مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (4) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (4) 2018

نظام ذكي لتقليص حجم وزمن عرض الفيديو

د. جـــورج كـــرّاز *

(تاريخ الإيداع 10 / 6 / 2018. قُبِل للنشر في 9/ 8 / 2018)

□ ملخّص □

تعتبر عملية الاختزال الذكية، لعرض محتويات الفيديو من الأمور الأساسية المطروحة في أدبيات الرؤيا الحاسوبية، لما لها من أهمية في تقليص الحجم اللازم لتخزين الفيديو في مختلف الوسائط، وبالأخص في الهواتف النقالة وكميرات المراقبة، وتقليص في الوقت اللازم لمشاهدة الفيديو.

تتلخص عملية الاختزال الذكية، ببناء برمجية قادرة على عرض وتخزين المحتوى الهام من المشاهد، التي تحتوي على تفاصيل متجددة، إما من ناحية الصورة، أو من ناحية الصوت المرافق، وحذف المشاهد ذات المحتوى المتكرر من التفاصيل.

تم في هذا البحث تقديم منهجية عمل جديدة لاستخلاص المشاهد ذات التفاصيل الجديدة في الصورة والصوت، دون التأثير على استمرارية الحركة ضمن الفيديو، وبشكل يضمن مشاهدة مستمرة؛ حيث اعتمدت منهجية العمل على خوارزميتين أساسيتين: الخوارزمية الأولى تعمل على استخلاص المشاهد ذات التفاصيل المتغيرة في الصورة، بالاعتماد على القيم الذاتية للمشاهد، التي تبدي تغير كبير يلازم أي تغير في تفاصيل المشهد، بينما الخوارزمية الثانية تعمل على استخلاص الصوت ذو التفاصيل المتغيرة، معتمدة على خوارزمية مقدمة عام 1985 من [1]، التي يمكنها أن تقنّع الإشارة المدروسة بغلاف ثنائي القيمة 1 أو 0، في منطقة الإشارة المحتوية على تفاصيل يأخذ القيمة 1 ، بينما في المنطقة غير المحتوية على تفاصيل يأخذ القيمة 0. يتم تنفيذ الخوارزميتين بشكل متزامن، وبالتالي يتم استخلاص المشاهد المتغيرة، والإشارة الصوتية المرافقة لها.

تمّ تطبيق منهجية العمل على مقاطع فيديو كبيرة ومتنوعة من حيث حركة الأغراض ضمنها، وحققت فعالية جيدة جداً، محققة دقة كبيرة في التزامن بين المشاهد، والصوت المرافق لها.

الكلمات المفتاحية: استخلاص المشاهد الهامة، القيم الذاتية، الهرم الغاوسي، نايكويست– شانون.

* مدرس في قسم الذكاء الصنعي ومعالجة اللغات الطبيعية ـ كلية الهندسة المعلوماتية- جامعة دمشق

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (4) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (4) 2018

Intelligent System to Reduce Size and Time of Video Display

Dr. George Karraz *

(Received 10 / 6 / 2018. Accepted 9 / 8 / 2018)

□ ABSTRACT □

Smart shorthand, to display video content, is one of the main problems in computer vision literature, because it is important to reduce the size of video storage in various media, especially in mobile phones and monitoring cameras, and reduce the time needed to watch video.

The smart shorthand process is to build software capable of displaying and save important content from the viewer, which contains new details, either in terms of the image or in the accompanying voice and deleting scenes with repeated content.

In this research, a new methodology was introduced to extract new scenes in the image and sound, without affecting the continuity of motion within the video, and in a manner that ensures continuous viewing. The methodology relied on two basic algorithms: the first algorithm works to extract scenes with variable details in the image, based on the eigenvalues of the scenes, which show a significant change in the details of the scene, while the second algorithm is based on the extraction of sound with variable details, based on the algorithm introduced in 1985 from [1], which can encode the sound signal with a double-value frame 1 or 0, in the signal area containing details that takes value 1, while in the non- Details takes value 0, the two algorithms are executed synchronously, and thus the variable scenes and the adjacent acoustic signal are drawn.

The methodology used to work on large video clips in terms of movement of objects within them has achieved very good effectiveness, great accuracy in synchronization between the scenes and sound adjacent to them.

Keywords: Important Scenes Extraction, Eigenvalues, Gaussian Pyramid, Nequist-Shannon.

* **Assistant professor- Department of Artificial Intelligence & Natural Languages Processing- Faculty of Information Technology Engineering- Damascus University.**

مقدمة:

في عصر السرعة الذي نحن فيه، قد لا يكون هناك متسع من الوقت للمرء لمشاهدة فيديو، قد يصل طوله لساعتين لمعرفة محتواه، إن كان ذلك فيلم على سبيل المثال، أو كان الفيديو عبارة عن تسجيل لكاميرا مراقبة، فإن مشاهدته كله، يعتبر هدر كبير في الوقت حتى نفهم طبيعة ما جرى خلال 24 ساعة ماضية، وهناك أمثلة عديدة على ذلك...، فكيف يمكن لنا من خلال وقت قصير فهم محتوى فيديو طويل؟

لقد قدمت لنا معالجة الفيديو العديد من الطرائق منها: ما هو لضغط الفيديو لتقليل حجم تخزينه، ومنها طرائق تقسيم الفيديو، أو تتبعه، أو طريقة فهرسته، ولعل طريقة فهرسة الفيديو هي الأفضل لتسهيل استرجاع المحتوى على نحو فعال، وتصفح المعلومات المرئية المخزنة في قواعد بيانات الوسائط المتعددة، ولإنشاء فهرس فعال، تختار مجموعة من الإطارات المفتاحية(key frames) النائبة، التي تلتقط محتوى الفيديو بالكامل، وتغلفه [2]، ورغم ذلك فإن عملية استخراج تلك الإطارات من الفيديو الأصلي، لم تؤدي الغرض منها في إظهار المحتوى الكامل للفيديو بشكل جيد؛ حيث كانت هناك محاولة لذلك من قبل [3]، إذ تم الاعتماد فقط في استخراج الإطارات الرئيسة، وكشف حدود اللقطة، على استخدام القيم الذاتية (Eigenvalues) لمصفوفة التباين المشترك، بدون الأخذ في الحسبان معالجة الحركة المستمرة، مما أظهر الديناميكية والحركية في الفيديو المنتج بشكل غير جيد.

كما قام [4] بتقديم تقنية لتلخيص الفيديو؛ تقضي بالتوليد التلقائي لفيديو مدمج، من خلال معالجة فيديو طويل، واستخلاص الأنشطة الهامة فيه، مع المحافظة على الحركة المستمرة، والديناميكية للفيديو الأصلي، فتكون نتيجة عرض الفيديو الجديد: هي أمثلة متعددة في ذات الوقت لكل نشاط في الفيديو الأصلي، بالرغم من أهمية هذه الفكرة، إلا أنها تحتاج لوقت طويل في التنفيذ، إضافة إلى عمليات معالجة كبيرة، وتستخدم في مجالات خاصة فقط.

قبل عام 2010 ، كان هناك طريقة لتتبع الأغراض تعتمد على القيم الذاتية، ومصفوفة اللحظة الثانية [5]؛ حيث تستطيع القيم الذاتية تمييز غرض من آخر، وذلك لأنها فريدة بالنسبة لكل غرض من الأغراض، وعليه يمكن اكتشاف عدد الأغراض ضمن النافذة المأخوذة، وبالاعتماد على التباين بين القيم الذاتية يمكن اختيار الإطارات المفتاحية من الفيديو.

تستخدم المنهجية المطورة في هذا البحث، القيم الذاتية، إضافة إلى الهرم الغاوسي لقياس التباين بين الأطر، كما تقوم بمعالجة الصوت للإطارات المفتاحية الهامة، ومن ثم تقوم بمزامنة الصوت المعالج مع الإطارات المفتاحية الهامة المستخرجة، وبذلك نحصل على الفيديو النهائي.

أهمية البحث وأهدافه:

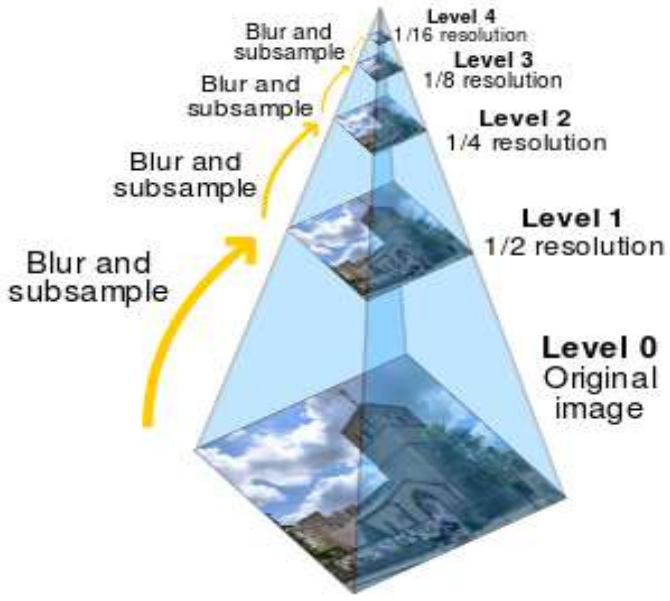
يتلخص الهدف من هذا البحث بإنتاج فيديو جديد يماثل تماما الفيديو الأصلي من ناحية الدقة والتفاصيل المعروضة بهما، ولكن أقل حجماً وأقصر زمناً أثناء عرضه، بالإضافة إلى معالجة الصوت الأصلي ليصبح مطابقاً، ومتزامناً مع المشاهد المعروضة في الفيديو الجديد.

. المبدأ النظري

1. الهرم الغاوصي Gaussian Pyramid –

التمثيل الهرمي الغاوسي: هو نوع من تمثيل الإشارة متعددة القياس، مطور من قبل المهتمين في الرؤية الحاسوبية، ومعالجة الصور، ومعالجة الإشارة [5]، وفيه تخضع الإشارة إلى تنعيم متكرر واخذ عينات جزئي، والتمثيل الهرمي هو أصل التمثيل بفضاء القياس (scale-space representation)، والتحليل متعدد التمييز (multi-resolution).

يوجد نوعان رئيسيان للهرم: هرم تمرير منخفض(low pass)، وهرم تمرير حزمة (band pass)؛ حيث هرم التمرير المنخفض يتم من خلال تنعيم الصورة بمرشح تنعيم مناسب، ثم إجراء اعتيان جزئي للصورة المنعمة بمعامل قدره 2، على طول اتجاه إحداثيات، ثم تخضع الصورة الجديدة للإجراء نفسه مرة أخرى، ومن الممكن عدداً من المرات، نسمي كل واحدة منها دورة ينتج عنها صورة جديدة، أكثر نعومة وأقل كثافة اعتيان مكاني، فإذا رتبنا الصور الناتجة والصورة الأصلية؛ بحيث تكون الصورة الأصلية في أسفل الشكل البياني، تتلوها الصورة الأصغر الناتجة عن أول دورة، ثم الناتجة عن الدورة الثانية، وهكذا مكدسة فوق بعضها، لتشكل بنية على هيئة هرم كما في الشكل (1).



الشكل (1) التمثيل المرئي لهرم الصورة بخمسة مستويات

هرم تمرير الحزمة يشكل من خلال إيجاد الاختلاف بين الصور، عند مستويات دقة متجاورة في الهرم، وإجراء نوع من استيفاء الصورة، عند تلك المستويات لحساب الفروقات من ناحية عنصر الصورة.

في الهرم الغاوسي يتم استخلاص الصورة التالية من الصورة، التي في المستوى الأدنى، بإخضاعها لمتوسط غاوص(blurring) مع تخفيض في حجمها؛ بحيث يكون كل عنصر في الصورة التالية، هو متوسط عناصر الصورة المحيطة به في الصورة الموجودة في المستوى الأدنى من الهرم.

2 .نظرية نايكويست-شانون في أخذ العينات

نظرية نايكويست – أو ما يعرف بنظرية أخذ العينات – مبدأ يتم اتباعه لتحويل الإشارات التماثلية إلى رقمية analog-to-digital converter (ADC)؛ حيث تصبح الإشارة الناتجة بعد التحويل إشارة رقمية، وتمر الإشارة التماثلية قبيل عملية التحويل بمرحلة تسمى عملية أخذ العينات sampling؛ حيث تؤخذ قيم الإشارة التماثلية في فواصل زمنية

متعاقبة متكررة ومتساوية، ونسمي عدد مرات أخذ العينات في الثانية بتردد أخذ العينات (sampling frequency)، أو معدل أخذ العينات (sampling rate)، وتعتبر هذه المرحلة ضرورية لإعادة إنتاج الإشارة التماثلية بموثوقية. [6].
تتألف أية إشارة تماثلية من مجموعة من المركبات الترددية، وكل مركبة ترددية منها عبارة عن إشارة جيبية، لها مطال معين وتردد معين– وقد اختيرت الإشارة الجيبية لأنها أبسط أنواع الإشارات التماثلية؛ حيث تعتبر كل طاقة الإشارة مركزة في تردد واحد– والمركبة ذات التردد الأعلى، هي من تحدد عرض المجال للإشارة التماثلية في حال كانت باقي العوامل ثابتة.

إذا افترضنا أن المركبة الجيبية ذات التردد الأعلى في الإشارة التماثلية ترددها F_{max} ، فوفقاً لنظرية نايكويست، فإن معدل أخذ العينات يجب أن يكون على الأقل مساوياً $2F_{max}$، أو أكبر من ذلك، أما إن كان أقل من ضعفي التردد الأعظمي، فإن المركبات الترددية العالية في الإشارة التماثلية لن يتم تحويلها إلى إشارة رقمية بشكل صحيح، وبالتالي عند تحويل الإشارة الرقمية مرة أخرى إلى تماثلية من خلال المحول DAC، ستظهر مركبات ترددية لم تكن موجودة في الإشارة التماثلية الأصلية، مما يؤدي إلى تشوه الإشارة، وهذا التشوه يعرف بالتزييف aliasing.

3 . العمل على الصورة

لقد اخترنا في بداية عملنا القيم الذاتية للعمل على الصورة ؛ وذلك لأن لها قيمة وحيدة وفريدة لكل غرض موجود ضمن الصورة، وعليه يمكننا تحديد أي الإطارات ضمن الفيديو هي الإطارات المفتاحية، ولكن كما هو معلوم، فإنه كلما كان حجم الصورة أكبر كانت القيم الذاتية أكبر، واحتاجت إلى زمن أطول لحسابها، لذلك لجأنا إلى استخدام الهرم الغاوسي لكل إطار؛ حيث باستخدامه نقلل من حجم الإطار، وبالتالي نخفف من عدد القيم الذاتية المحسوبة، ونوفر في الوقت اللازم لذلك، فلو كان لدينا حجم الإطار الواحد 300x300 ، فإن القيم الذاتية لكل إطار قد تكون 300 قيمة، وستكون درجة المعادلة اللازمة لذلك كبيرة، أما باستخدام الهرم الغاوسي سيصبح حجم الإطار 100x 100 ، وعليه قد تكون القيم الذاتية 100 قيمة فقط.

3 – 1. صعوبات العمل على الصورة

أهم الصعوبات التي واجهتنا، أنه لا يمكن حساب القيم الذاتية إلا لإطار ذي أبعاد مربعة (ارتفاع الإطار يساوي عرضه)، وأغلب أطر الفيديو مستطيلة، إضافة إلى أنه عند حساب الهرم الغاوصي لإطار ما، قد يكون في ذلك الإطار تفاصيل هامة تضيع نتيجة تطبيق هذا الهرم، وكذلك الحركة ضمن الفيديو. لحل هذه المشكلات هناك العديد من الطرائق يمكن سردها قبل أن نطرح الحل الذي نراه مناسباً:

1 – لا يمكن إضافة أصفار حشو إلى أعمدة أو صفوف الإطار لجعله مربعاً، لأن ذلك سيضيف شكلًا أو غرضاً غير موجود أصلاً في الإطار الأصلي.

2 – إذا كان الإطار مستطيلا مثلا ذي أبعاد 400×300، يمكن أن نأخذ جزءً مربعاً منه كنافذة أبعادها 300×300 وحساب القيم الذاتية لها، ومن ثم أخذ نافذة أخرى مزاحة بمقدار عنصر صورة واحد، ثم حساب القيم الذاتية وهكذا، وفي النهاية أخذ اجتماع كل القيم الذاتية، إلا أن تلك الطريقة تحتاج زمناً طويلاً، علاوة عن الكم الهائل من القيم الذاتية الناتجة عن الإطار الواحد، مما يستهلك موارد كبيرة من الذاكرة.

3 بفرض أن الأطر مربعة، أي أن مسالة حساب القيم الذاتية محلولة، فإن عدد القيم الذاتية في الإطار الواحد تعبر عن عدد الأغراض الموجودة في هذا الإطار، وبمقارنة القيم الذاتية بين الإطار الحالي والسابق، نستطيع أن نختار الإطار المفتاح، عندما يكون الفارق بين الإطارين في القيم الذاتية كبيراً نسبة لعتبة نحن نختارها، وإلا فإن الإطارين

تقريباً يحويان ذات الأغراض أو الأشياء، ويمكننا من خلال تلك العتبة، تغيير معيار اختيار الأطر المفتاحية، فلو كان لدينا في أحد المشاهد تفاصيل دقيقة يجب إظهاره، ككلمات صغيرة في مقطع ما من الفيديو الأصلي، وأردناها أن تظهر في الفيديو الجديد، فإننا بتصغير العتبة يمكننا أن نحصل على الأطر الحاوية لتلك الكلمات.

لكن هل هناك رابط بين العتبة والمستوى في الهرم الغاوسي؟ إن التفاصيل الدقيقة والحركة في الفيديو تتطلب العديد من الأطر، لذلك لا يجب تقليل حجم الإطار، وبالتالي سيكون مستوى الهرم الغاوسي صغيراً، وعندها سيكون عدد القيم الذاتية في كل إطار كبيراً، أما العتبة فيمكن أن تكون كبيرة أو صغيرة اعتمادا على محتوى الفيديو، ودقته وعوامل أخرى، لتحديد هذه الرابطة هناك طريقتان:

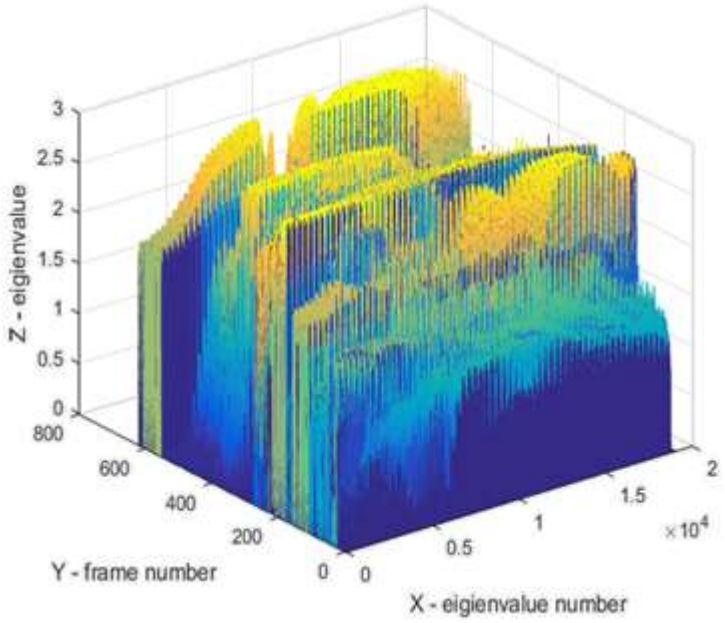
أ- تحديد عدة أنواع للفيديو؛ حيث يتمتع كل نوع بمجموعة من المزايا، تقتضي اختيار عتبة مقارنة ومستوى هرمي معينين، فمثلا لو كان الفيديو بقياس $n \times m$ ، وفيه الكثير من التفاصيل، فإن هذه شروط معينة تميز هذه الفيديو على أنه من النمط type1، وتقتضي باستخدام عتبة T1 للمقارنة بين القيم الذاتية، و L1 كمستوي للهرم الغاوسي، لكن هذه الطريقة تتطلب وقتا طويلا في البحث في كل الأنواع، وهناك احتمال كبير لنسيان إدراج بعض الأنواع.

ب- استخدام خوارزميات التعلم التلقائي، لتوقع قيمة العتبة، ومستوى الهرم المناسب لفيديو محدد، ولكن هذه الطريقة تحتاج لوقت طويل، بالإضافة إلى الكم الهائل من الفيديوهات الواجب مشاهدتها لتنفيذ أداة التعلم التلقائي الذكية.

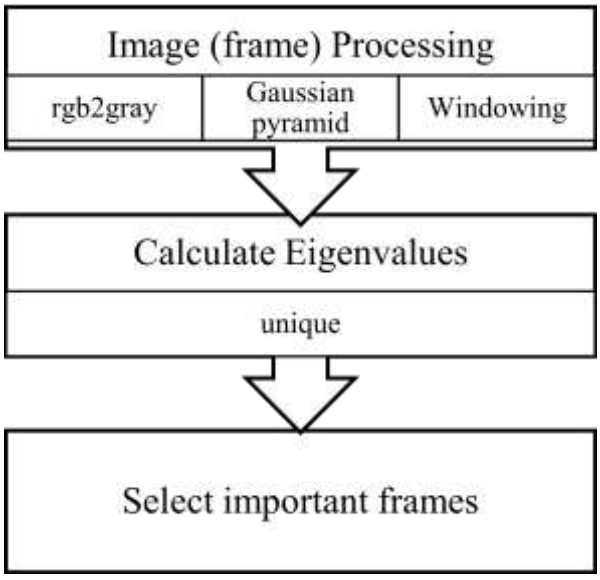
3 – 2 . الطريقة المقترحة في الفيديو

لحل تلك المشكلات السابقة، قمنا بإضافة مرحلة هامة، وهي تقسيم كل إطار إلى مجموعة من النوافذ الصغيرة المربعة (نافذة بقياس 3×3)، وبهذا نكون قد حللنا مشكلة الأطر المستطيلة، عندها كل نافذة ستمتلك 3 قيم ذاتية، وبما أن الإطار الواحد قد يحوي ذات الغرض بأماكن مختلفة، فإن القيم الذاتية قد تتكرر لأكثر من نافذة ضمن الإطار الواحد، وذلك لأن القيم الذاتية للغرض ذاته تكون فريدة، مهما كان مكان الغرض ضمن الإطار، ويمكننا بتقريب معين حذف القيم المتكررة؛ بحيث يبقى لدينا ضمن كل إطار من 7 إلى 8 قيم ذاتية فقط، ويظهر الشكل(2) القيم الذاتية المحسوبة وفق هذه الطريقة لجميع الأطر ضمن الفيديو.

بعد حساب القيم الذاتية للإطارات، نقوم بمقارنة القيم بين الإطار الحالي والسابق، وذلك لاستنتاج الأطر المفتاحية، التي ستكون في الفيديو الجديد، وهذه الطريقة تضمن أن جميع الحركات والأحداث، التي كانت في الفيديو الأصلي ستظهر في الفيديو الجديد، ويبين الشكل (3) ذلك.



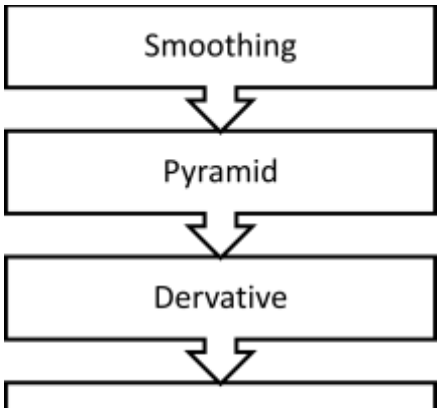
الشكل (2) القيم الذاتية لجميع الأطر في الفيديو وفق الطريقة المقترحة



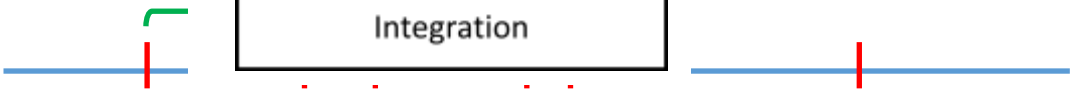
الشكل (3) الخوارزمية المقترحة لاستخلاص الأطر المفتاحية

2 – 4 . العمل على الصوت

اعتمدنا لمعالجة الصوت على نظرية نايكويست في الاعتيان؛ حيث يجب أن يكون معدل الاعتيان ضعفي أكبر تردد في الصوت $2\times f_{max}$، وقد تبين بأن أفضل تردد يمكن اعتماده من قبلنا هو ربع تردد الاعتيان الذي يعتمد في الملفات الصوتية، والمساوي 44100Hz. لقد اقترحنا لمعالجة الصوت خوارزمية مؤلفة من ست خطوات مبينة في الشكل (4).



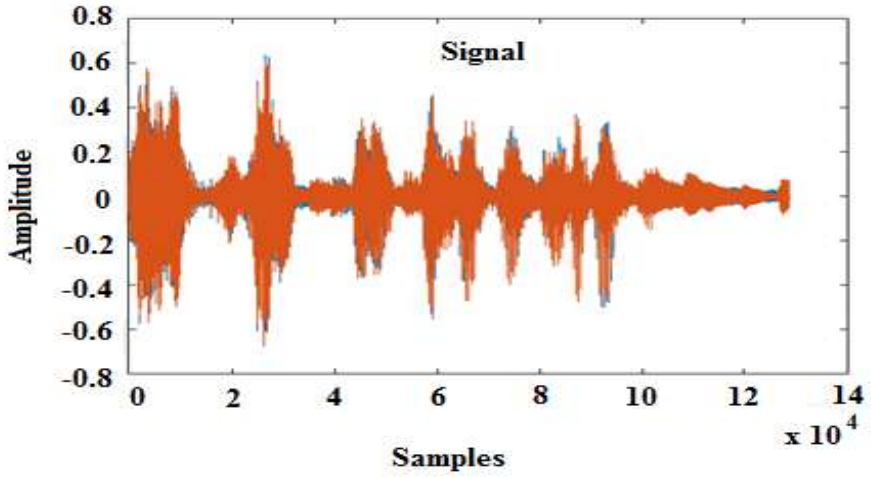
الشكل (4) الخوارزمية المقترحة لمعالجة الصوت



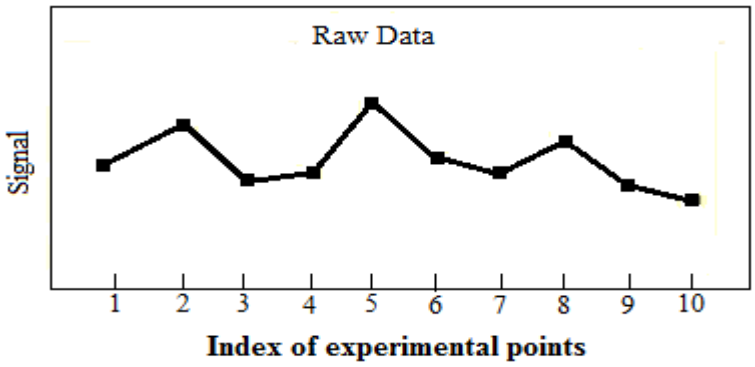
الشكل (5) رسم تمثلي للفيديو الأصلي والأطر المفتاحية المختارة.

تبدأ الخطوة الأولى في هذه الخوارزمية من حيث انتهينا في معالجة المشاهد، أي بعد مرحلة استخلاص الإطارات المهمة، فكما مبين في الشكل(5)، فإن الفيديو الأصلي ممثل على شكل خط أفقي، بينما الخطوط الشاقولية القصيرة تمثل مناطق الإطارات المهمة، ستتم معالجة الصوت الموجود فقط بين الإطارات المفتاحية الحمراء المختارة؛ بينما الصوت الموجود قبل الإطار المفتاحي الأول فلا حاجة لمعالجته لأن الإطارات المرتبطة معه تم حذفها، وكذلك الأمر ينطبق على الصوت الموجود بعد الإطار المفتاحي الأخير، كما يبين الشكل (6) مثال على إشارة الصوتية الأصلية، المراد معالجتها بالخوارزمية المؤلفة من الخطوات الست التالية:

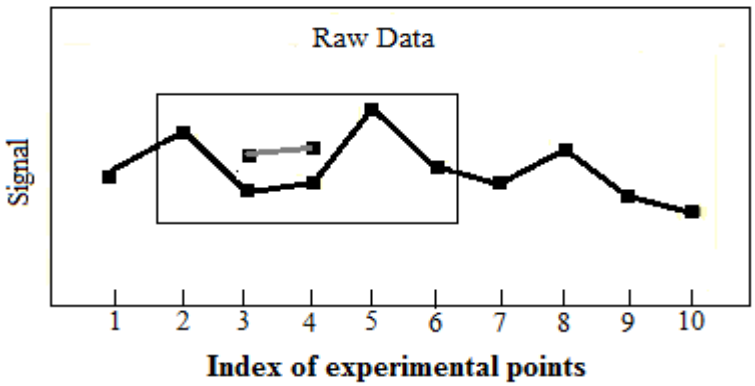
1 . التنعيم smoothing: التقنية الأبسط في تنعيم الإشارة الصوتية المعتانة بمسافات زمنية متساوية، هو المتوسط المتحرك moving average، حيث بواسطته تتحول الإشارة المضجة إلى إشارة منعمة، فالنقطة المنعمة y_k: هي متوسط لمجموعة النقط المتتابعة من الإشارة المضجة قبل وبعد تلك النقطة، والتي عددها $2n+1$ والممثلة على الشكل $(y_k,y_{k+1},\dots,y_{k+n-1},y_{k+n}, y_{k-n},y_{k-n+1},\dots,y_{k-1},)$، وفي الخوارزمية المقترحة ستكون $n=w/2$ ؛ حيث $w=F_s/4$، وسنسمي w عرض المرشح ذي المتوسط المتحرك، وكلما كان عرض المرشح كبيراً كان التنعيم أفضل، وتشرح الأشكال المتعاقبة (7 و8 و 9) هذه التقنية في التنعيم.



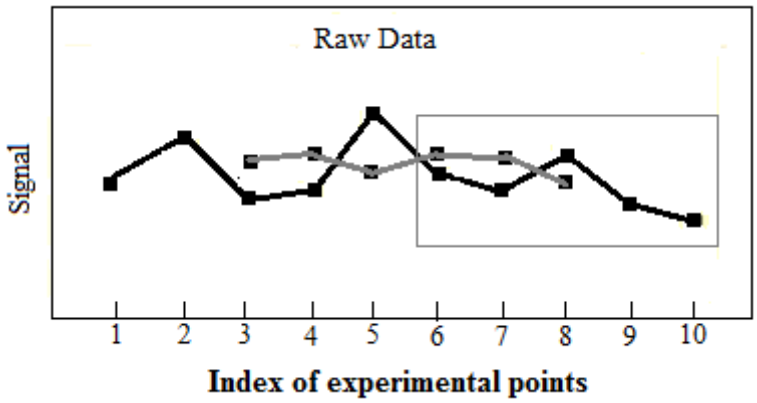
الشكل (6) إشارة الصوت الأصلية بعد أن تم إجراء تنظيم لقيم السعات لتكون محصورة في المجال [1 -1]



الشكل (7) الإشارة المضجة

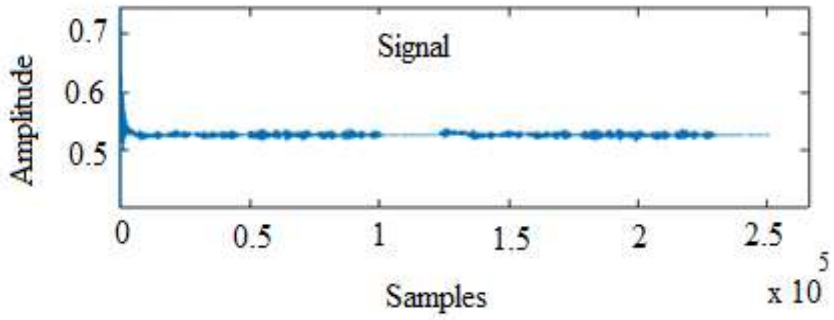


الشكل (8) المتوسط المتحرك بالتقسيم إلى نوافذ.



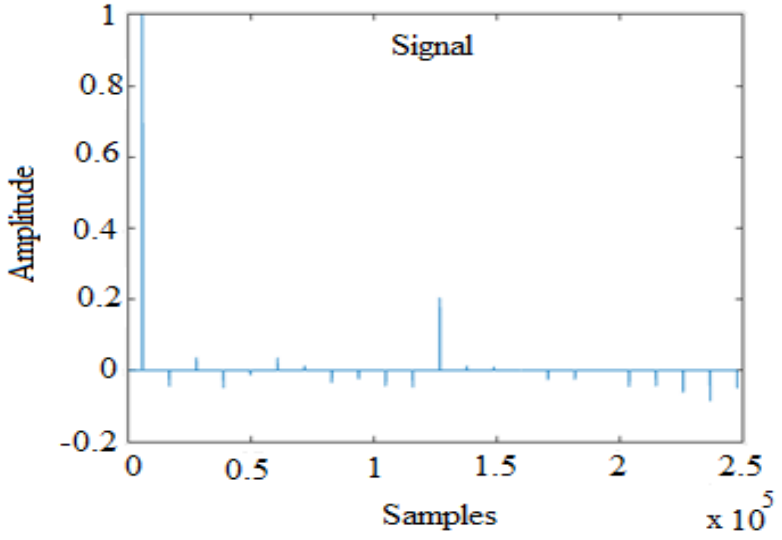
الشكل (9) الإشارة المنعمة الناتجة بلون أخضر

في هذا المثال عرض المرشح w=5، فالنقاط الخمسة الأولى من الإشارة المضجة الممثلة بمربعات سوداء، تدخل ضمن المستطيل، الذي يمثل نافذة المرشح ذات العرض 5، والتي تمثل مرشح المتوسط المتحرك؛ حيث تؤخذ قيمة المتوسط لتلك النقاط، وترسم تلك القيمة كنقطة خضراء ترتيبها ثلاثة، وهكذا تزاح نافذة المرشح بمقدار عينة واحدة؛ بحيث يؤخذ متوسط العينات من 2 إلى 6 ، ويرسم متوسطها كنقطة خضراء ترتيبها 4، وهكذا من أجل جميع القيم، حتى نحصل على الإشارة المنعمة النهائية الممثلة بمربعات خضراء، تكون فيها المركبات ذات التردد العالي ضعيفة المطال، بينما الترددات المنخفضة مركباتها أعلى مطالاً، وذلك لأن المرشح ذي المتوسط المتحرك مرشح تمرير منخفض. ويبين الشكل (10) الإشارة المنعمة الناتجة من تمرير الإشارة الأصلية في الشكل (6) على مرشح التنعيم.



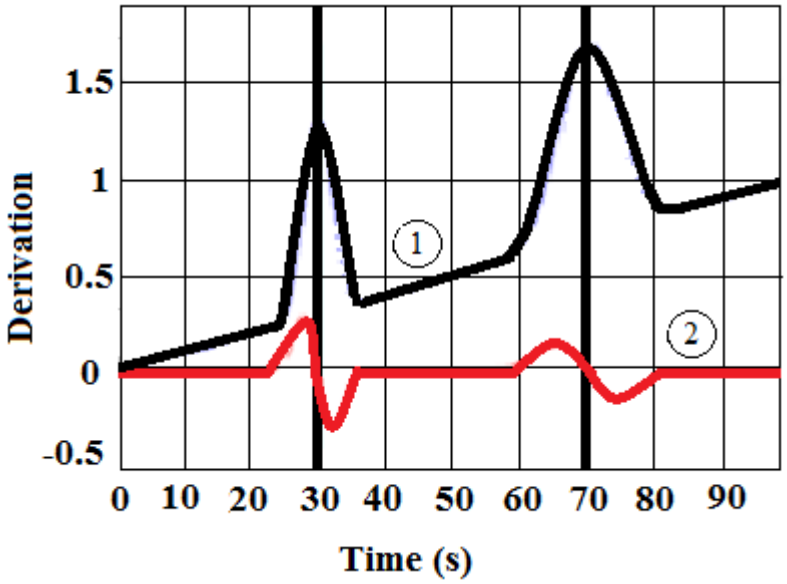
الشكل (10) إشارة الصوت المنعمة النهائية

2 الهرم: هذه المرحلة يطبق الهرم الغاوسي على الإشارة المنعمة، الناتجة من المرحلة السابقة، تماماً كما في الصورة، ولكن هنا لبعد واحد؛ حيث سيقلَ بنتيجة هذه المرحلة عدد عينات الإشارة، تماماً كما في مرحلة التنعيم، ولكن بدل التحرك عينة عينة، تتم معالجة كل مجموعة من العينات؛ بحيث تمثلها عينة واحدة، بالنتيجة نحصل على إشارة رقمية بعدد عينات أقل، مما سيجعل المعالجة أسرع في المراحل القادمة، ويبين الشكل (11) مرحلة الهرم التي تلي مرحلة التنعيم.



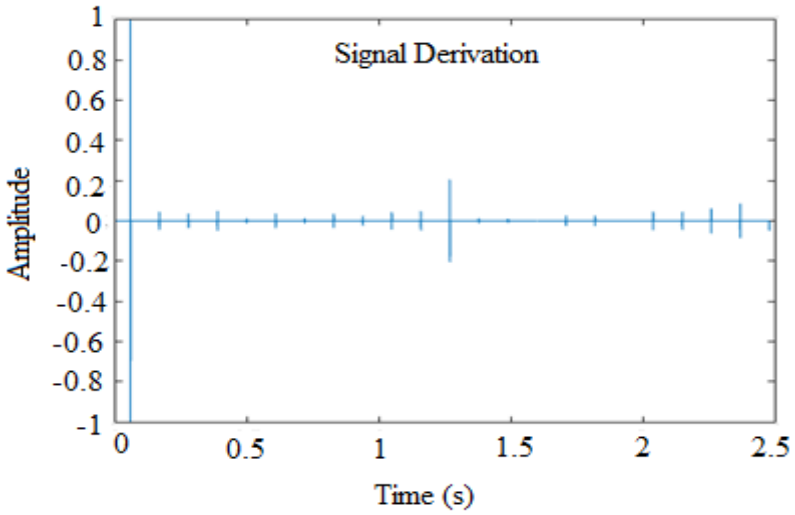
الشكل(11) تطبيق مرحلة الهرم على الإشارة المنعمة

3 . الاشتقاق: يستخدم المشتق عادة لمعرفة القمم في الإشارة، فالمشتق الأول عند قمة من قمم الإشارة ينحدر ماراً بالصفر عند تلك القمة، وهذا ما يساعد على تحديد قيمة الإحداثي x لتلك القمة (في الإشارة الصوتية الإحداثي x هو الزمن)، ويبين الشكل (12) هذا الأمر، فإذا لم يكن هناك ضجيج في الإشارة، فإن أية قيمة في الإشارة، أعلى من التي قبلها والتي تليها تدل على أن هذه النقطة هي قمة عظمى.



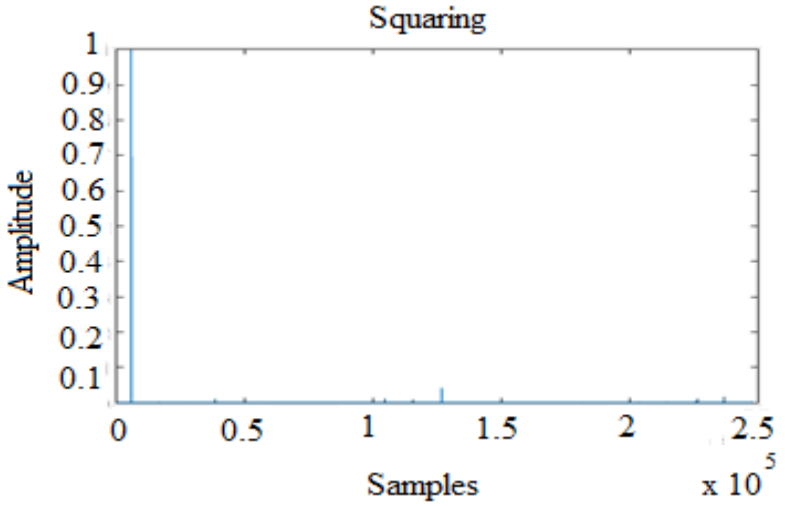
الشكل(12) مواقع القمم العظمى، (2) يمثل المشتق الأول للتابع (1).

ويبين الشكل (13) الاشتقاق للإشارة الناتجة من مرحلة الهرم.



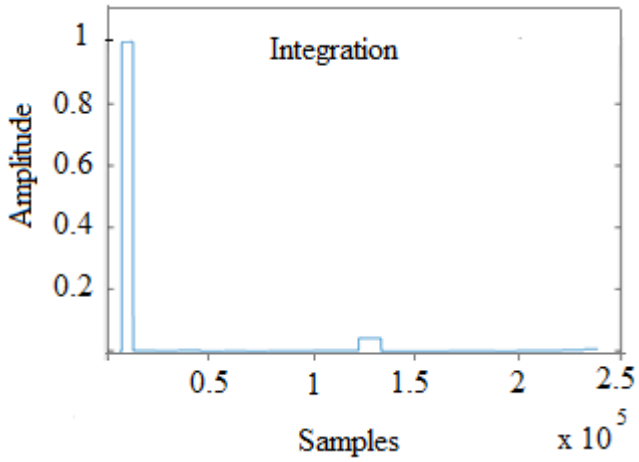
الشكل (13) الإشارة الناتجة من مرحلة الاشتقاق

4 . التربيـــع: تطبـق هـذه المرحلـة علـى الإشـارة الناتجـة مـن الاشـتقاق، وذلـك لزيـادة الفـارق بـين العينـات؛ حيـث تتحـول القـيم السـالبة للعينـات بتربيعهـا إلـى قـيم موجبـة، بينمـا تصـبح القـيم الصـغيرة (التي بين 1- و 1) أصغر، ويبين الشكل (14) الإشارة الناتجة من مرحلة التربيع.



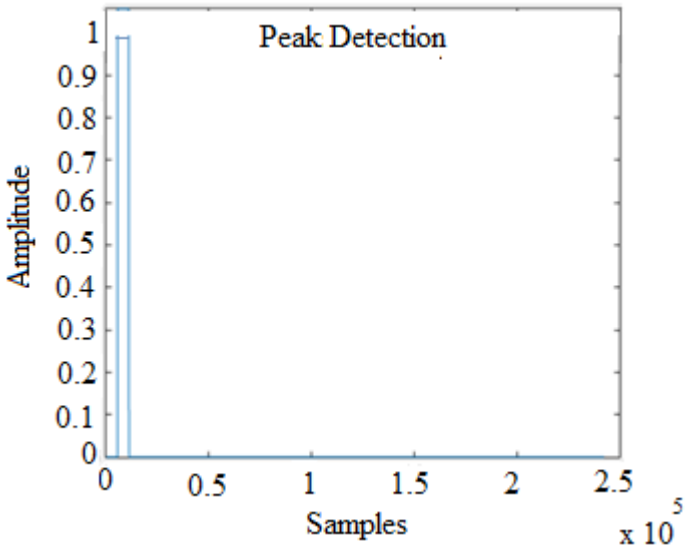
الشكل (14) الإشارة الناتجة من مرحلة التربيع

5 . التكامل: عادة ما يستخدم التكامل لمعرفة مقدار المساحة، فإذا أخذنا نافذة كما في حالة التنعيم، وحسبنا المتوسط، فإن كل نقطة ناتجة عنه تعبر عن المساحة المحسوبة ضمن تلك النافذة، وكلما كان مقدار المساحة أكبر، كان احتمال أن تكون تلك النافذة عقدة، ويظهر الشكل (15) مرحلة التكامل بعد التربيع.



الشكل (15) مرحلة التكامل

6 . كشف القمة: بتحديد عتبة معينة، ومقارنتها مع جميع العينات بعد مرحلة التكامل، سيجعل بالإمكان تحديد العينات التي ستبقى، والعينات التي ستعتبر ضجيجا، وبالتالي يجب حذفها، مع العلم أن كل عينة من العينات التي ستبقى، هي ممثلة (نائبة) لمجموعة العينات المختارة، بنافذة التكامل في المرحلة السابقة، وبالتالي يجب العودة بعملية عكسية، لإيجاد تلك العينات، والتي سنطبق عليها أيضاً عملية عكسية لأنها عينات نائبة في مرحلة الهرم، وكذلك نطبق عملية عكسية على العينات الناتجة، لأنها عينات نائبة في عملية التنعيم، أي أننا سنطبق عملية عكسية ثلاث مرات تبعاً لكل مرحلة، في النهاية نحصل على الصوت النهائي للأجزاء المهمة بين الإطارات المفتاحية، ويبين الشكل (16) القمم النهائية قبل إجراء العمليات العكسية.



الشكل(16) كشف القمة

2 – 5 . ملء الصوت والصور

في بعض الفيديوهات توجد مشاهد –مجموعة أطر– بدون صوت، أو صوت بمشاهد ثابتة (إطارات لا توجد فيها حركة)، لذلك يجب معالجة هذه المشكلة بإضافة صوت في أماكن الإطارات التي لاتحوي صوت، وإضلفة إطارات ثابتة في المناطق المتزامنة مع إشارة الصوت حيث لايوجد أطر الثابتة.

2– 5 – 1. ملء الفيديو

في هذه المرحلة تضاف إطارات فيديو كإطارات مفتاحية، في حال كان لدينا صوت، ولم يكن لدينا إطارات مرافقة ومرتبطة بهذا الصوت. فلو فرضنا بأنه لدينا 100000 إلى 200000 عينة مرتبطة بأطر ثابتة، فإنه عند إجراء الخوارزمية، سيكون هناك إطار مفتاحي واحد نائباً عن تلك الإطارات، وبهذا سيكون زمن الصوت أطول من زمن عرض هذا الإطار، لذلك نجري عملية ملء الإطار المفتاحي عدداً من المرات تحسب كما يلي:

$$Nf = \frac{Ns \times F_{rate}}{F_s} \qquad (1)$$

حيث Nf: هو عدد الأطر المفتاحية المراد إضافتها، Ns: هو عدد عينات الصوت المرتبط بالأطر الثابتة، F_{rate}: هو معدل الإطار عادة ما تكون قيمته 25 إطاراً في الثانية، بينما F_s: معدل اعتيان الإشارة الصوتية وعادة ما تكون قيمته 44100Hz.

2– 5 – 2 . ملء الصوت

عندما يكون لدينا إطارات صامتة، فإننا نحتاج لملء تلك الإطارات بعينات صوتية صامتة، أي بمركبات ترددية صفرية، وذلك لجعل زمن الصوت مساوياً لزمن الأطر المفتاحية، ونملأ الصوت بناءً على المعادلة التالية:

$$Ns = \frac{Nf \times F_s}{F_{rate}} \qquad (2)$$

حيث N_s : عدد العينات الصوتية المراد إضافتها لكل الإطارات الصامتة N_f. أي أن كل إطار صامت سيحصل على عدد من العينات مقدارها:

$$s = \frac{F_s}{F_{rate}} \qquad (3)$$

ولكي تكون العينة الصوتية الصامتة في وسط الإطار، فإننا نضيف عدد من العينات الصوتية مقدارها S/2 إلى يسار العينة، و S/2 إلى يمينها.

2 – 6 . دمج الفيديو والصوت

بعد جعل كل من الصوت والفيديو بنفس الطول الزمني، بإمكاننا دمجهما معاً، لتشكيل الفيديو النهائي الذي لا يكون فيه الصوت يسبق الإطارات المرتبط معها، أو يتأخر عنها، وهذا الفيديو الناتج الجديد سيكون أقل عدداً من الأطر، والعينات الصوتية، وبالتالي أقل حجماً من الفيديو الأصلي.

3 . مرحلة التصميم والإنجاز

في مرحلة التصميم قمنا باستخدام الماتلاب للتحقق من عمل الخوارزمية المقترحة بشكل جيد، بعدها قمنا باستخدام البرمجة بلغة جافا لصناعة البرنامج النهائي؛ حيث استخدمنا مجموعة من المكتبات المدرجة مع الجافا، والخاصة بالفيديو والصوت وهي:

- OpenCV: تستخدم لمعالجة الفيديو، وتحتوى توابع التحويل من ملون إلى رماديات وتابع الهرم الغاوصي، وتابع حساب القيم الذاتية وتوابع حفظ الفيديو وقراءته.

- JAMA(Java Matrix): تحتوي جميع التوابع التي تطبق على المصفوفات، مثل ضرب المصفوفات، مقلوب المصفوفات ...إلخ، وذلك لتسهيل العمل على المصفوفات.
- JAVE(Java Audio Video Encoder): تستخدم هذه المكتبة لقراءة الصوت من الفيديو كملف wav، وذلك لكي يكون من المكن معالجته.
- WavFile وWavFileException: هما صفان يستخدمان لقراءة العينات من أو كتابتها في ملف الصوت من نوع Wav، يستخدمان في مرحلة معالجة الصوت.
- FFmeg: تستخدم التوابع في هذه المكتبة لدمج الصوت والفيديو الجديدين معاً.
- ImShow: هو صف يستخدم في إظهار الإطارات في الزمن الحقيق قبل حفظها على شكل فيديو.

الفيديو الناتج يحفظ بلاحقة avi، حتى يكون الفيديو الجديد أصغر حجماً من الأصل، يجب أن يكون الأصل من ذات اللاحقة avi، إذا كان الفيديو الأصلي من غير لاحقة فقد لا نجد فرقاً كبيراً في الحجم.

النتائج والمناقشة

لقد كان الهدف الرئيسي من هذا البحث، تقليص زمن الفيديو الأصلي من خلال اقتراح خوارزمية لتحقيق ذلك؛ حيث اختبرت هذه الخوارزمية بعد تنفيذها على عدد من الفيديوهات عددها 100 فيلم تتراوح مدتها بين 2 دقيقة، وحتى 10 دقائق متنوعة في المحتوى، وقد استخدمنا القيمة المتوسطة لتقييم الخوارزمية، يبين الجدول (1) ثلاث عينات فلمية من أصل 100 المطبق عليها الخوارزمية الجديدة.

في الجدول(1)، العامود الثاني يحوي عدد الإطارات الكلية، والعدد بين الأقواس يمثل أعداد الإطارات قبل معالجة الصوت، وهو الخطوة الأولى في الخوارزمية، ويظهر في العامود الثالث العدد الكلي للإطارات في الفيديو الناتج، وذلك بعد المزامنة بين الصوت والصور، أما العامود الرابع فيبين العدد الكلي لعينات الصوت في الفيديو، والعامود الخامس يبين عدد العينات الصوتية بعد المعالجة حسب المنهجية المقترحة، في العامود السادس تظهر فترة عرض الفيديو الأصلي، أما ما بين الأقواس فهي فترة الفيديو المعالج ولكن قبل معالجة الصوت(المرحلة الأولى)، بينما يظهر العامود السابع فترة الفيديو الناتج النهائي بعد الدمج مع الصوت، أما الحقلين الثامن والتاسع فيظهران نتائج القياس، فدرجة الفهم تستخدم لمعرفة مدى كون الفيديو الناتج مفهوماً مثل الفيديو الأصلي، دون ضياع المعلومات في كل من الإطارات والصور، بينما درجة التزامن تبين مدى التزامن بين الصوت والإطارات في الفيديو الناتج.

الجدول (1) بعض الفيديوات التي تم اختبار منهجية البحث عليها

1	2	3	4	5	6	7	8	9
Video	**exact frames (before sound)**	**frames after**	**samples before**	**samples after**	**Exact duration (before adding)**	**Dura-tion after**	**Understanding degree (%)**	**Accuracy of synchronization between audio and video (%)**
V1	7543 (1237)	5333	13876224	9763879	5:14 (00:51)	3:42	100%	100%
V2	9642 (2457)	5782	14189568	8494135	5:21 (1:21)	3:12	85%	100%
V3	7412 (2358)	5249	9619814	7288171	5:09 (1:38)	3:38	95%	95%

V1- What causes kidney stones - Arash Shadman.
V2- What Does Lead Poisoning Do To Your Brain.
V3- How does asthma work? - Christopher E. Gaw.

لقد قمنا بتجربة الخوارزمية المقترحة على 100 فيديو؛ حيث تبين النتائج أن القيمة الوسطى لدرجة الفهم تساوي 86.65%، بينما درجة التزامن 99%، وكلما كان الفيديو كبيراً بالحجم كانت الدقة أفضل، لكن هذا يتطلب ذاكرة كبيرة لكي يستطيع الحاسوب معالجته.

الاستنتاجات والتوصيات

لقد عرضنا في هذه المقالة منهجية جديدة لتقليص الفيديو باستخدام القيم الذاتية، ونظرية نايكويست في أخذ العينات، لقد اتسمت الطريقة بعدم تعقيد المعادلات الرياضية، التي كانت تحتاجها طرائق تقليص الفيديو، بالإضافة إلى سرعتها في المعالجة، ولكنها تحتاج ذاكرة كبيرة، وهذه الطريقة أفضل من غيرها من ناحية تزامن الصوت مع الفيديو، بالإضافة إلى فعاليتها في معالجة الصوت.

تمتاز الخوارزمية المطورة بأهمية اقتصادية كبيرة، خاصةً في أنظمة الهواتف النقالة وكميرات المراقبة، فهي توفر الحجم التخزيني اللازم لتخزين الفيديو الطويل وتوفر الوقت اللازم لمشاهدته؛ حيث أن الفيديو الناتج عن تطبيق الخوارزمية يركز بشكل أساسي على المشاهد ذات المحتوى الهام والمتغير من التفاصيل، ويتجنب تكرار المشاهد الثابتة التي تحتوي على نفس التفاصيل.

سيكون عملنا المستقبلي على هذه الخوارزمية في إدخال المرحلة الدلالية عليها، وذلك لتوليد وصف مختصر تلقائياً حول الفيديو؛ حيث بقراءة هذا الوصف، نوفر الوقت اللازم لمشاهدة الفيديو بأكمله، كما يساعدنا ذلك في تسريع عملية البحث عن الفيديو المفيد من بين العديد من الفيديوهات المعروضة. يمكننا استخدام تقنيات معالجة اللغات الطبيعية، وقاموس كبير كفاية يحتوي صورا ومعانيها، إضافة إلى تقنية تعتمد على متوسط القيم الذاتية، وذلك لتوليد هذا الوصف، لكن يجب العمل على هذه التقنية لضمان عملها الصحيح.

المراجع

1. PAN, J., TOMPKINS, W. J. A real-time QRS detection algorithm. IEEE Trans. Biomed. Eng., BME-32(3), 1985 pp 230-236.
2. VIJEETKUMAR, B., DINESH, R., PUNITHA, .P, RAO, V. Key frame extraction and shot boundary detection using Eigenvalues. International Journal of Information and Electronics Engineering vol. 5, no. 1, India, 2015 pp 40-45.
3. ESAKKIRAJAN, S., JAYARAMAN, S., VEERAKUMAR, T. Digital Image processing. Tata McGraw Hill, New Delhi, 2015 pp 719.
4. RODRIGUEZ, M. Compact Representation of Actions in Movies. Computer Vision and Pattern Recognition (CVPR) Conference, San Francisco, CA, USA, 2010 pp 3328 – 3335.
5. SIMON, J.PRINCE, D. Computer Vision. Cambridge University Press, 2012 p580.
6. SMITH, S. W., Digital Signal Processing. Elsevier Science, 2013 p 672.

7. KARRAZ, G., MAGENES, G. Automatic Classification of Heartbeats using Neural Network Classifier based on a Bayesian Framework. Proceeding of the 28th IEEE EMBS Annual international Conference, New York City, USA, FrEp8.10, 2006 pp 4016-4019.
8. PAULETTI, M., MARCHESI, C. Model Based Signal Characterization for Long Term Personal Monitoring. IEEE Computer in Cardiology, 28, 2001 pp.413- 416.