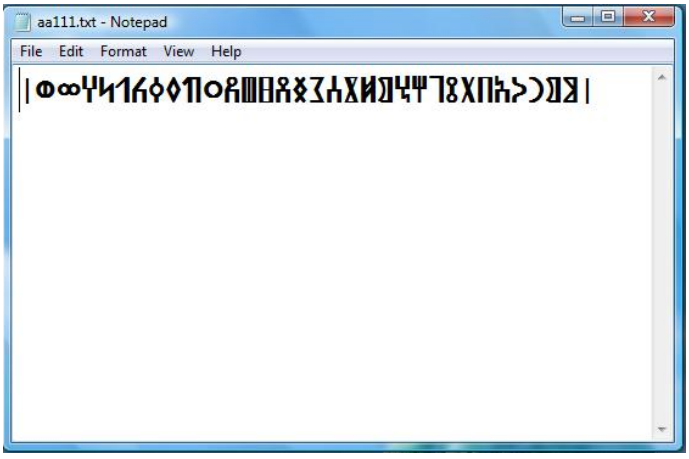
هذا الاختيار يقوم بتكوين ملف نصي يحتوي على الحروف السبئية وكما موضح في الشكل (16).



الشكل (16) الحروف السبئية

2.2.3.7 التحويل:

هذا الاختيار يقوم بتحويل محتويات الملف النصي للنتائج الموضحة في الشكل (17) الى ما يقابله من الحروف السبئية الموضحة في الشكل اعلاه حيث ان الشكل التالي يمثل النتائج النهائية لعملية التمييز.



الشكل(17) النتائج النهائية لعملية التمييز

8- الاستنتاجات

ان النظام الحاسوبي الذي تم بناؤه بلغة ال(Matlab) يقوم بقراءة الملف الصوري للحروف السبئية حيث يتم تقطيع الملف الصوري الى الحروف السبئية للحصول على حرف منفرد ثم استخلاص صفات الحرف السبئي بالاعتماد على العزوم الثابتة اللاعتمادية والتي تنتج قيما حقيقية والتي تكون كجيل ابتدائي في الخوارزمية الجينية. وقد تم تمييز الحروف السبئية بما يعادلها من الحروف الانكليزية وخزنها في ملف نصي وذلك لعدم امكانية (Matlab) على

طباعة الاحرف السبئية ولمعالجة هذه المشكلة تم بناء الجزء المكمل للنظام الحاسوبي بلغة (Visual Basic.Net) للتعامل مع الحرف السبئي حيث يقوم باستقبال الملف النصي الناتج من لغة (Matlab) ومقارنته مع جداول تم بناؤها انفاً فتنتج لنا مايعادله من الحروف السبئية وخزنها في ملف نصي اخر وقد كانت نسبة التمييز جيدة جدا تبلغ نسبتها ما بين (85%-91%) وذلك لان الفرد الذي ينتقى عشوائيا يسبب هذا التراوح في نسبة التمييز.

وتم استخدام المعادلة التالية لحساب نسبة التمييز:

نسبة التمييز= عدد الحروف التي تم تمييزها بصورة صحيحة لكل المستندات / عدد الحروف الكلي

وقد تم استخدام 20 وثيقة وكانت نسب التمييز كالاتي:

رقم الوثيقة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
نسبة النجاح%	86 %	87 %	85 %	88 %	85 %	89 %	86 %	87 %	91 %	88 %	85 %	87 %	91 %	86 %	90 %	85 %	89 %	88 %	87 %	% 85

9- التوصيات المستقبلية

1- امكانية توسيع هذا العمل واستخدام الارقام السبئية فضلا عن الاحرف التي تم استخدامها في هذا العمل.

2- تهجين الخوارزمية الجينية بالمنطق المضبب لزيادة نسبة التمييز.

3- امكانية استخدام الشبكات العصبية بدلاً من الخوارزمية الجينية في عملية التمييز.

المصادر Reference

1. الجميلي، عامر، الكتابة السبئية، بحث قيد النشر ، جامعة الموصل.
2. الروقي، هند،خوارزمية الجينات، (2009) على الرابط:

http//knol.google.com

3. السيباني، سليمان،قصة موت الحضارة اليمنية، (2006) على الرابط:

http//www.do3n.com

4. الصباحي، نور، الخوارزميات الجينية، (2008).

http//www.schwarztiger’s weblog.com

5. العمري، هدى عبدالله باشعالم، كروموسومات جينية لتجزئة الصور.رسالة ماجستير،كلية علوم الحاسبات والرياضيات،جامعة الموصل، (2004).
6. المحمدي، قائد،خط المسند واهميته، (2008) على الرابط:
http//forums.ibb7.com/ibb35059.html
7. الموسوعة العربية، تمييز الانماط ، (2010) على الرابط:
http//www.marefa.org
8. بافقيه، محمد عبد القادر وآخرون، مختارات من النقوش اليمنية القديمة، منشورات المنظمة العربية للتربية والثقافة والعلوم، ادارة الثقافة، تونس، 1985، ص17.
9. خروفة، شهلة حازم احمد، استخدام الخوارزمية الجينية المهجنة لتصنيف صور الاقمار الصناعية، رسالة ماجستير،كلية علوم الحاسبات والرياضيات،جامعة الموصل، (2005).
10. سيد مجيد، د. جمال ومحمد صديق، شيرين"هندسة برمجية لخوارزمية الاختيار النسيلي المناعية والمهجنة بالخوارزمية الجينية لتمييز خط اليد"، بحث مقبول للنشر في مجلة التربية والعلم- كلية التربية-جامعة الموصل، رقم (2830) في 2011/10/19.
11. شبكة المنهل التعليمية،مقدمة في تمييز الانماط ومعالجة الصور، (2010) على الرابط:
http//www.c4arab.com
12. عفيفي، فوزي سالم، نشأة وتطور الكتابة الخطية العربية ودورها الثقافي والاجتماعي، الطبعة الاولى، نشر وكالة المطبوعات، الكويت 1980، ص65.
13. علي، جواد، تاريخ العرب قبل الاسلام، الطبعة الاولى، الجزء السابع، مطبعة المجمع العلمي العراقي، 1957، ص37 .
14. محمد، اسامة ياسين ، "تمييز صور الطبية للدماغ باستخدام تقنيات البايومعلوماتية"،رسالة ماجستير، جامعة الموصل، 2010.
15. هبو، احمد، الابجدية، نشأة الكتابة واشكالها عند الشعوب، الطبعة الاولى، مطابع الف باء- الاديب، الناشر دار الحور للنشر والتوزيع، سوريا-اللاذقية، 1984، ص 90.
16. ولفنسون، تاريخ اللغات السامية، الطبة الاولى، لجنة التأليف والترجمة والنشر، مطبعة الاعتماد بشارع حسن الاكبر بمصر، 1929، ص245.
17. ويكيبيديا الموسوعة الحرة،خط المسند، (2009) على الرابط:
http//www.ar.wikipedia.org

18. ويكيبيديا الموسوعة الحرة،خوارزميات وراثية،(2009) على الرابط:
http//ar.wikipedia.org

19. ويكيبيديا الموسوعة الحرة، Image Moment، (2009).
http//ar.wikipedia.org

20. AL-Dulaimi,Buthaina;Ali,Hamza, Enhanced Traveling Salesman Problem Solving by Genetic Algorithm Technique (TSPGA),world Academy of Science, Engineering and Technology ,Page 296. (2008).
21. Exavation at Ur.1922 p. 98.
22. Gonzalez, R. C. et al., Digital Image Processing, second edition.(addison-wesly problishing company, Inc.), 1987.
23. www.marefa.org/ index.php.