

تقنية مقترحة لكبس الوثائق النصية العربية

ميسون خضر النعيمي

Email: [Maysoon\\_khider@yahoo.com](mailto:Maysoon_khider@yahoo.com)

د. خليل ابراهيم السيف

Email: [khalil\\_alsaiif@yahoo.com](mailto:khalil_alsaiif@yahoo.com)

قسم علوم الحاسبات | جامعة الموصل | العراق

الخلاصة

تعد دراسة الكبس بأنواعه كافة موضوعا مهما في جميع المجالات ، إذ أن تقليل المساحة الخزن نية في الذاكرة المستخدمة لخرن البيانات فضلا عن الوقت الذي تتطلبه تلك البيانات في النقل دفع العديد من الباحثين إلى عمل خوارزميات لكبس البيانات ومن ثم إعادة فك كبسها. و نظرا لذلك فقد تم في هذا البحث دراسة صور الوثائق النصية العربية ومحتوياتها لغرض إمكانية كبسها من خلال اقتراح خوارزمية خاصة بصور الوثائق النصية العربية اعتمادا على استخلاص المعاملات الخاصة بصفات الوثيقة.

وقد تضمنت أفكار الخوارزمية المقترحة العمل على استئصال الأسطر ثم استئصال الأعمدة الغير مستخدمه من الوثيقة النصية لغرض اختزال حجم الوثيقة النصية إلى اقل ما يمكن وذلك لاختزال فترة تنفيذ عملية الكبس وتسهيلها ، و من خلال تطبيق مفاهيم خوارزميات التقسيم الشجري الرباعي على محتويات الوثيقة المختزلة ليتم الحصول على وصف كامل لنسيجها ومن ثم إنشاء ملف جديد يحوي على المعلومات الخاصة بوصف الصورة النصية.

وجدت انه عند تطبيق تلك الخوارزمية على ملفات نصية بهيئات مختلفة كانت نتائج الكبس مشجعة لغرض اعتمادها من قبل برامجيات الكبس ، فقد تم الحصول على نسبة كبس جيدة ، فضلا عن ذلك تم الحصول على وضوحية جيدة وعالية عند فتح كبس صور تلك الوثائق النصية. علما أن نسبة الكبس تراوحت من 1:5.

Abstract

The study of compression types is regarded as an important subject in all fields. Minimizing the storage area in the memory used in storing data and the time required by the data on transferring have urged many researchers to prepare algorithms to compress data and then to repeat untying compression. In the current research, the pictures of Arabic textual documents together with their contents have been studied in an attempt to compress them through the suggestion of an algorithm specific to the pictures of Arabic textual documents based on the deduction of coefficients of the features of the document.

The ideas of the algorithm suggested subsumes working to uproot the lines and then the columns of the textual document so as to minimize the size of the textual document to its minimum degree by minimizing the period of executing and by facilitating the process of compression. All that is done by applying the concepts of the algorithms of the four-fold tree division to the contents of the minimized document so as to get a full description of the texture of that minimized document, then building up a new file that contains the data limited to the description of the textual picture.

We have also found that, on applying that algorithm to the textual files in different shapes, the outcomes have been very encouraging so as to be adopted by compression programmes. We have also obtained a good compression ratio and also a better and higher vividness on opening the compression of those textual documents. The compression ratio ranged between 1:5

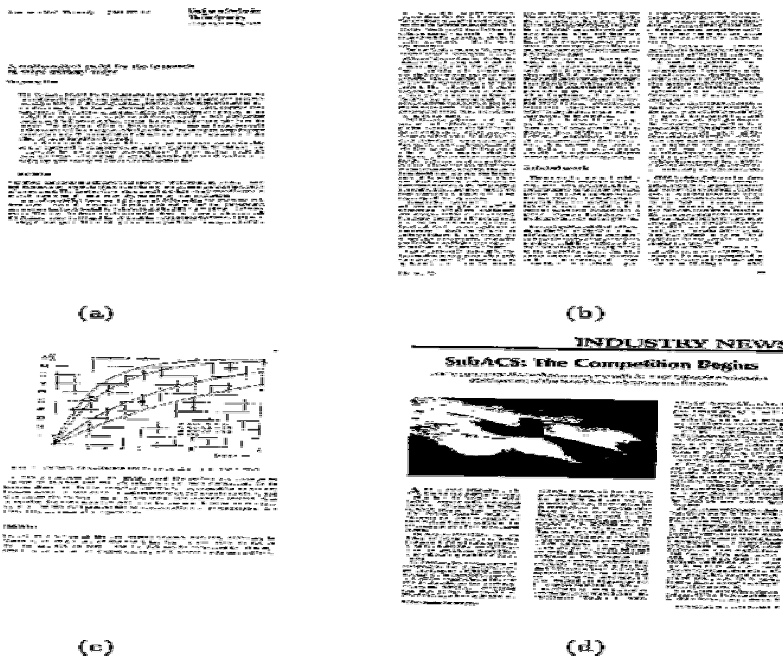
## 1- تمهيد :

موضوع كبس الصور يدخل في حقل معالجة الصورة، و لتحليل الصورة يجب أن يكون هناك قدرة لمعالجة النتائج المتمثلة مباشرة . وقد ازدادت الأهمية للكبس بشكل كبير نتيجة للنمو السريع في قدرة الحاسبة ، والنمو الذي قابله في مجال تطبيق الأوساط المتعددة ( Multimedia ) والمحاسن في الشبكة العالمية للانترنت والتي تمكن الجميع من الوصول إلى أي موقع ، فضلا عن محاسن تقنية الفيديو وتشمل الإمكانيات العالية لاستخدامات التلفزيون والتي تتطلب الجديد والأفضل والأسرع لخوارزميات الكبس [7].

فكبس البيانات لا يزال يعد حقلًا مهمًا جدًا في التطبيقات العديدة مثل نقل المعلومات خلال شبكة الاتصالات الرقمية ( Networks )، فضلا عن أهميته في تقليل مساحات الخزن في الذاكرة المستخدمة لخرن هذه البيانات ، وأهميته في الاستخدام الأمثل للإمكانيات والموارد المتوفرة مما يؤدي ذلك إلى تقليل الكلفة بشكل عام من جهة ويؤدي إلى تقليل وقت النقل المستخدم وتقليل حزمة الإرسال ( Bandwidth ) من جهة أخرى. فالتقدم السريع جدا زاد في كثافة كمية الخزن وتسريع المعالج ، حيث أن الحاجة لتطوير كثير من الطرق لتحليل الإشارة والصور لا تكفي ولكن يجب أن نضع ونطور الطرق لكبسها [6].

## 2\_ خصائص الوثيقة ( Document characteristic ) :

صور الوثائق المكتسبة خلال الماسح الضوئي ( scanned documents ) في كثير من الحالات هي pseudo - binary وهي غنية بالمحتويات النصية ، و بالإمكان تعريف صورة الوثيقة على إنها صورة تحتوي على عناصر تمثل رموز اللغة [13]. حيث يمثل الشكل ( 1 ) صورة لوثائق بصيغ مختلفة . وقد مثلت بشكل مناسب بواسطة صور تم اكتسابها إلى الحاسبة بوضوحية 300 نقطة لكل انج مربع ( 300 dpi ) ، والتي تبدو بهيكل خارجي عالي المستوى يحتوي على تناظر كبير بأشكال الرموز الموجودة في داخلها .



الشكل (1) صور لوثائق متنوعة

و نظرا للاستخدامات الواسعة للوثائق في عدد كبير من المجالات ، عليه يجب أن يعار اهتمام عالي جدا على مقدار تأثير الصورة بواسطة الكبس ، وذلك لكون الوثيقة النهائية سوف تقرا بواسطة البشر ، فمن الضروري لهيكل الكبس أن يحافظ على أشكال العناصر كي يتمكن من إدراكها وفهمها بواسطة القارئ بعد الاسترجاع ، و أن إعادة إنتاج هذه الوثيقة واسترجاعها يجب أن تكون بصيغة قريبة من الأصل .

### 3- الهدف من البحث :

يهدف البحث إلى وضع خوارزمية لها القابلية في التعامل مع ملفات صور الوثائق لاجل كبسها من خلال استخلاص الخواص لكل مجموعة من البيانات الداخلة إلى الحاسبة بأسلوب يؤدي إلى اختزال وتقليل كبير في حجم الملف والذي بدوره يؤدي إلى تقديم خدمة كبيرة في عملية ترسل البيانات والوثائق عبر وسائل الاتصالات المختلفة ، وكذلك سرعة التراسل، وتقليل حجم الخزن، والكلفة المادية ومع إمكانية إعادة بناء الملف الأصلي بشكل بسيط وسهل .

### 4- معالجة الوثائق :

لغرض معالجة الوثائق هنالك عدد من العوامل التي يجب مراعاتها كي تكون المعالجة ناجحة وكفوءة ومن تلك العوامل توفير ثلاث أساسيات أولية (رئيسية) هي [10]:-

1- حجم الوثيقة 2- تجهيز وصول للوثيقة 3- أن تكون الوثيقة صالحة للقراءة

### 5 - المبادئ من وراء الكبس :

الخصائص المشتركة لمعظم الصور إن النقاط المتجاورة تكون مترابطة بشكل كبير ولهذا السبب تحتوي على معلومات متكررة بشكل كبير . و الفعالية المهمة في ذلك هي إيجاد تمثيل جديد لتلك النقاط بأسلوب يقلل من التكرارات وهو مبدأ رئيسي في عملية الكبس [12]. و تقليل التكرار يتم من خلال عملية مسح التكرار من الإشارة ( image / video ) الأصلية ، كما يتم النقل من خلال حذف أجزاء من الإشارة والتي سوف لن تلاحظ بواسطة مستقبل الإشارة .

عموما توجد ثلاثة أنواع من النقل في الصور الرقمية والصورية يمكن تحديدها:-

- 1- النقل الخاص ( Spatial Redundancy ) أو الارتباط ( correlation ) بين قيم النقطة المتجاورة .
  - 2- النقل الطيفي ( Spectral Redundancy ) أو الارتباط بين مستويات الألوان المختلفة أو بين زمرة ( شريط ) الطيف ( Spectral bands ) .
  - 3- النقل الزمني ( Temporal Redundancy ) أو تقليل الارتباط بين الهياكل ( الإطارات ) ( frames ) المتقاربة في سلسلة من الصور .
- معظم أهداف بحوث كبس الصورة هي تقليل عدد الـ bits المحتاجة لتمثيل الصورة بمسح التكرارات الخاصة والطيفية أن أمكن.

## 6- الاستعراض المرجعي:

بداية لا يعد الكبس من المواضيع الجديدة ، فقد قام العديد من الباحثين في السابق بالعمل في هذا المجال حيث استخدم الكبس في كثير من المجالات منها: - في مجال تسهيل النقل ، وزيادة سرعة البيانات المنقولة ، كذلك في تقليل حجم البيانات... الخ. و إن الحديث عن هذا الموضوع يجر إلى الحديث عن الطرق المستخدمة في الكبس وعن البحوث التي أنجزت في هذا المجال.

ان كبس صورة الوثيقة أصبح معروف كحالة خاصة من الكبس حيث أن كثير من العلماء والباحثين عملوا بمجال الكبس و أنجزوا بحوث واطاريح جيدة في هذا المجال من ضمن هذه الاطاريح هي أطروحة " Document image compression and analysis " التي قام بها Omid Kia Ebrahimi حيث تضمنت هذه الأطروحة بكبس صورة الوثيقة بالاعتماد على مكتبات من النماذج وجدول بموقع العناصر وكذلك تطبيق طريقة التجميع ( clustering ) على هذه النماذج بدلا من طريقة تقليل التكرار القديمة. و هذه الصيغة من التمثيل هي أساس الكبس في ذلك البحث وهي تجهز وصول مباشر لعناصر الصورة. واستطاع اوميد أن ينجز عمل جيد في استخدام عدة خوارزميات [11].

أما الباحث Hui Cheng فقدم أطروحته وهي Document image segmentation and compression إذ قام هذا الباحث بعمل تقطيع للصورة باستخدام خوارزمية معينة كان هذا الجزء الأول من الأطروحة أما الجزء الثاني منها فهو خوارزمية كبس بعدة مستويات . و خوارزمية الكبس هذه أولا تقطع صورة الوثيقة المكتسبة بواسطة الماسح الضوئي scanner إلى عدة مستويات و من ثم كبس كل مستوى باستخدام خوارزمية خاصة به [9].

وقد قام الباحث جرجيس احمد خضر في سنة 1996 قدم بحثه Fractal vector quantization of image الذي على تطوير طريقة تكميم الكتل الصورية (VQ) باستخدام تقنية الكبس الكسري Fractal compression . و تستخدم تقنية الكبس الكسري لكبس دليل الرموز الصورية (Codebook) ، وهي نتائج عرضية تنتج عن الكبس بطريقة تكميم الكتل الصورية مما يؤدي ذلك إلى زيادة سرعة الإرسال المباشر لدليل الرموز و إلى تقليل المساحة الخزنية اللازمة لخزنه. علما انه قد تم تبني خوارزمية (Linde-Buzo- LBG Gray) في طريقة التكميم الصورية . بينما تم تبني خوارزمية نظام الدالة المتكررة (Iterated IFS Function System) في طريقة الكبس الكسري [10].

وتوجد عدة كتب نشرت في هذا المجال من معالجة الصورة ومن هذه الكتب هو كتاب computer vision and image processing للدكتور سكوت Scott E. Umbaugh والذي تكلم عن طرائق الكبس المقسمة على أساس حفظ المعلومات [14]. كذلك كتاب Advanced Data communication and Network من قبل W. Buchanan الذي تكلم عن طرائق الكبس وعن فإيلات الكبس بمختلف أنواعها [4]. أما كتاب Digital Image Processing للمؤلفان Rafael C. Gonzalez و Richard E. Woods فقد تطرق الجزء الثاني الذي صدر سنة 2002 بشكل تفصيلي عن طرق الكبس [3].

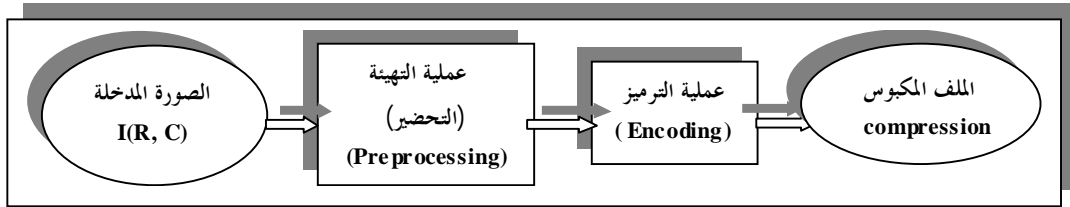
## 7\_ نماذج أنظمة الكبس : compression system models

معظم نماذج أنظمة الكبس تتألف من الجزئين الرئيسيين التاليين والذين من خلالهما يتم الحكم في معظم الأحيان على ذلك النموذج [1] [8] .

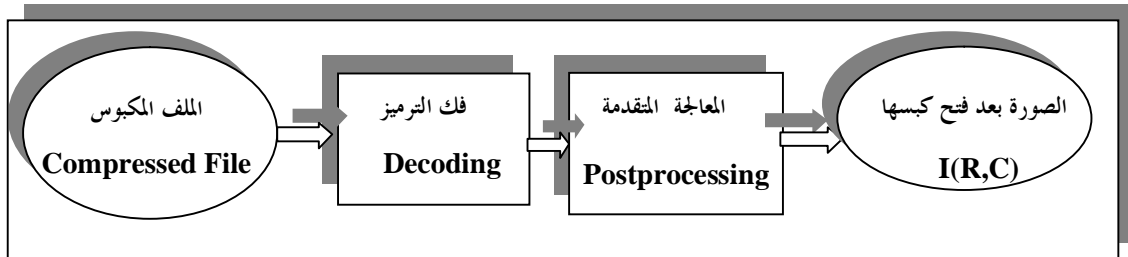
1- الكابس (compressor). بالإمكان اعتبار مرحلة الكبس متألفة من مرحلة التهيئة للعملية (preprocessing stage) ، و مرحلة التحليل أو الترميز (encode).

2- محلل (فك) الكبس (decompressor). يتألف فك الكبس من مرحلة إعادة التحليل (decode stage) بمرحلة المعالجة المتقدمة (postprocessing) .

وكما هو موضح في الشكل (2).



أ - الكبس



ب - فك الكبس

الشكل (2) يمثل مراحل الكبس وفك الكبس

تبدأ عملية الترميز بالمعالجة الأولية وذلك من أجل تهيئة الصورة لغرض ترميزها والتي قد تتضمن أي عدد من العمليات التي يحددها التطبيق، وبعد الكبس تتم عملية فتح الترميز حيث تعمل المعالجة المتقدمة على استبعاد بعض المواقع الغير مرغوب فيها كتلك التي تم الحصول عليها بواسطة عملية الكبس ، و إن اغلب خوارزميات الكبس المطبقة عملياً هي اتحاد (أو دمج) عدد من تقنيات الكبس المفردة. ومن الممكن أيضاً تجزئة الكابس (compressor) إلى مراحل مبيّنة بالشكل (3).

## 8\_ تصنيف الكبس

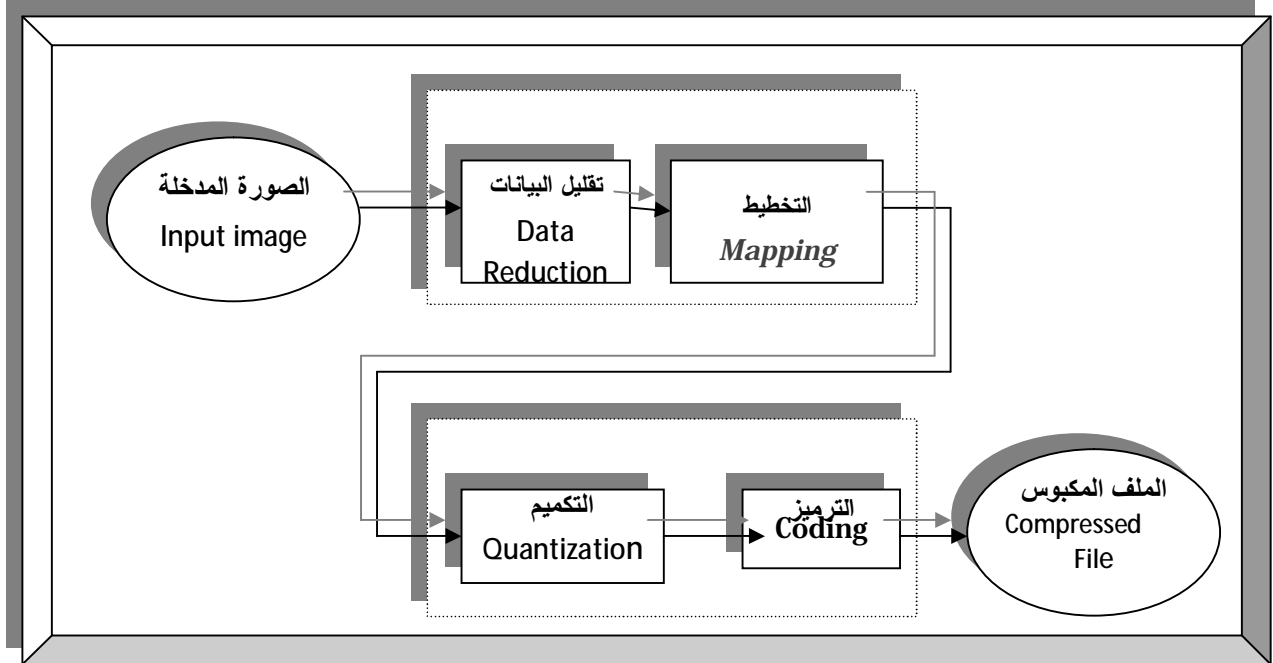
الكبس ممكن أن ينجز بعدة طرائق وأساليب وحسب هذه الطرائق تصنف خوارزميات الكبس إلى صنفين :-

## 1-8 طرق الكبس مقسمة على أساس الحفاظ على المعلومات

### • الكبس بفقدان في البيانات (lossy compression)

ومن طرق الكبس بهذا الأسلوب :

- Wavelet Coding.
- Vector Quantization.
- Quadtree.



الشكل (3) الكابس (compressor)

### • الكبس بدون فقدان في البيانات (lossless compression)

ومن طرائق الكبس بهذا الأسلوب :

- 1- طريقة هوف مان Huffman Coding
- 2- طريقة الترميز الرياضي Arithmetic Coding
- 3- طريقة LZW Coding

## 2-8 طرائق الكبس مقسمة على أساس أسلوب الاستخدام

يوجد صنفين من طرائق الكبس وذلك حسب أسلوب استخدامه وهما: [5]:

- طرائق (off-Line or) static :
- طرائق (on-Line) dynamic :

## 9\_ الخوارزمية المقترحة للكبس

لإنجاز كبس صورة الوثيقة النصية بالاعتماد على تطبيق طريقة اختزال الأسطر والأعمدة الفارغة والتي تم الحصول عليها من خلال إيجاد لون الخلفية ثم تطبيق خوارزمية التقسيم الشجري الرباعي (Quadtree) ونتيجة الكبس للصورة النصية تكون ملف يحتوي على المعاملات المهمة للصورة المكبوسة ، حيث أن الملف الذي تم تكوينه يبدأ بالرأس (Header) الذي يقدم وصف كامل لتوزيع البيانات في الصورة، أما البيانات فسوف تقع بعد منطقة الرأس حيث يمكن تقسيم الملف إلى الأجزاء التالية:

الرأس (Header) ويتكون من ثلاثة اسطر في كل سطر أربعة أرقام وكما يلي:

### السطر الأول:

الرقم الأول والثاني : يمثل أبعاد الصورة الأصلية (العرض و الطول).

الرقم الثالث والرابع: حجم الصورة القياسية في هذا البحث تم اخذ الأبعاد القياسية التالية (256x256 , 228x128).

### السطر الثاني:

الرقم الأول: يمثل عدد الأسطر التي تحتوي على مواقع الأسطر الفارغة داخل الصورة الأصلية والتي تلي الأسطر الحاوية على مصفوفة التقسيم الشجري الرباعي.

الرقم الثاني: يمثل عدد الأسطر التي تحتوي على مواقع الأعمدة الفارغة داخل الصورة الأصلية والتي تلي الأسطر الحاوية على مصفوفة التقسيم الشجري الرباعي والأسطر الحاوية على مواقع الأسطر الفارغة.

الرقم الثالث: يحتوي على طول المصفوفة الناتجة من تطبيق خوارزمية التقسيم الشجري الرباعي.

الرقم الرابع: غير مستخدم.

### السطر الثالث:

الرقم الأول والثاني : يمثل أبعاد الصورة المختزلة (أي الخالية من الأسطر والأعمدة).

الرقم الثالث والرابع: غير مستخدمة (للاستخدامات المستقبلية).

وفيما يلي مخطط يوضح هيكلية الملف المتكون نتيجة الكبس، شكل (4):

في حين تتناول المبحث الثاني معالجة الملف الناتج من عملية الكبس و الذي يحتوي على معاملات الصورة المكبوسة ثم فتح الكبس بالاعتماد على تلك المعاملات.

## 10 - مراحل تنفيذ الخوارزمية المقترحة للكبس

### المرحلة الأولى:

تمثل هذه المرحلة عملية اكتساب الصورة إلى الحاسبة الإلكترونية ، إذ يتم استخدام الماسح الضوئي أو الكاميرا الرقمية أو الفيديوية لإدخال صورة الوثيقة النصية الملونة ، ومن ثم يتم تخزين بيانات الصورة في ملف بصوري من نوع Bmp ، ليتم التعامل معها وإجراء المعالجة الخاصة عليها والشكل (1- ملحق) يمثل الصورة المدخلة .

### المرحلة الثانية:

تتم في هذه المرحلة تحويل الصورة الرقمية التي تم الحصول عليها من المرحلة الأولى إلى ثلاث صور ذات تدرج رمادي (الأحمر و الأخضر و الأزرق) كما في الشكل (2- ملحق) الذي يقدم وصف كامل لهذه المرحلة. المرحلة الثالثة:

يتم إيجاد لون الخلفية للصورة الرمادية بالاعتماد على قيمة العتبة من خلال المدرج التكراري للصورة حيث يتم اعتماد الخوارزمية التي تم ذكرها في الفصل الثاني (المبحث الرابع).

### المرحلة الرابعة:

تمثل هذه المرحلة مرحلة المعالجة الأولية ، والتي يجب أن يتم إنجازها على الصورة قبل البدء بإجراء المعالجات اللاحقة عليها ، تتضمن هذه المرحلة أولاً اختزال جميع الأسطر و الأعمدة الفارغة والتي سوف تقلل من الحجم ثم الاحتفاظ بإحداثيات مواقع تلك الأسطر والأعمدة ووضعها في مصفوفة مؤقتة ، حيث تم بناء دالة خارجية لإنجاز هذا العمل وإضافتها إلى الدوال الخارجية لبرامجيات Matlab وبالشكل التالي:

```
Function [trankimage.zerorows.zerocols] = trank (image1)
```

حيث أن:

**image1**: هي الإدخال للدالة ويمثل صورة نصية بمواصفات اعتيادية.

**Trankimage**: هي إخراج الدالة ويمثل صورة نصية تم اختزال جميع الأسطر والأعمدة منها.

**Zerorows**: تمثل مصفوفة احادية تحتوي على الأسطر الفارغة (أي الخالية من الكتابة أو الصور).

**Zerocols**: تمثل مصفوفة احادية تحتوي على الأعمدة الفارغة (أي الخالية من الكتابة أو الصور).

و الأشكال (3 و 4 - ملحق) تمثل مثال تطبيقي على عمل تلك الدالة:

وفيما يلي المخطط الصندوقي لعمل تلك الدالة التي تم بنائها لإنجاز عملية استئصال الأسطر والأعمدة الفارغة

من الصورة النصية ، شكل (5) :

### المرحلة الخامسة:

تتضمن هذه المرحلة تحويل الصورة المختزلة والنتيجة من المرحلة الرابعة والتي تكون خالية من الأسطر والأعمدة الفارغة إلى أبعاد قياسية ، أي إعادة تعيين حجم الصورة بأبعاد مربعة (صورة مربعة)، وذلك لان



خوارزمية التقسيم الشجري الرباعي تتطلب أن تكون الصورة ذات أبعاد قياسية (مربعة) ، حيث تم اعتماد الأبعاد (256x256) في هذا البحث ، و الشكل (5- ملحق) يبين الصورة الناتجة من هذه المرحلة:

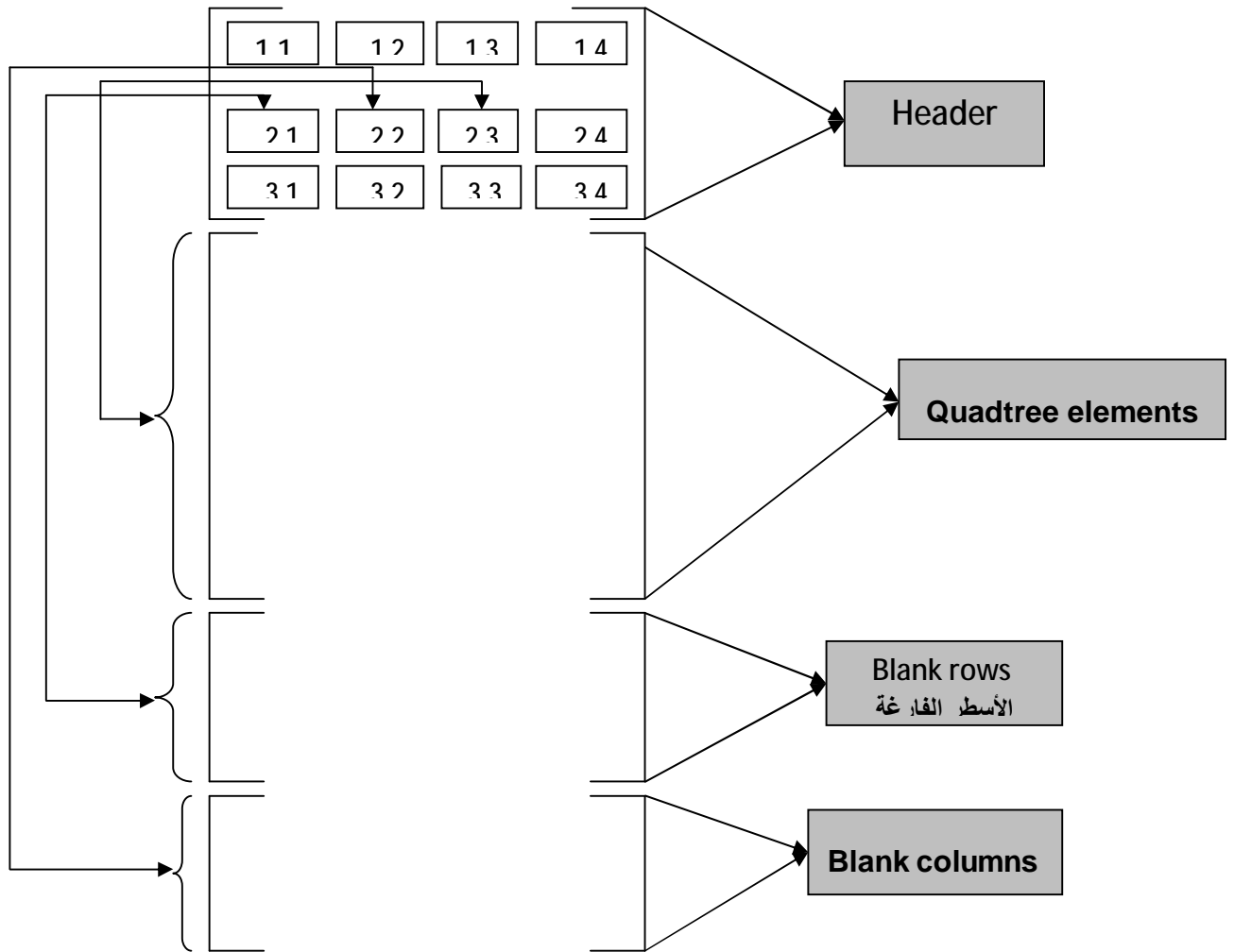
### المرحلة السادسة:

في هذه المرحلة يتم اختيار معيار أو مقياس للتجانس ، طريقة الاختيار لهذا المعيار تعتمد على الهدف من المعالجة المطلوبة للصورة الرقمية ، و لأجل تطبيق معيار التجانس على الصورة التي يتم تقطيعها إلى مناطق بقيم شدة متجانسة ، يجب أن يتم اختيار أو حساب قيمة العتبة الخاصة بمعيار التجانس للتقسيم الشجري الرباعي، وقد تم الاعتماد على معيار التجانس الموضح بالمعادلة التالية:  
حيث أن:

$$|Max - Min| \leq 30$$

Max : تمثل أعلى قيمة في الصورة.

Min: تمثل اقل قيمة في الصورة.



الشكل (4) هيكلية الملف الناتج

30 : تمثل قيمة العتبة التي تم الاعتماد عليها في هذا البحث والتي تم الحصول عليها عمليا من خلال اختيار مجاميع كبيرة من الأرقام من قبل باحثين سابقين [2].

### المرحلة السابعة:

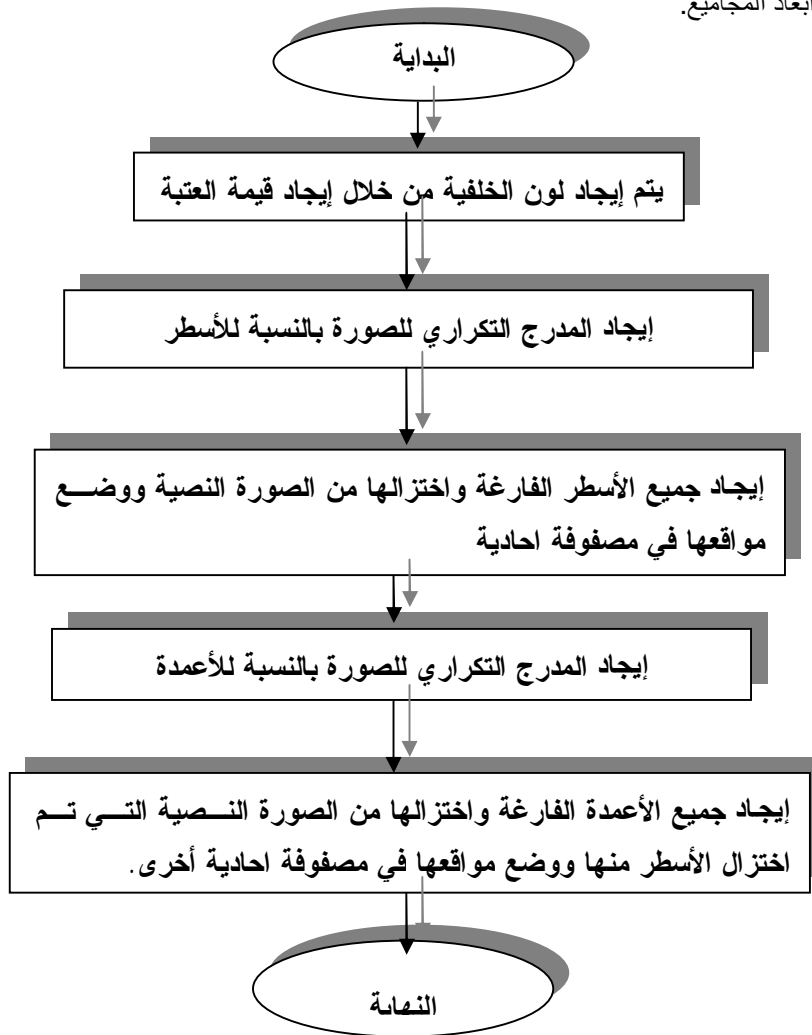
يتم تطبيق التقسيم الشجري الرباعي على الصورة القياسية (المربعة) ذات التدرج الرمادية الناتجة من المرحلة الخامسة والحصول على معاملات لتلك الصورة المربعة حيث يتم في هذه المرحلة تكوين مصفوفة تحتوي على أربعة أعمدة وكما موضحة بالملحق\ الشكل (6-ملحق) وهي تضم :

العمود الأول والثاني: يحتوي على إحداثيات مواقع المجاميع التي تم الحصول عليها من الصورة بعد تطبيق خوارزمية التقسيم الشجري الرباعي.

العمود الثالث: يحتوي على معدل الشدة لكل مجموعة الذي تم إيجاده بعد عملية التقسيم الشجري الرباعي

على الصورة.

العمود الرابع: يحتوي على أبعاد المجاميع.



الشكل (5) يمثل المخطط الصندوقي لعمل دالة اختزال الأسطر والأعمدة الفارغة

### المرحلة الثامنة:

يتم تكوين الملف الذي يحتوي على الرأس (Header) كما مر ذكره في بداية هذا الفصل ، وكذلك يحتوي هذا الملف على كل المعلومات الضرورية لاستعادة الصورة الأصلية والتي تشمل مصفوفة التقسيم الشجري الرباعي وإحداثيات مواقع الأسطر والأعمدة التي تم اختزالها وتأتي هذه المعلومات بعد الرأس مباشرة و الشكل (7- ملحق) يمثل الملف الناتج من كبس صورة الشكل (1- ملحق) والتي تم معالجتها خلال خطوات التنفيذ: وقد تبين من خلال خزن محتويات الملف المتكون أن حجم الملف يساوي (37.7 KB) بينما كان حجم الصورة قبل الكبس يساوي (187 KB) وبذلك تم حساب نسبة الكبس كما يلي:

$$\text{نسبة الكبس} = \frac{\text{حجم الملف قبل الكبس}}{\text{حجم الملف بعد الكبس}}$$

$$5 \text{ KB} \approx 4.96 \text{ KB} = \frac{187 \text{ KB}}{37.7 \text{ KB}} = \text{نسبة الكبس}$$

أي أن : نسبة الكبس = 1:5 أي أن الصورة تم كبسها إلى: 0.2 من حجمها الأصلي

وقد تم حساب مقياس المصدقية على أساس (Root-mean-square error) وكانت النتيجة: eRMS = 0.915

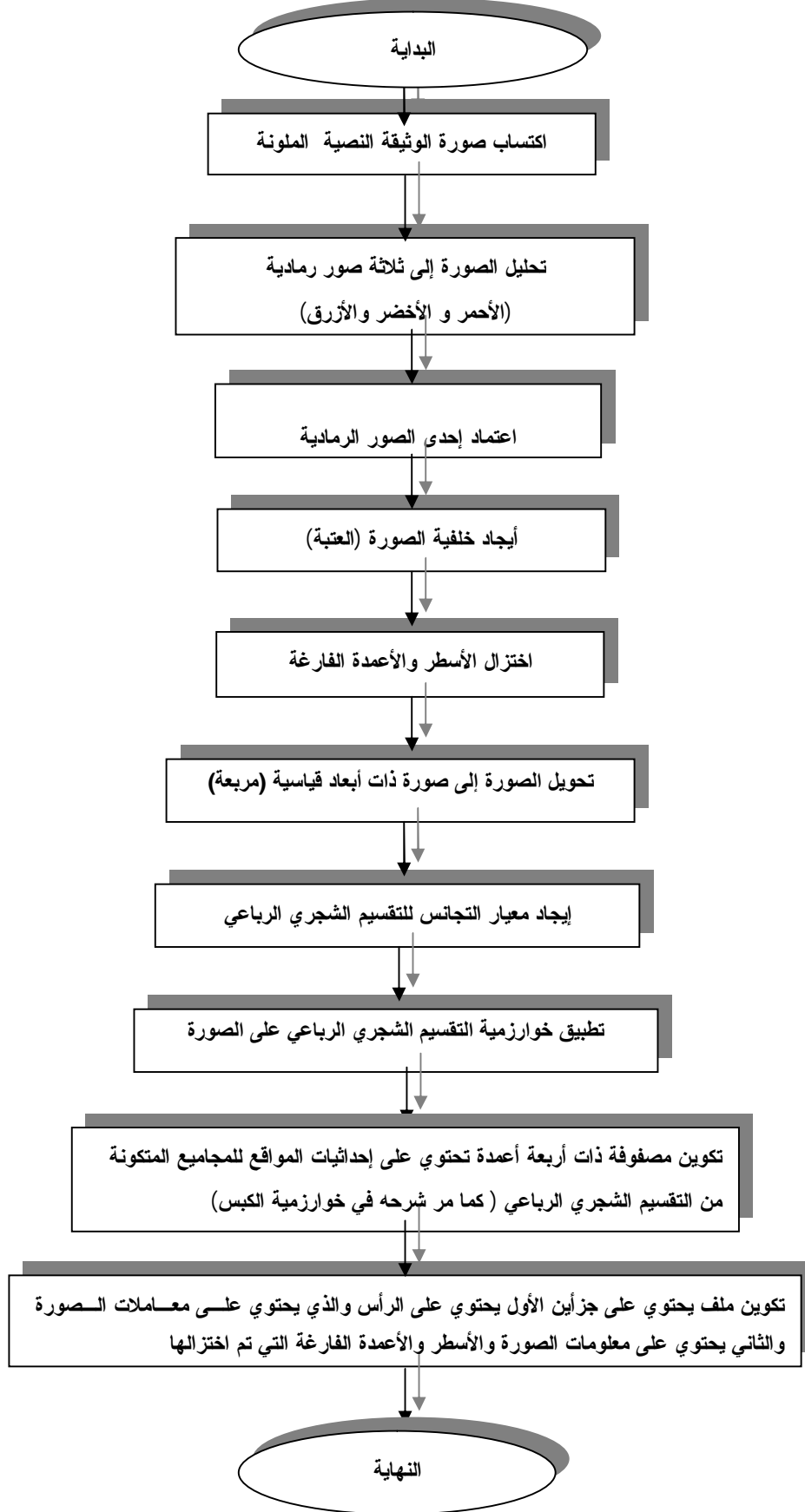
وجداول (1 - ملحق) يبين النتيجة بشكل واضح للمثال المذكور ضمن الملحق.

ولغرض تتبع مراحل تنفيذ الخوارزمية المقترحة للكبس بشكل سهل تم تقديم مخطط صندوقي متكامل وكما في الشكل (6). أما عملية فك الكبس للملف الناتج من كبس صورة الوثيقة النصية فتكون عملية معاكسة لعملية الكبس تقريبا ويظهر ذلك واضحا في الأشكال (8-9-10-11-12 - ملحق).

### 11\_ الاستنتاجات

تم في هذا البحث تطبيق ناجح للخوارزمية المقترحة في مجال كبس صور الوثائق النصية حيث طبقت على نماذج كثيرة ومتنوعة جميعها أدت إلى الحصول على نسبة جيدة للكبس ونوعية جيدة وواضحة ، وقد تبين أن الخوارزمية المقترحة أعطت أسلوبا جديدا في إمكانية كبس صور الوثائق النصية و إمكانية فتح كبس تلك الوثائق من خلال اعتماد معاملات الصورة ، وذلك بوصف كامل لشكل النسيج المكون لصورة الوثيقة النصية العربية.

وقد أعطت هذه الخوارزمية نتائج ذات نسبة كبس عالية عند تطبيقه على صور مختلفة حاوية على نص فقط أو نص وصورة أو صورة فقط تم خزن هذه الوثائق على الحاسبة من خلال كاميرات رقمية أو من خلال اكتسابها بواسطة الماسح الضوئي ، ولم يثبت أي خلل في كبس تلك الصور المختلفة ، وعند فتح كبس جميع تلك الصور كانت الوثائق النصية المسترجعة بصورة صحيحة وكذلك الحصول على نسبة جيدة للكبس بالإضافة إلى الحصول على وضوحية جيدة بالنسبة للصور المعاد فتح كبسها.



الشكل (6) المخطط الصندوقي لتنفيذ مراحل كبس صورة الوثيقة النصية

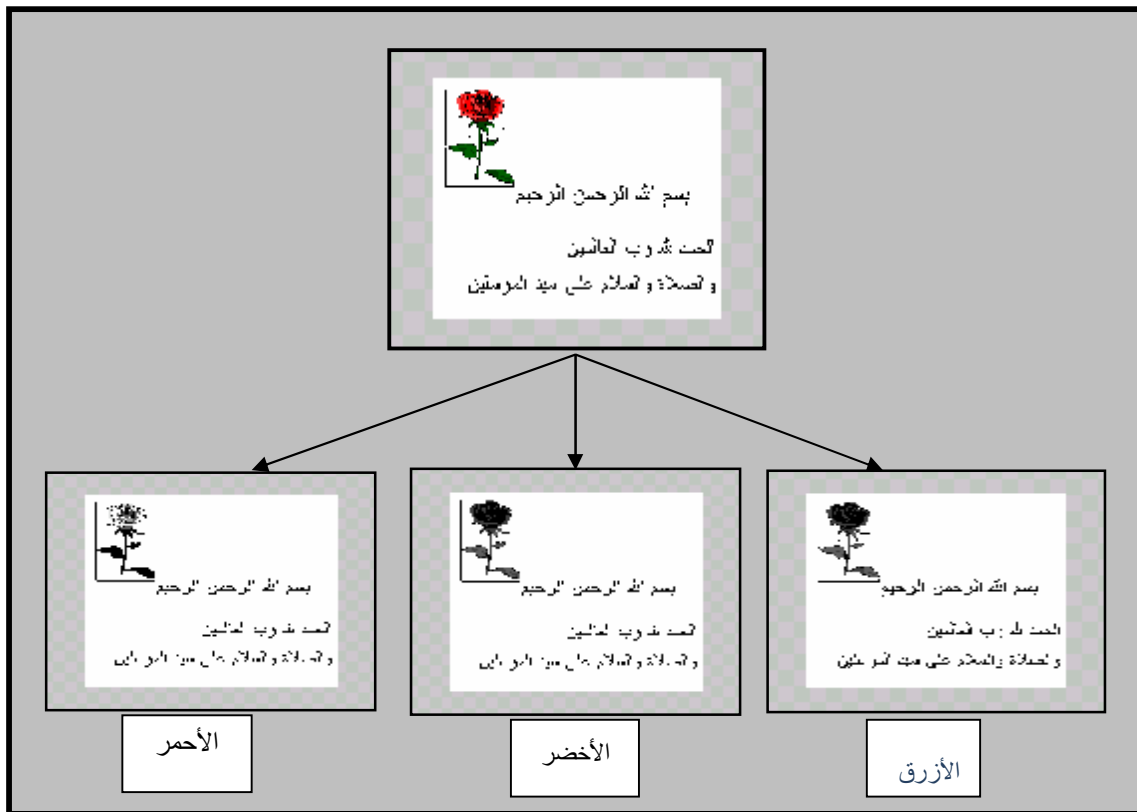
12\_المصادر:

1. حسن، زهور حسين ، 2001 : " تطبيق حافظه الشاشة لنظام التشغيل ويندوز 98 " كلية علوم الحاسبات والرياضيات ، جامعة الموصل ، الموصل.
2. عبدالله، لمى اكرم الصفار ، 2002: " تمييز اللهب في الصور الرقمية الملونة " كلية علوم الحاسبات والرياضيات ، جامعة الموصل ، الموصل.
3. Brigger, Patrick, Muller Frank, Klaus Illgner, Unser Michael, 1999: " Centered Pyramids ", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.8NO.9, and 1057-7149/99 \$10.00.
4. Buchanan, W., 1997: " Advanced Data Communications and Networks " first edition, Prentice Chapman and Hall.
5. Debra, A. Lelewer, Daniel S. Hirschberg, 2003: " Data Compression ".  
[www.ics.uci.edu/~dan/pubs/datacompress.html](http://www.ics.uci.edu/~dan/pubs/datacompress.html).
6. Edleno, S. de Moura, Gonzalo N., Nivio Z., 1997:"Indexing Compressed Text "Dep. Of Ciencia da Computer, University, Federal, Minas, Gerais, Brazil.  
[www.dcc.uchile.cl/~gnavarro/abstracts/wsp97.4html](http://www.dcc.uchile.cl/~gnavarro/abstracts/wsp97.4html)
7. Eero, S., Jaana K., Jussi K., Kalervo J., 2001:" Document text characteristics affect the ranking of the most relevant documents by expanded structured queries" © NEC Research Institute.
8. Gonzales, Rafael C.; Woodes, R. E., 2002: " Digital image processing ", 2<sup>nd</sup> Edition, Publisher: Prentice.
9. Hui, Cheng, 1999:" Document Image Segmentation and Compression " Pp.D., Purdue University.
10. Jarjees, Ahmed Khidir, 1996:" Fractal Vector Quantization of Images "M.Sc., thesis, computer science, College of Science, University Of Mosul, Mosul.
11. Kia, Omid Ebrahimi, 1997:" Document Image Compression and Analysis ".
12. Saha, Subhasis, 77136 hits, 2000: " Image Compression – from DCT to Wavelets: A Review ".
13. Sargur, N. S., Stephen W. L., Venu G., Rohini K.S.and Jonathan J. Hull, 2001: "Document Image Understanding " State University of New York at Buffalo, report from.  
<http://www.cedar.buffalo.edu/publications/techreps/survey/survey.htm/>
14. Umbaugh, Scott E, TA1634.U48, 1998: " Computer vision and image processing:" a practical approach using CVIPtools Prentice Hall PTR.

ملحق



الشكل (1) صورة ملونة



الشكل (2) مستوى ألوان الصورة  
{ 129 }



الشكل (3) الصورة مختزلة من الأسطر الفارعة



الشكل (4) الصورة مختزلة من الأسطر والأعمدة



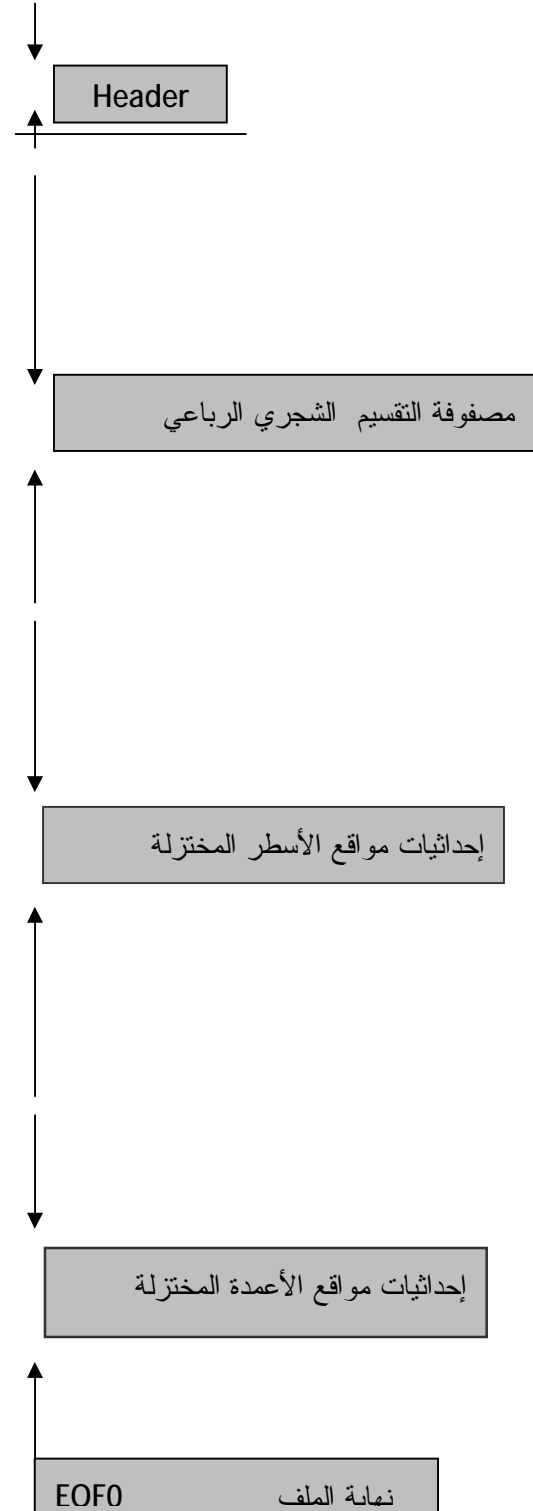
الشكل (5) الصورة بأبعاد قياسية (مربعة)

التسلسل	x	y		البعد
1.	93	1	255	2
2.	158	1	255	1
3.	159	1	255	1
4.	160	1	255	1
5.	161	1	255	32
6.	193	1	255	32
7.	225	1	255	16
8.	241	1	255	16
9.	1	2	255	1
10.	2	2	255	1
11.	3	2	255	1
12.	4	2	255	1
13.	5	2	255	1
14.	6	2	255	1
15.	7	2	255	1
16.	8	2	255	1
17.	9	2	255	1
18.	10	2	255	1
19.	11	2	255	1
20.	12	2	255	1
21.	13	2	255	1
22.	14	2	255	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	237	256	255	1
1	238	256	255	1
1	249	256	255	1
1	250	256	255	1
1	251	256	255	1
1	252	256	255	1

الشكل(6) المصفوفة الناتجة من عملية التقسيم الشجري الرباعي ( Quadtree )



	231	277	256	256
2	20	8	5121	0
3	153	248	0	0
4	93	1	255	2
5	158	1	255	1
6	159	1	255	1
7	160	1	255	1
8	161	1	255	32
9	193	1	255	32
	225	1	255	16
	241	1	255	16
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	1	2	3	4
1	5	6	7	8
1	9	10	124	125
1	126	127	128	129
1	130	131	132	133
1	134	135	136	137
1	138	139	140	141
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	1	2	3	4
1	5	6	7	8
1	9	10	11	12
1	122	123	126	127
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	273	274	275	276
5152	69	79	70	0



الشكل (7) الملف الناتج من عملية الكبس والذي يتكون من الرأس ومعلومات الصورة

رقم الشكل	حجم الصورة قبل الكبس	حجم الملف الناتج	نسبة الكبس	eRMS
(1)	963 KB	65 KB	1:15	0.922

الجدول (1) يبين نسبة الكبس و eRMS للمثال السابق



الشكل (8)

الصورة المتكونة من مصفوفة التقسيم الرباعي الشجري والموجودة بعد الرأس



الشكل (10) الصورة المختزلة بعد إعادة الأسطر المختزلة إليها



الشكل (9)

الصورة المختزلة (الصورة أي بأبعادها الأصلية قبل إجراء التقسيم الشجري الرباعي عليها)



الشكل (12) الصورة بعد تجميعها و إعادة الألوان إليها



الشكل (11) الصورة بعد إعادة الأسطر والأعمدة الفارغة إليها