

د. محمد الكنهل
د. عبد الله الراجح

د. عبد الملك السلطان
د. يوسف العوهلي

عناصر العرض

✓ المقدمة.

✓ نظام برايل.

✓ استعراض للأعمال السابقة في مجال برايل.

✓ مراحل التعرف الضوئي على برايل
العربي.

✓ نتائج التجارب.

✓ الخاتمة.

المقدمة

نتيجة لتزايد استخدام نظام برايل في أكثر دول العالم وانتشار مطبوعات برايل فقد أصبحت هناك حاجة لوجود أنظمة أتمتة لها. وقد ظهرت عدة أنظمة للتعرف الضوئي على برايل لعدة لغات إلا أنه لا يوجد حتى الآن نظام للتعرف على برايل العربي.

إن الهدف من نظام التعرف الضوئي على برايل العربي هو أتمتة التعامل مع مطبوعات برايل وتحويلها إلى نص عربي.

هناك مرحلتين من المعالجة في نظام التعرف الضوئي على برايل العربي هما:

١. مرحلة تحويل مطبوعات برايل إلى صيغة رقمية يمكن التعامل معها.
٢. تحويل هذه الصيغة الرقمية والتي تمثل رموز برايل إلى نص عربي مقروء.

المقدمة

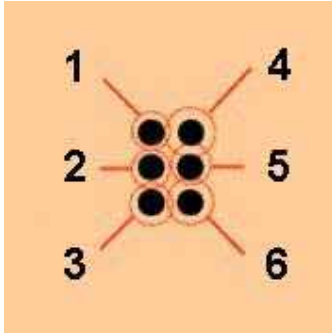
هناك العديد من الفوائد لوجود نظام التعرف الضوئي منها:

١. إمكانية الاحتفاظ بنسخ إلكترونية من مطبوعات برايل.
٢. إعادة إنتاج مطبوعات برايل حيث هناك العديد منها قديمة وموجودة في المكتبات لذا تحتاج إلى إعادة نسخ، ووجود مثل هذا النظام يسهل هذه العملية.
٣. مساعدة من ليس لديهم القدرة على التعامل مع مطبوعات برايل بترجمتها لهم، ممن لهم علاقة في ذلك.

ولكي يكون نظام التعرف الضوئي عملي يجب أن لا يأخذ وقتاً طويلاً في التعرف إضافة إلى أنه يجب أن يكون هناك نوعاً من المرونة في التعامل مع بعض المشاكل التي قد تحدث عند التقاط الصفحة إضافة إلى عوامل الإضاءة ووضعيات الصفحة نفسها.

نظام برايل

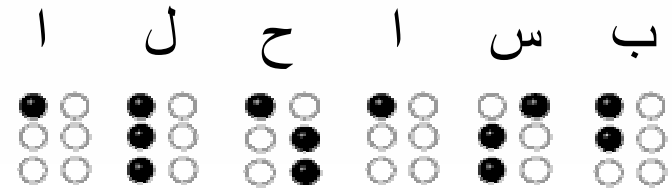
لقد ابتكر لويس برايل المولود في عام ١٨٠٩م هذا النظام، والذي كان قد فقد بصره في وقت مبكر من حياته، وأخذ هذا النظام بالانتشار على مستوى العالم لسهولة وسرعته في القراءة والكتابة. وقد أصبح اليوم نظام برايل هو نظام القراءة والكتابة المعتمد على مستوى العالم للمكفوفين إلا أنه لا يوجد حتى الآن معيار موحد لبرايل في العالم العربي.



- **ونظام برايل يعتمد على حاسة اللمس ويتكون من:**
- مجموعة من الخلايا (cells).
- كل خلية عبارة عن شبكة من ستة نقاط (dots).
- من خلال عدد النقاط وأماكنها يمكن تكوين ٦٣ خلية مختلفة.
- الجدير بالذكر أن نظام برايل يقرأ من اليسار إلى اليمين في جميع اللغات.

نظام برايل

مثال على طريقة كتابة "الحاسب" بترميز برايل:



إن نظام برايل ليس لغة وإنما هو ترميز لجميع اللغات، وكما يظهر في الشكل تمثيل الحروف العربية:

	ا	ب	ت	ث	ج	ح	خ	د	ذ	ر	
#	١	٢			٠			٤			
	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	
#										٦	
	ق	ك	ل	م	ن	هـ	و	ي			
#						٨		٩	٣	٥	٧

استعراض للأعمال السابقة في مجال برايل

إن غير المبصر أو من لديه ضعف شديد في ذلك يحتاج إلى مجموعة من الأجهزة الخاصة لكي يتفاعل مع الحاسب. ومن ذلك:

- * لوحة مفاتيح برايل.
- * السطر لإلكتروني (Braille display).
- * *تقنية التعرف الصوتي.
- * *قارئ الشاشة (Screen reader).
- * *طابعات برايل (Braille Embossers).
- * *مكبر الشاشة (Screen magnifier).
- * *برامج كتابة برايل والترجمة.

هناك نوعان من الأنظمة المتعلقة بترميز برايل وهي:

١. أتمتة المطبوعات عن طريق التعرف الضوئي عليها.
٢. معالجة الصيغة الرقمية لبرائيل إما بتحويلها إلى نص مقروء أو تحويل النص إلى برايل.

إن ما يهمنا في هذه الورقة النوع الأول وهناك عدد قليل من الأنظمة التي تقوم بالتعرف الضوئي على برايل، إلا أنه وبحسب علمنا يوجد نظام تجاري واحد في هذا المجال وهو:

استعراض للأعمال السابقة في مجال برايل

يعتمد نظام (Neovision OBR) على تقنيات عالية في معالجة الصور ومن ذلك الظل الناتج من نقاط برايل، إضافة إلى أنه يقوم بالاعتماد على وضع ورقة شفاف خاصة على مطبوعات برايل.

وهناك ميزات عديدة لهذا النظام منها:

- * دقة التعرف لديه عالية حيث فاقت ٩٩%.
- * يتعرف على مطبوعات برايل ذات الوجهين من خلال مسح واحد.
- * معالجة الورقة عندما تكون عمودية أو أفقية أو مائلة أو مقلوبة بشكل آلي.
- * معالجة مطبوعة برايل الكبيرة بمسح كل جانب على حدة ومن ثم الدمج بينهما آلياً.

إلا أن من أبرز جوانب القصور في هذا النظام أنه لا يدعم برايل العربي

مراحل التعرف الضوئي على برايل العربي

إن عملية التعرف الضوئي على برايل تقوم على نقطة جوهرية وهي تحديد أماكن النقاط (dots) عن طريق ظلها. وقد تبين لنا من التجربة أن استخدام الشفاف مع ورق برايل يزيد من وضوح الصورة خصوصاً مع الورق الأبيض كما يظهر في



الشكل التالي:
صورة لورق
برايل بدون
ورقة الشفاف



صورة لورق
برايل بورقة
الشفاف

وكما هو واضح فإن لدينا ثلاثة أجزاء (ساطعة، داكنة، ليست ساطعة ولا داكنة). ولهذا فإن الجزء العلوي من النقاط البارزة ساطع والجزء السفلي داكن بعكس النقاط الغائرة.

مراحل التعرف الضوئي على برايل العربي

أما بعد التقاط صور مطبوعات برايل فهناك تسع مراحل للتعرف عليها:

١. تحويل الصورة إلى لون رمادي (Gray

(Level

الصور الملونة تخزن في مصفوفة ذات ثلاثة أبعاد (D³). لهذا لابد من تحويلها إلى مصفوفة ذات بعدين (D²) - لون رمادي- ليسهل التعامل معها بحيث يكون تمثيل الألوان لأي بكسل (pixel) في الصورة من (٠-٢٥٥).

٢. تهذيب إطار الصورة

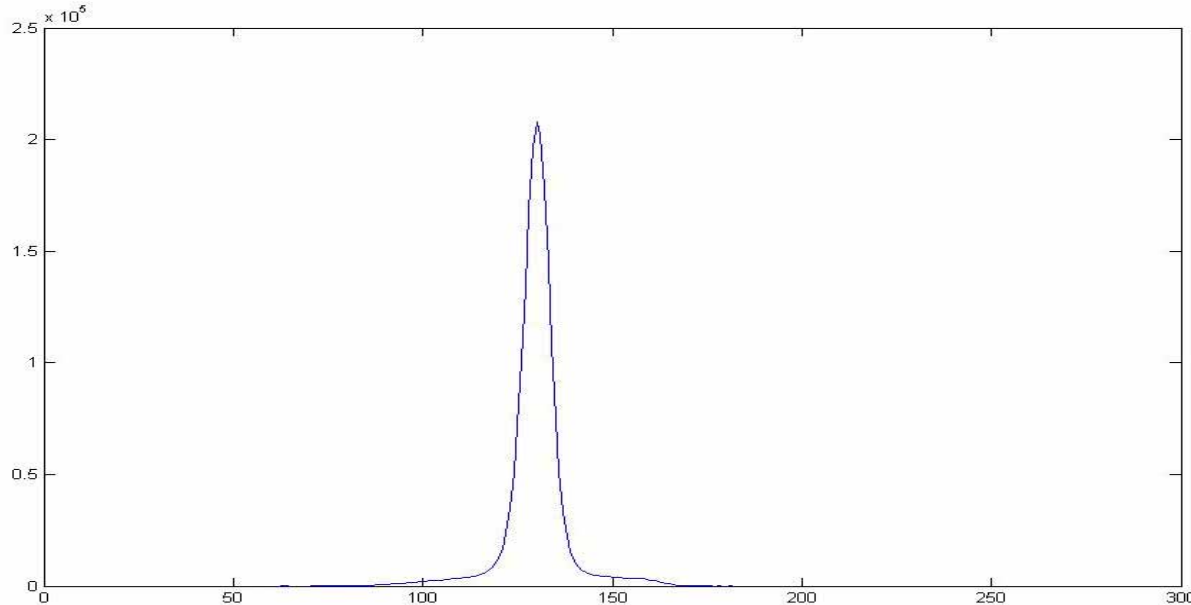
(Cropping)

أحياناً يوجد سواد أو بياض في إطارات الصفحة من الماسح الضوئي أو لون مختلف بسبب أن الشفاف لم يغط كامل الورقة، لهذا لابد من إزالة ذلك، لأن وجوده يؤثر على التعرف.

مراحل التعرف الضوئي على برايل العربي

٣. تصنيف الصورة

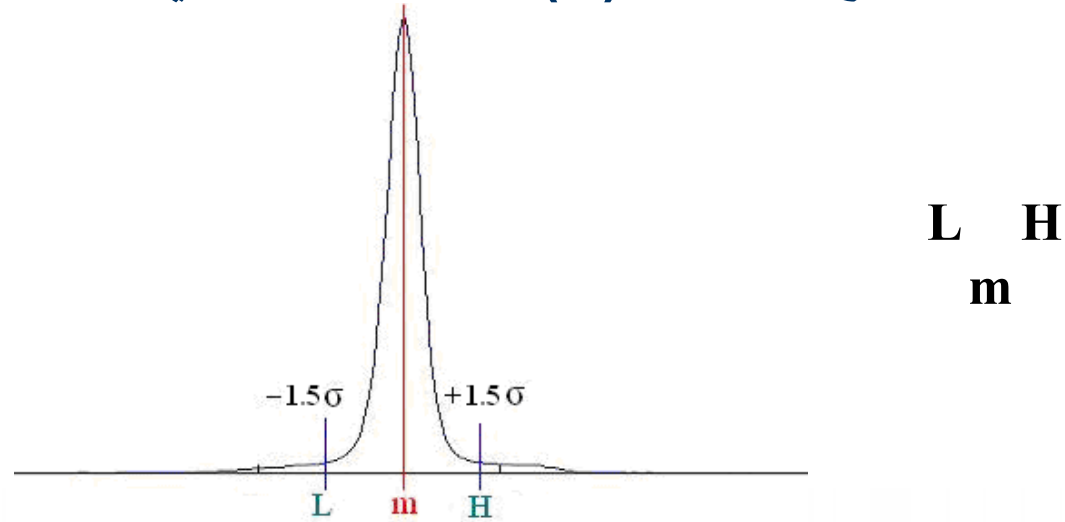
من خلال وضعنا للرسم البياني لتكرار الألوان للورقة (histogram) كما في الشكل التالي، لاحظنا أنه يظهر على شكل توزيع طبيعي (normal distribution)، لهذا فمن المقبول استخدام التوزيع الطبيعي لتصنيف الورقة



مراحل التعرف الضوئي على برايل العربي

٣. تصنيف الصورة

ومن خلال التجربة وجد أن $(\sigma \times 1,5)$ هو الحد الأمثل، حيث أن (σ) هي الانحراف المعياري، والتي من خلالها نوجد قيمة سفلى (L) و قيمة عليا (H)، بحيث يكون ما فوق (H) جزء ساطع وما دون (L) جزء داكن كما في الشكل التالي:

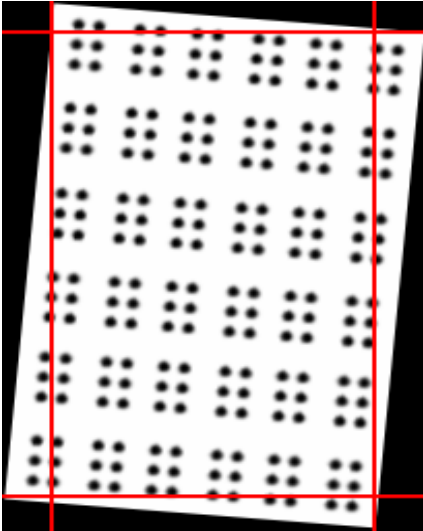


ملاحظة: تحديد معامل (σ) بقيمة ١,٥ ليس ثابتاً تماماً وإنما يعتمد على عدة متغيرات كنوع المساح الضوئي

مراحل التعرف الضوئي على برايل العربي

٤. تعديل انحراف الصورة (Deskewing)

١- عندما تكون الصورة منحرفة أثناء التقاطها لابد من معرفة درجة انحرافها لتعديلها عن طريق خوارزمية معينة، علماً بأننا نفترض أن أقصى انحراف يُتوقع حدوثه هو ٤ درجات من اليمين أو اليسار، وأن دقة التعرف على الانحراف هي (١/١٦) من الدرجة.



٢- بعد تحديد درجة الانحراف نقوم بعمل تدوير للصورة الأصل (الرمادية) في المرحلة الأولى من مراحل المعالجة ومن ثم نطبق المرحلة الثانية والثالثة مرة أخرى عليها، وذلك لأنه في حالة انحراف الصورة فإن عملية التهذيب (Cropping) في المرحلة الثانية قد تقوم بحذف بعض أجزاء من نص برايل مما يؤثر على التعرف كما في الشكل المقابل:

مراحل التعرف الضوئي على برايل العربي

٥. التعرف على أجزاء النقطة (Dot)

من خلال التجارب السابقة للتعرف على النقطة خصوصاً في حال كانت الورقة ذات وجهين (double side) كانت طريقة التعرف على أجزاء النقطة -وليس النقطة ككل- هي الأفضل.

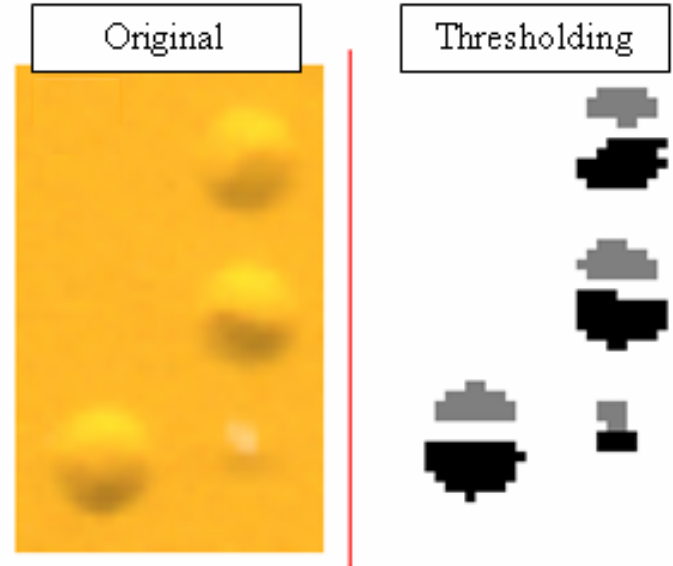
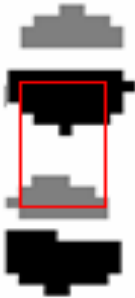
وعند عمل تصنيف لصور برايل يتبين أن أي نقطة إنما هي عبارة عن منطقة ساطعة ومنطقة داكنة. أما بالنسبة لتحديد نوع النقاط هل هي بارزة أو غائرة فذلك يعتمد على تواجد المنطقة الساطعة والداكنة، فإذا كانت المنطقة الساطعة في الأعلى والداكنة في الأسفل فهذا يدل على أن النقطة بارزة وإذا كانت عكس ذلك فهي غائرة.

ملاحظة: نود أن نشير إلى أنه ليس على الدوام تكون المنطقة الساطعة في الأعلى والداكنة في الأسفل للنقطة البارزة والعكس للغائرة، إلا أننا سنفترض ذلك للتسهيل.

مراحل التعرف الضوئي على برايل العربي

٥. التعرف على أجزاء النقطة (Dot)

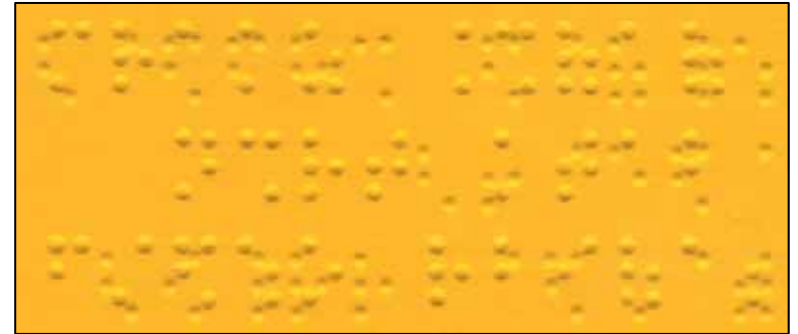
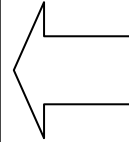
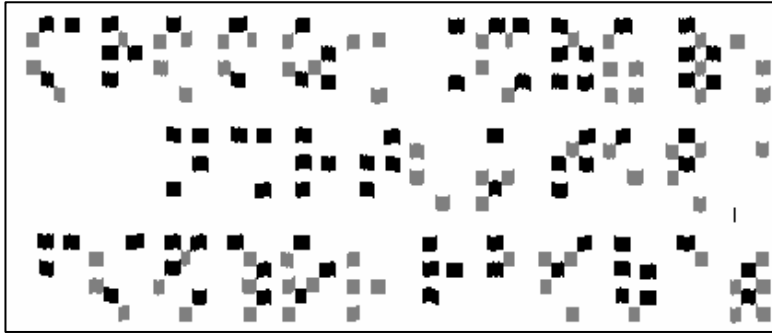
ملاحظة: خلال البحث العمودي عن أجزاء النقاط المناسب لا بد من القيام بخطوة مهمة جدا في ضبط التعرف، وهي القفز مقدار معين عن التعرف على جزء من نقط. لأنه بدون هذه الخطوة قد يتعرف على عمود غير صحيح هو في الواقع الجزء الأسفل من النقطة الأولى والجزء الأعلى من النقطة الثانية التي في الأسفل، كما في الشكل التالي:



مراحل التعرف الضوئي على برايل العربي

٥. التعرف على أجزاء النقطة (Dot)

بعد الانتهاء من البحث ستظهر لنا النقاط البارزة والغائرة كما في الشكل التالي، ومن ثم نقوم بفصلهما في مصفوفتين مختلفتين.



مراحل التعرف الضوئي على برايل العربي

٦. تصحيح التعرف على أجزاء النقطة (Dot)

خلال التعرف على أجزاء النقطة فإنه قد يحدث هناك خطأ بسبب وجود بعض الشوائب حول النقاط، ولتجاوز هذه المشكلة نقوم باستخدام مقياس عام (Global Measure) على كامل الورقة. حيث نقوم بتصنيف اعمدة الورقة كالتالي:

١. أعمدة لنقاط بارزة.

٢. أعمدة لنقاط غائرة.

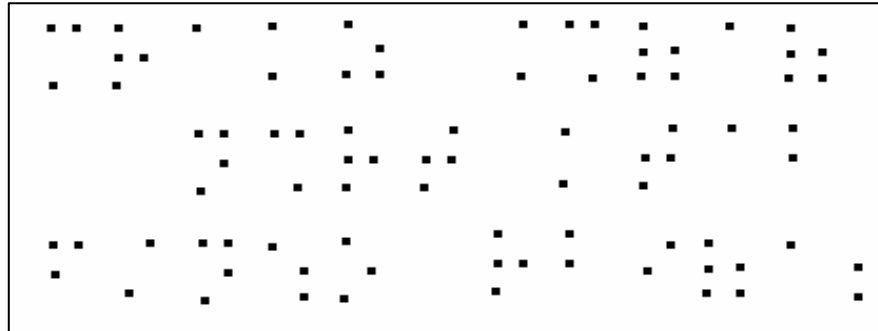
٣. أعمدة غير محددة.

وذلك لأن من خصائص ورق برايل أن عمود النقاط من نوع واحد إما نافر أو غائر. ويكون هذا التصنيف بناءً على عدد أجزاء النقاط الأكثر على طول العمود في الورقة، بحيث يكون هناك معالجة مبدئية لتصنيف الأعمدة، ومن ثم معالجة نهائية بحيث لا يكون التعرف إلا على أجزاء النقاط التي من نفس العمود، أما بالنسبة للعمود غير المحدد نوعه فيكون التعرف على كلا نوعي أجزاء النقاط البارزة والغائرة خلال البحث العمودي.

مراحل التعرف الضوئي على برايل العربي

٧. التعرف على كامل النقطة (Dot)

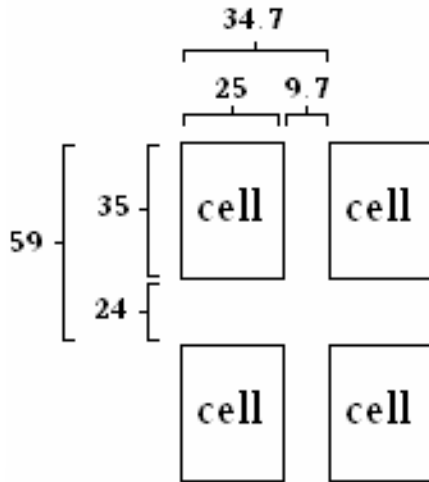
- هناك عدة خطوات للتعرف على كامل النقطة وهي:
١. النقاط المعتبرة هي التي تحوي ثلاثة أعمدة فأكثر، أما أقل من ذلك فليس بنقطة.
 ٢. ننشء مصفوفة خالية، وتكون مماثلة للمصفوفة التي تحوي أجزاء النقاط، لحفظ النتائج.
 ٣. نقوم بالبحث عن النقاط بشكل عمودي في المصفوفة من أعلى إلى أسفل بحيث أنه عند وجود ثلاثة أجزاء من نقطة، ثم نقوم بإزالة هذه المنطقة التي تحوي أجزاء النقاط لكي لا تدخل معنا في بحث آخر، وهذه الخطوة (Detect & Remove) مهمة جداً لجودة التعرف.
 ٤. بعد الانتهاء من البحث ستظهر لنا النقاط كما في الشكل التالي:



مراحل التعرف الضوئي على برايل العربي

٨. التعرف على الخلايا (Cells)

١. في هذه المرحلة نحدد المنطقة التي تحوي جميع النقاط (Dots)، حيث أنه لا يمكن أن نجد نقاط خارج هذه المنطقة. وذلك عن طريق جمع صفوف المصفوفة على حدة، وجمع أعمدة المصفوفة على حدة. ومن ثم نأخذ موقع القيمة الموجبة الأولى والأخيرة للمحورين x و y .



٢. نحدد عدد أسطر الخلايا في المنطقة وعدد الأعمدة. ومن خلال دراسة أطوال الخلايا والمسافة بينها، تم معرفة متوسط عرض السطر والعمود كما في الشكل المقابل. ولهذا فإن حساب عدد الأسطر والأعمدة يكون على النحو التالي:

$$\text{linNum} = (y\text{Max} - y\text{Min}) / 59$$

$$\text{colNum} = (x\text{Max} - x\text{Min}) / 34.7$$

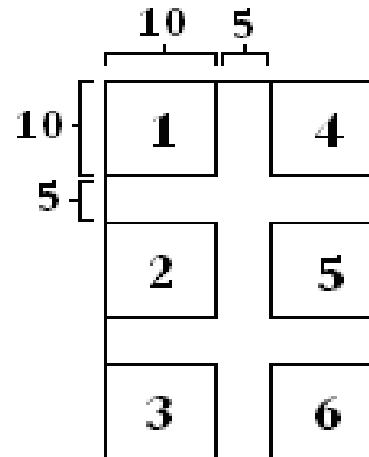
٣. بعد معرفة عدد الأسطر والأعمدة، يمكن الوصول لأي

خلية من خلال رقم السطر والعمود

مراحل التعرف الضوئي على برايل العربي

٨. التعرف على الخلايا (Cells)

ولتحويل الخلية إلى (binary code) نقوم بتقسيمها إلى ستة أقسام كما في الشكل التالي، ومن ثم يوضع الرقم إذا يوجد نقطة أو صفر إذا كانت المنطقة خالية.



بعد تحويل الصورة إلى ترميز ثنائي نقوم بالتحويل إلى الترميز العشري:

$$\text{decimalCode} = b1 + b2*2 + b3*4 + b4*8 + b5*16 + b6*32$$

مراحل التعرف الضوئي على برايل العربي

٩. تصحيح وضعية الورقة

في كثير من الحالات تكون الصفحة معكوسة بمقدار درجة عند وضعها في الماسح الضوئي، لهذا لابد من إعادتها إلى الوضع الصحيح ليتم التعرف عليها. ولعمل ذلك نقوم بحساب نسبة وجود الأحرف الهجائية (بدون الرموز والعلامات الأخرى) في الصفحة.

تجدر الإشارة إلى أنه في بعض الماسحات الضوئية تكون النقاط البارزة ظلها لأعلى والأخرى ظلها لأسفل، وعلى هذا سيكون لدينا أربع احتمالات لوضع الورقة كما في الشكل التالي، حيث أن ظهور الانعكاس الأفقي والعمودي هو بسبب عدم تحديد مكان الظل للنقطة.

ومن خلال التجربة عندما تكون الصفحة في وضعها الصحيح فإن نسبة الأحرف الصحيحة أكثر من باقي الأوضاع الأخرى.

P

q

الوضع الصحيح

انعكاس أفقي

d

b

انعكاس ١٨٠ درجة

انعكاس عمودي

نتائج التجارب

وبعد تطبيق جميع الخوارزميات السابقة لمعالجة ورق برايل، كان متوسط نسبة التعرف ، %، وقد ظهر أثر هذه الخوارزميات الواضح على جودة التعرف، وفيما يلي جدول يبين نتائج هذه التجارب علماً بأن ورقة برايل كاملة ذات وجه واحد (single side) تحوي حرف وذات الوجهين (double side) تحوي حرف:

، %		، %			
		، %			
، %		، %			
		%			

الخاتمة

لقد قمنا في هذه الورقة بعرض الخوارزميات والتي قمنا بتطويرها من أجل استعمالها في المرحلة الأولى من نظام التعرف الضوئي على برايل العربي والمتعلقة بمعالجة الصور وتحويلها إلى صيغة رقمية، ولقد أظهرت النتائج المنفذة نسباً عالية في التعرف.

أما بالنسبة للأعمال المستقبلية فستكون بمحاولة التعرف على مطبوعات ذات الحجم الكبير وإيجاد الحلول المناسبة لذلك، إضافة إلى إكمال الجزء الثاني من نظام التعرف الضوئي وهو تحويل خلايا برايل التي تم التعرف عليها إلى نص (عربي/إنجليزي)، ولذا يلزم التعرف على اللغة وعلى قواعد التحويل.