معالجة دقّة النصوص العربيّة الرقميّة لمزيد من الوضوح و الفاعليّة

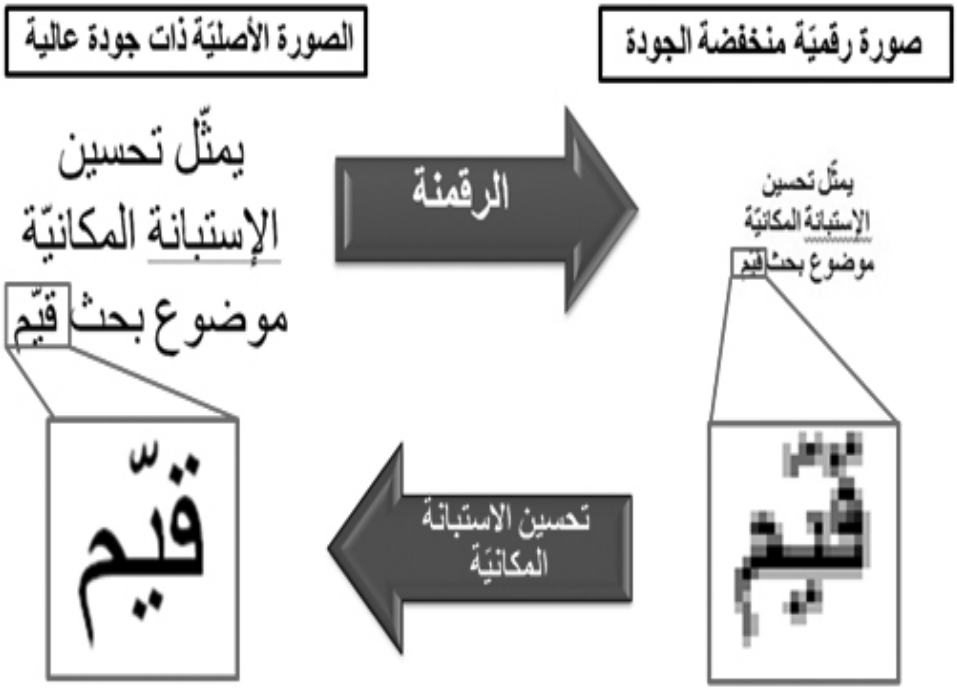
ريم ولها، فدوى دريرة، فرنك لبورجوا ، محمد عادل العليمي، كرستوف فرسيا

١. المقدّمة

غزت الوثائق المرقّمة جلّ معاملاتنا في عصر تميّز بتطوّر التّقنيات الحديثة للمعلوماتيّة و الإتّصالات. و يجدر بالذكر إلى أنّ أجهزة الهواتف الذّكيّة أو الكمبيوترات اللوحيّة المنتشرة على نطاق واسع اليوم هي من أبرز هذه التّقنيات و أحدثها. إلّا أنّ العديد من العوامل أسفرت عن ترويج صور رقميّة منخفضة الجودة حدّت من إستعمالاتها. و تعود هذه العوامل بالخصوص إلى أوّلا محدوديّة تكنولوجيّة بعض الأجهزة المستعملة في مشاريع الرّقمنة، و ثانيا اللّجوء إلى إختيارات للرّقمنة تركّز على التخفيض من الكلفة الجمليّة (الحدّ من تكلفة التّخزين، سهولة معالجة و تبادل كميّات كبيرة من الصّور) على حساب الجودة في الإستعمالات (انظر رسم ١ كمثال).

وفي هذا الإطار، يتناول هذا البحث موضوع معالجة الصّورة الرّقميّة و يهتّم أساسا بالصّورة التّي تحتوي على نصّ مكتوب باللّغة العربيّة. هدفنا الأساسيّ هو تحسين الاستبانة المكانيّة الّتي تشير إلى كثافة النّقاط في وحدة مساحة الصّورة الرّقميّة. فالإستبانة المكانيّة تحدّد دقّة وجودة الصّورة الرّقميّة. فمن المعلوم، التّخفيض في الإستبانة المكانيّة عند عمليّة التّصوير الرّقميّ يؤدي إلى إنحلال الصّورة و تردّيها و عدم وضوحها. و يتفاقم هذا المشكل خاصّة في حالة الصّور النصيّة ذات حروف ضئيلة السّمك. لكنّ الإشكاليّة لا تقتصر على المشاهد، فأنظمة التعّرف الضوئيّ على الحروف تواجه هي الأخرى مشاكل كبيرة لمعرفة الأحرف في الصّورة و تحويلها إلى نصّ حاسوبيّ. في هذا الصّدد، يقدّم هذا البحث نظام تحسين الإستبانة المكانيّة للصّور النّصيّة العربيّة بالإعتماد على طريقة التّرميز المتناثر التّي سنعرضها فيما بعد. لقد قمنا بتقييم نتائج هذا النّظام المقترح بصريّا و كميّا على صور نصيّة مختلفة تحتوي على كتابات عربيّة مطبوعة و أخرى مكتوبة بخطّ اليد و لقد توصّلنا إلى نتائج مثيرة للاهتمام.

بقية محتوى هذا التقرير سيكون كما يلي: القسم الثاني سيقدم استعراض موجز للأعمال ذات صلة بهذا البحث. بعد ذلك، سيهدف القسم الثّالث إلى تقديم تفاصيل نظام تحسين الإستبانة المكانيّة للصّور النّصيّة العربيّة المقترح. أمّا القسم الرّابع، فهو سيعرض دراسات تجريبية على صور نصيّة مختلفة و مقارنة مع نتائج تولّدها أنظمة أخرى قائمة. وأخيرا، إستنتاجات وبعض وجهات النّظر الواردة سيتم إعطاؤها في القسم الخامس.



رســـــم ١ : تأثير الرّقمنة على جودة الصّورة الأصليّة و دور تحسين الإستبانة المكانيّة

٢. لمحة عامّة على الأنظمة الموجودة

يمثّل تحسين الإستبانة المكانيّة موضوع بحث قيّم بسبب سرعة تزايد الحاجة إلى صورة رقميّة عالية الجودة في العديد من التطبيقات كتحليل الصورة و رؤيا الحاسوب. و لقد تمّ في الأبحاث السابقة إقتراح العديد من الطّرق لحلّ هذا الإشكال لكنّ معظمها تطرّق إلى تحسين الصّور الطّبيعيّة. و يجدر بالإشارة إلى أنّ تطبيقها المباشر على الصّور النصيّة العربيّة ليس فعّالا بما فيه الكفاية نظرا للخصائص التي تميّز هذه الصّور على الصّور الطّبيعيّة. فالصّور النّصيّة العربيّة لها من الخاصيّات ما يميّزها عن غيرها من الصّور مثل وجود عدّة أجزاء هيكلية (نذكر على سبيل المثال الحواف و الزّوايا، قطاعات الخطّ و علامات التشكيل) تشكّل الحروف و الّتي ينبغي الحفاظ عليها. و اللّجوء إلى إستبانة مكانيّة غير كافية قد يؤدي إلى فقدان مجموعة من المعلومات. بالتّالي، الصّور النّصيّة العربيّة هي فئة متميّزة من الصّور و ينبغي تحديد أساليب مناسبة لها.

جلّ التّقنيّات الّتي تمّ إقتراحها لتحسين الإستبانة المكانيّة يمكن إدراجها ضمن قسمين رئيسيين : القسم الأوّل يعتمد على الحفظ و يستخدم بالأساس تقنية القاموس الذي يحتوي على أزواج تجمع بين عيّنات من الصّورة العالية الجودة و ما يوافقها في حالة إنخفاض الجودة. و يجدر بالذكر أنّ إستخراج هذه الأزواج من العيّنات يجب أن يكون مسبوقا بعمليّة حفظ و تدريب دقيقة جدّا تختلف ملامحها بإختلاف الطّرق و الأساليب المستعملة -١،٢،٣- . أمّا القسم الثّاني فهو لا يقوم على أساليب التعلّم و إنّما يرتكز على الخوارزميّات لتحسين دقّة الصورة. فمن بين هاته التّقنيّات يمكن أن نذكر طريقة الاستيفاء بأقرب النّقاط المجاورة، الاستيفاء المترابط ذو بعدين و الاستيفاء المترابط ذو ثلاث أبعاد. هذه الأساليب هي بسيطة وسريعة، ولكنّها تدخل الضّبابية على طول الحواف. من أجل الحدّ من هذا التّدهور لجودة الصّورة الرّقميّة، تمّ اقتراح طريقة الاستيفاء التّكييفي التي