متابعة حية وغير حية لهدف متحرك باستخدام ملف فيديوي MPEG

د. مهى عبد الرحمن حسو نادية معن محمد شهد عبد الرحمن حسو

قسم علوم الحاسوب / كلية علوم الحاسوب والرياضيات

جامعة الموصل

الاستلام القبول

١٩ / ٠٩ / ٢٠١٠ ٠٧ / ١١ / ٢٠١٠

Abstract

The process of diagnosis and classification of moving targets occupies a major place in applications that use a computer. Facts in our world is a visual moving objects deal with video scenes that contain an unspecified number of moving targets. Here, the idea was to cut and diagnosis and follow-up to these goals through sequential video images.

The idea of this paper is to take a video file (MPEG) and convert it to a set of images (JPG format) for the purpose of setting the goal of any moving object tracking, i.e. a moving target within a video file if it's a person, a car or a ball, etc. For Recognition the facility of subtracting between the frames is used. where Tracking is executed using Rectangle algorithm.

Programs have been written in "Visual C#" as it considered one of the most efficient visual languages.

الملخص

ان عملية التشخيص والتصنيف للاهداف المتحركة تشغل حيزاً كبيراً في التطبيقات التي تستخدم الحاسوب . حيث ان الحقائق في عالمنا البصري هي عبارة عن اشياء متحركة تتعامل مع مناظر فيديوية والتي تحتوي على عدد غير محدد من الاهداف المتحركة . وهنا كانت الفكرة بتقطيع وتشخيص ومتابعة هذه الاهداف من خلال فيديو متعاقب من الصور.

تقوم فكرة البحث على أخذ ملف فيديوي (MPEG) وفتحه وتحويله إلى مجموعة من الصور من نوع (JPG) وذلك لغرض تحديد هدف Object متحرك أي تتبع هدف متحرك داخل ملف فيديوي كأن يكون شخص أو سيارة أو كرة ..الخ. لغرض التمييز تم استخدام خاصية الطرح ما بين Frames. اما التتبع فقد تم باستخدام خوارزمية المستطيل.

كتبت البرامج بلغة "Visual C#" باعتبارها إحدى اللغات المرئية الكفوءة.

1. المقدمة

تنجز عملية تشخيص او تحديد هدف معين في فيديو متعاقب عادة بواسطة اصطفاف ازواج من الصور وعمل مقارنة بين الصور . الهدف المتحرك من الممكن ايجاده خلال عملية المقارنة بين الصور باستخدام خوارزمية مثل Clustring (العنقدة) او (العناقيد). ولسوء الحظ فان اصطفاف الصور عادة قد لا تأتي بالضبط بنفس نقطة الصورة. [1]

قد يؤدي الاختلاف الكبير بين الصور الى تنبيه كاذب عن وجود الهدف . بالإضافة الى ان هناك مشكلة اخرى وهي ان عمليتي اصطفاف الصور والطرح بينها تاخذ وقتا كبيرا عند التنفيذ، كذلك ان اختيار كل نقطة في الصورة تاخذ وقتا خلال عملية المقارنة . اضافة الى ذلك ان تطبيق خوارزمية الـ Clustring تحتاج وقتا كبيرا للتنفيذ. ولهذا اذا حاولنا تطبيق هذا العمل على انظمة تعمل بالوقت الحقيقي Real Time فلابد من وجود اجهزة عتاد Hardware ذات مواصفات خاصة يتطلبها هذا العمل. [2][3]

2. خزن الصور الرقمية

تخزن الصور في الحاسوب على شكل بيانات صورية في ملفات ذات هيئة خاصة تسمى ملفات الرسم (Graphic File Format). هناك عدة طرائق لخزن الصور الرقمية في الحاسوب حسب حاجة وطبيعة التعامل مع الصور، والطريقة المناسبة لخزن ملفات الصور لاجل المعالجات الحاسوبية وامكانية اجرا ئها بوسائل متعددة بعيدة عن تخصص كل وسيلة هي (Joint Picture Expert Group (JPG)) والتي تختار لكونها تعطي وصفا فيزيائيا جيداً عن الصورة، ويمكن ان يكون الكبس فيه كبيراً او صغيراً ولهذا السبب يكون حجم الملف مناسباً نسبة الى باقي انواع ملفات الصور مثل BMP وGIF وغيرها. [4]

يطلق على الصورة التي تخزن في ملف من نوع JPG بالصور النقطية (Raster Image) والتي يمكن تمثيلها بنموذج مصفوفة ثنائية بالشكل $f(r,c)$.

حيث ان:- f: متغير يشير الى اسم المصفوفة

r: عدد اسطر المصفوفة c: عدد اعمدة المصفوفة

هذه المصفوفة تخزن بيانات الصورة وعليه تكون الصورة النقطية هي مجموعة من النقاط حيث ان لكل نقطة لونا وموقعا محددا مسبقا. عندما يتم ادخال الصورة من هذا النوع من قبل أي نظام تتم المعالجة على نقاط هذه الصورة التي يطلق عليها عادة بمعلومات الصورة (Image Information) وليس على الاشياء المكونة للصورة. [5]

3. الملف الفيديوي MPEG

تعتبر طريقة MPEG (Motion Picture Expert Group) من الطرائق المعروفة والمشهورة للتعامل مع الفيديو او الصور المتحركة . حيث لاقت اقبالاً شديداً على استخدامها من مختلف المجالات كمحطات الاذاعة المرئية وسوق الافلام والانترنيت.

تعتمد هذه الطريقة على حذف المعلومات المتكررة وبالأخص ما بين الصور بالاضافة الى تلك المتكررة في الصورة الواحدة . وكذلك تعتمد على حذف المعلومات غير المهمة في الصور . وبهذا الحذف يقل حجم البيانات للتخزين او النقل.

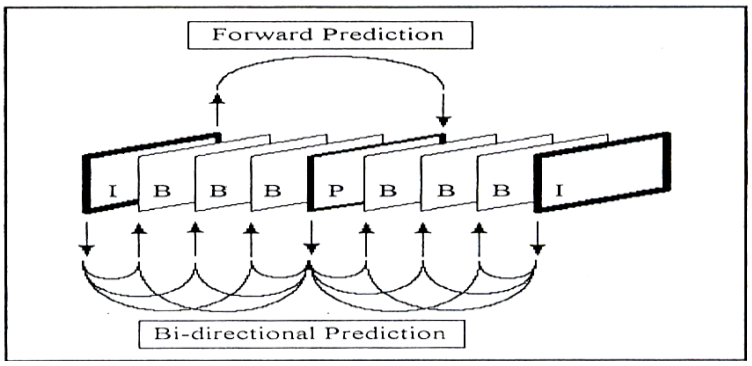
تشترك الصور المتتابعة في الفلم عادةً مع بعضها البعض في كثير من المكونات . فعلى سبيل المثال، صورتين متتابعتين في فلم ما لهما اغلب المعلومات المكونة للصورة الاولى– ان لم تكن جميعها–وهي نفسها في الصورة التي تليها . وفي حال اختلفت الصورتان، فعادةً ما تختلفان في الاجزاء المتحركة القليلة والتي قد انزاحت قليلاً بالفترة الزمنية القصيرة ما بين الصورتين، والتي لا تتعدى عشر الثانية . ففي هذه الحالة، تبحث طريقة MPEG عن الاجزاء المتحركة لتحفظ تلك المعلومات التي تفيد الحركة فقط بدلاً من الاحتفاظ بالصورة كاملة، اذا ما تم الاحتفاظ بالصورة التي تسبقها . ومثال على ذلك : معلومات الحركة في مناطق الفم لمذيع الاخبار . وتتم عملية MPEG بتقطيع الصور الى وحدات كبيرة نسبياً والتي تكون بحجم pixel 16x16. [6]

قبل عملية المقارنة والبحث تلجأ طريقة MPEG الى الاستفادة من ضعف حساسية العين للتغير في الالوان . فعين الانسان تكون اكثر حساسية للتغير في الاضاءة لمساحة صغيرة من الصورة من التغير في اللون في تلك المنطقة من الصورة . لذلك تسعى الطريقة الى التقريب في الالوان وحذف التفاصيل اللونية التي لا تلحظها العين . وبذلك يقل حجم الصورة اذا ما استبدل التكرار في الالوان باللون والعدد بدلاً عنهما . كذلك فان تقليل عدد الالوان يسهل عملية البحث عن الاجزاء المتحركة في الصورة والتي تتميز بها طريقة MPEG.

تتحول الصور المتتابعة في الفيديو باستخدام طريقة MPEG الى سلسلة من الصور المضغوطة (المختزلة) تكون كل صورة احدى الانواع الثلاثة التالية: [6][7]

- صورة – I I - Frame
- صورة – P P - Frame
- صورة – B B - Frame

وترتب تلك الصور في مجموعات تسمى "مجموعة من الصور Group of Pictures" GOP". ويتكرر الترتيب في المجموعة الواحدة تلو الاخرى . يمثل الشكل (1) مثالاً على مجموعة GOP. وتكون اول صورة في اية مجموعة GOP دائماً من النوع I-Frame. وفيما يلي وصف للأنواع أعلاه:



شكل رقم (1): مجموعة من الصور

أ. I-Frame

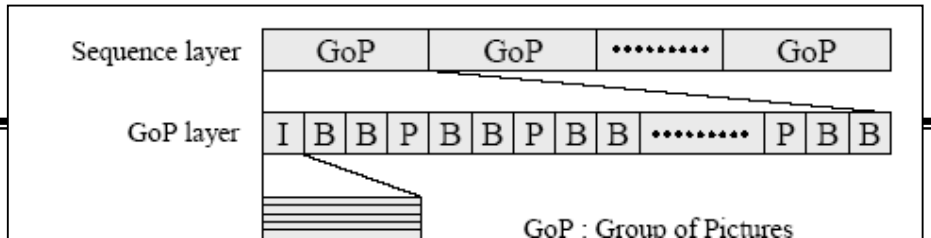
وهو اختصار Intra-frame encoding أي الترميز الذي يعتمد على معلومات الصورة نفسها فقط. حيث يعتمد هذا النوع على اختزال الصورة مستقلة عما يسبقها او يليها من الصور في الفيديو . وتتكّون الصورة مجال مجموعة الالوان YCbCr، حيث ان Y تمثل الاضاءة و CbCr تمثل سويةً التلويزة الحمراء والزرقاء على الترتيب.

ب. P-Frame

وهو اختصار Predictive frame encoding اي الترميز المتنبىء. يعتمد هذا النوع على معلومات الحركة ان وجدت . وتسمى متجه الحركة (MV) Motion Vector. وتكون لكل وحدة كبيرة من وحدات الصور والتي بحجم pixel 16x16 ان وجد لها مثيل في الصورة التي تسبقها. فان وجدت كانت لها معلومات تفيد الحركة MV حتى وان لم تتحرك.

ج. B-Frame

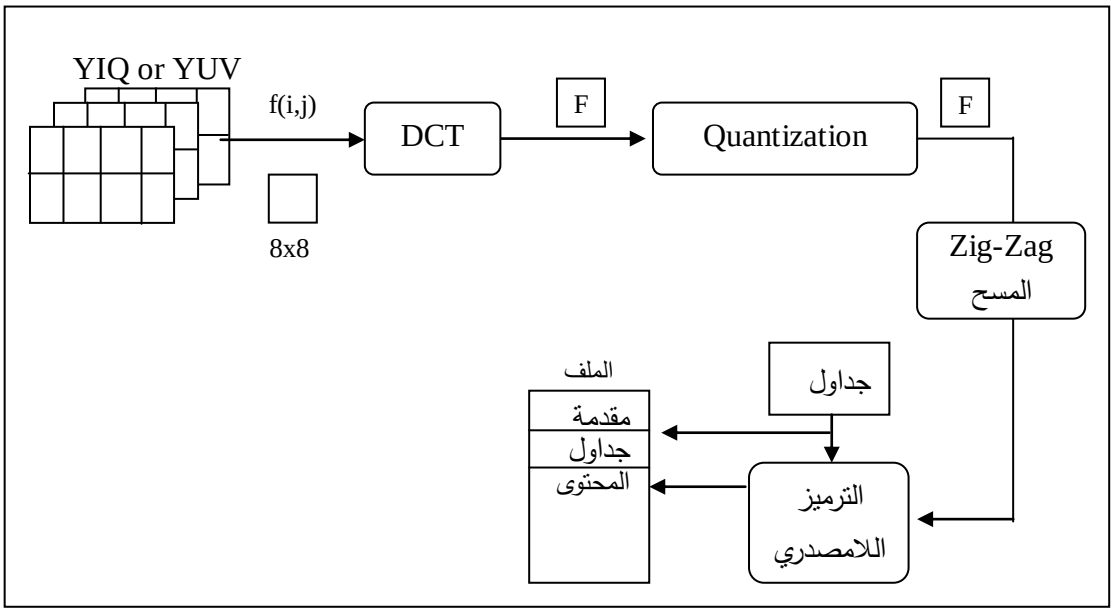
وهو اختصار Bidirectional Predictive frame encoding أي التنبؤ ذو الاتجاهين. يأتي هذا النوع من الصور المختزلة ليحد من زيادة حجم البيانات في حال فشل عملية البحث عن الوحدة الشبيه ة. فيعمل B-Frame ذو الاتجاهين، على البحث عن الوحدة المعنية من الصورة الحالية في كلا الاتجاهين الزمنيين، (الماضي والمستقبل) على ان تكون الصورة المقارن بها من النوع I او P. والشكل (2) يوضح مقطعاً تفصيلياً عن ملف الفيديو من نوع MPEG.



4. بيانات الفيديو (Frame): JPG

تكون الصورة مكونة من 24 bit لكل Pixel، عادة ما تمثل RGB أي 8 bit لكل لون . فتحول الى مجموعة الوان YIQ او YUV. وكأن الصورة الواحدة تتمثل بثلاث طبقات من الصور، واحدة لكل عنصر من العناصر الثلاث YIQ او YUV.

وقد اختيرت مجم وعة الألوان YIQ أو YUV لان عنصر Y يمثل الإضاءة . وذلك للاستفادة من ضعف حساسية العين للتغير في الالوان بعكس الإضاءة. فنستطيع تمثيل الطبقة الثانية والثالثة Q و I او U و V لكل وحدة بحجم 4x4 pixel بدلاً من 8x8 pixel لان كل pixel تغطي مساحة 2x2 pixel في الصورة الأصلية . والخطوات الأساسية لطريقة JPG (Joint Picture Expert Group) سوف تطبق على كل طبقة بالمثل . لاحظ الشكل (3). [6]



شكل رقم (3): المراحل الأساسية لعملية JPG

5. أعمال سابقة ذات علاقة

يمكن تلخيص البحوث التي سبقت هذا العمل والتي تعنى بنفس الموضوع وكما يلي:

قام الباحثان باسل الخياط وعبي ر عبد الخالق في العام 2005 بتقديم خوارزمية مقترحة للتعقب الفيديوي وتطبيقها على كرة السلة، وحصلوا على نتائج جيدة في حينها [1].

وفي العام 2006، استطاع الباحث .Yilmaz, A وآخرون من تقديم مرجع على قدر كبير من الاهمية يتعلق بموضوع التعقب الفيديوي شمل العديد من الطرائق والخوارزميات المعتمدة لذلك [2]. وجاء العام 2008 وعلى ايدي الباحثان ,.Zheng, Y و ,.Meng, Y ليشهد عرض خوارزمية ديناميكية للتعقب تتعلق بذكاء سرب الحشرات عند الحركة [9].

كذلك في العام 2009 تمكن الباحث .Utkarsh, K. S من تطبيق طريقته في التعقب عن طريق الأجزاء [8]. اما الباحث .Cecilia, C.M.B وآخرون فقد تمكنوا من تطوير فكرة التعقب الاعتيادي ليصبح تعقب بطريقة التوازي وذلك في العام 2010 [10].

6. التعقب الفيديوي الحي

كجزء من البحث لابد من تطبيق خصائص التعقب للجهاز والتي تنفذ بالكامل على لوح الجزء المادي (hardware board)، حيث ان تصميم الاشياء يفيد كثيرا وبشكل خاص على قطعة الاجهزة المادية والتي تاخذ جهدا ووقتا ، ولكي نتجنب التحديدات المعقدة للخوارزميات التي تطبق على الواح الاجهزة المادية ولضمان كون جميع الخوارزميات قد صممت بشكل صحيح لذا سيتم كتابة (window application) لتقليد بيئة التقاط frame من كاميرا (web) وكاميرا اليد بالتعقب. [9][8]

تم بناء Class يجهز تطبيقات واجهات برمجية للوصول الى اجهزة تجميع الفيديو ومنها (Web Camera) والتي استخدمت مع هذا التطبيق. لكي نكون قادرين على تتبع Object من خلال فيديو حي سنكون بحاجة لـ (Frames) مستقلة، وللحصول على ذلك تم استخدام Cap) (Set Callback On Frame وكما يلي:

BOOL CapSetCallbackOnFrame (HWND hwnd, FrameCallback fpProc)

HWND hwnd * : Handle لنافذة التقاط الفيديو.

FrameCallback fpProc * : مؤشر لعرض دالة الاستدعاء الارجاعي. وتخصص لها القيمة (NULL) لمنع تمكين تثبيت دالة الاستدعاء الارجاعي.

حال دخول Frame إلى الإجراء (callback) سيتم اقترانه بنافذة الالتقاط لتبدأ عملية التعقب.

ان اختيار فضاء اللون الصحيح للتطبيق المحدد سيبسط وبشكل ملحوظ عملية الحساب ، وفي هذا التطبيق هي خاصية السطوع (Brightness). لكي نكون واثقين من ان فضاء الألوان الذي نتعامل معه يمتلك مكونات السطوع عندئذٍ يكون YUV هو احد فضاء الألوان الذي يملك هذه الخاصية. وبما ان YUV قد لا يكون بالضرورة الصيغة التي توفرها (web camera)، ولذا سنكون بحاجة لعملية التحويل من النموذج (RGB 24) الى صيغة (YUV).

يمكن توضيح العلاقة بين RGB و YUV من خلال المجموعة التالية: [6]

$$\begin{bmatrix} Y \\ U \\ Y \\ Y \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 & 0.063 \\ -0.148 & -0.291 & 0.439 & 0.500 \\ 0.439 & -0.368 & -0.072 & 0.500 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \\ L \end{bmatrix}$$

هذه المصفوفة ناتجة من مفهوم تغيير النواقل في الجبر الخطي، حيث انه في مثل هذه الحالات يُعتمد على تدوير مكعب الألوان وكانه ناقل جديد يمتلك مكونات مع ملكية فردية R=G=B .

الان بإمكاننا الوصول مباشرةً الى سطوع كل pixel، والخوارزمية التي تم استخدامها لتعقب سطوح الكائن (object) هي خوارزمية المستطيل، والتي تحافظ على تعقب أربعة نقاط في كل (frame)، الجزء الأعلى top والأسفل bottom واليمين right واليسار left، ومعظم هذه النقاط تكون قد تجاوزت بسطوعها نقطة threshold المعينة.

بالإمكان بناء المستطيل من هذه النقاط والتي تخبرنا عن سطوع الكائن ، كما يمكن استبدال لون اطار المستطيل باللون الذي نرغب به.

تمتاز خوارزمية المستطيل ببعض النقاط هي: [10]

- اعطاء موقع Object لمقطع كامل على الشاشة.
- عدم الاحتفاظ باي معلومات عن شكل Object.
- عدم القدرة على تحديد وسط Object.
- عدم القدرة على متابعة اكثر من Object في وقت واحد.

7. النتائج والمناقشة

كما لاحظنا مما ذكر سابقا فان العمل يقوم على جزأين رئيسيين هما :

1.7 برنامج Object Tracking

يمكن تلخيص عمل هذا البرنامج بالتسلسل التالي :

أ. سوف نختار ملف فيديو بصيغة (MPEG) قد تم تسجيلة مسبقاً، حيث يتم فتح هذا الملف من داخل البرنامج باستخدام الأمر (Open).

ب. تحويل ملف MPEG إلى مجموعة من (Frames) بهيئة صور (JPG) وكما موضح في الشكل رقم (4).

ج. قراءة هذه الصور من خلال (Header) الخاص بها وهكذا نكون قد حصلنا على صور ملونة من نوع JPG .

د. تحويل الصور الملونة إلى صور ذات تدرجات رمادية (Gray Level).

هـ. طرح جميع الصور (واحدة تلو الأخرى) من الصورة الأولى والتي تعتبر خلفية ثابتة وان ناتج هذه العملية سيكون ظهور (Object).

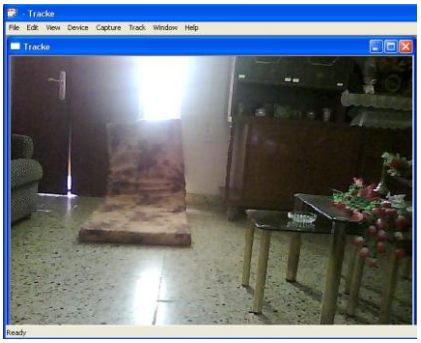
و. وبما أن (Object) الناتج من عملية الطرح يحمل التدرج الرمادي من الأبيض إلى الأسود، اذن سيتم البحث عن قيمة يتم تحديدها داخل البرنامج تقع ضمن التدرج الرمادي (Object) وهذه ستشكل النقطة الدالة على Object.

ز. أخذ إحداثيات تلك النقطة جميع النقاط ذات قيم لونية أعلى من النقطة المحددة.

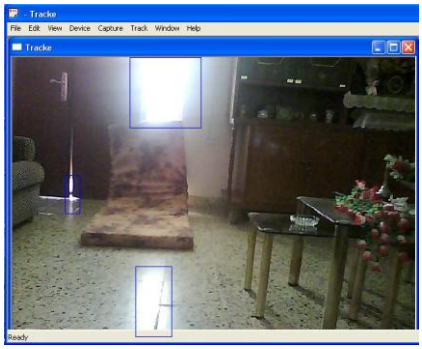
ح. البحث عن:

- اكبر قيمة بالإحداثيات بالنسبة (Y-axis) والتي تشكل ارتفاع Object.
- اقل قيمة بالإحداثيات بالنسبة (Y-axis) وتشكل الحد السفلي Object.
- اكبر قيمة بالإحداثيات بالنسبة (X-axis) وتشكل بداية أبعاد Object.
- اقل قيمة بالإحداثيات بالنسبة (X-axis) وتشكل نهاية أبعاد Object.

رسم مربع أو مستطيل بالاستفادة من الاحداثيات السابقة (حسب أبعاد Object) وإسقاط هذه الإحداثيات على الصورة الأصلية (الصورة الملونة قبل تحويلها). لاحظ الشكل رقم (5) والذي يوضح الصور التنفيذية للبرنامج مع تحديد الهدف.

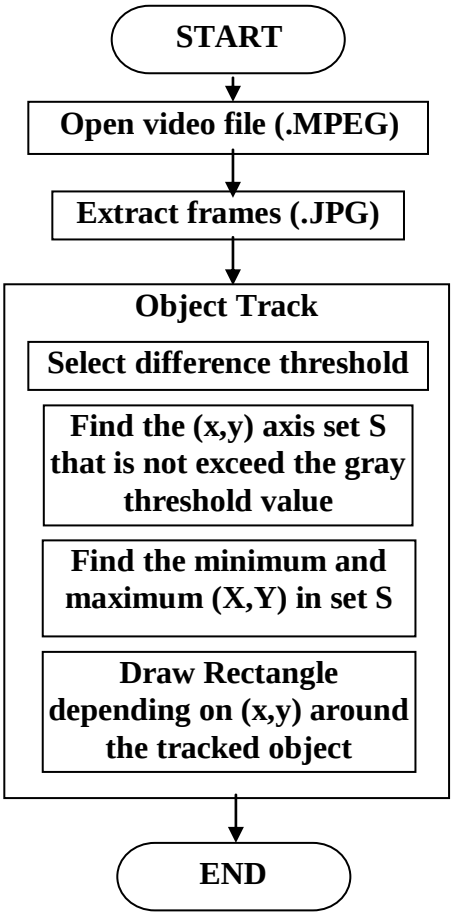


(a)



(b)

شكل رقم (5): (a،b) تحديد هدف ضمن ملف فيديو



شكل رقم (4): مخطط يوضح متابعة هدف من ملف فيديوي نوع (MPEG)

2.7 برنامج التعقب الحي Live

يعمل برنامج التعقب الحي (والموضح بالشكل (6)) على:

أ. متابعة Object متابعة حية حيث يبدأ البرنامج بعد الدخول الى القائمة الموضحة بالشكل رقم (7).

ب. اختيار المتابعة والتعقب (Track).

ج. الدخول الى الاختيار (Brightness) والذي سيطلب تحديد قيمة (Brightness) التي ستعتمد عليها عملية المتابعة.

د. تحديد نوعية احاطة Object عن طريق الاختيار مابين (الأول أو الثاني).

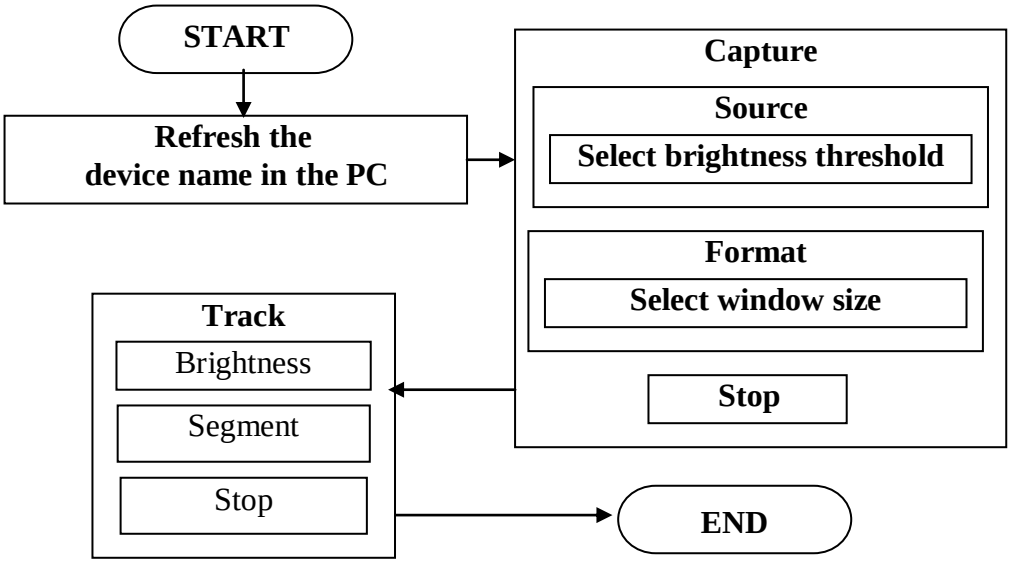
- فللأول سيعمل على احاطة Object بشكل مستطيل أو مربع حسب أبعاد Object وكما موضح في الشكل رقم (8).
- اما الثاني فسيعمل على إحاطة Object بمجموعة نقاط (منتظمة وغير منتظمة) وحسب شكل Object، انظر الشكل رقم (9).

هـ. اختيار (Ok) في نفس القائمة مما يعني بداية عملية المتابعة والتعقب (Track).

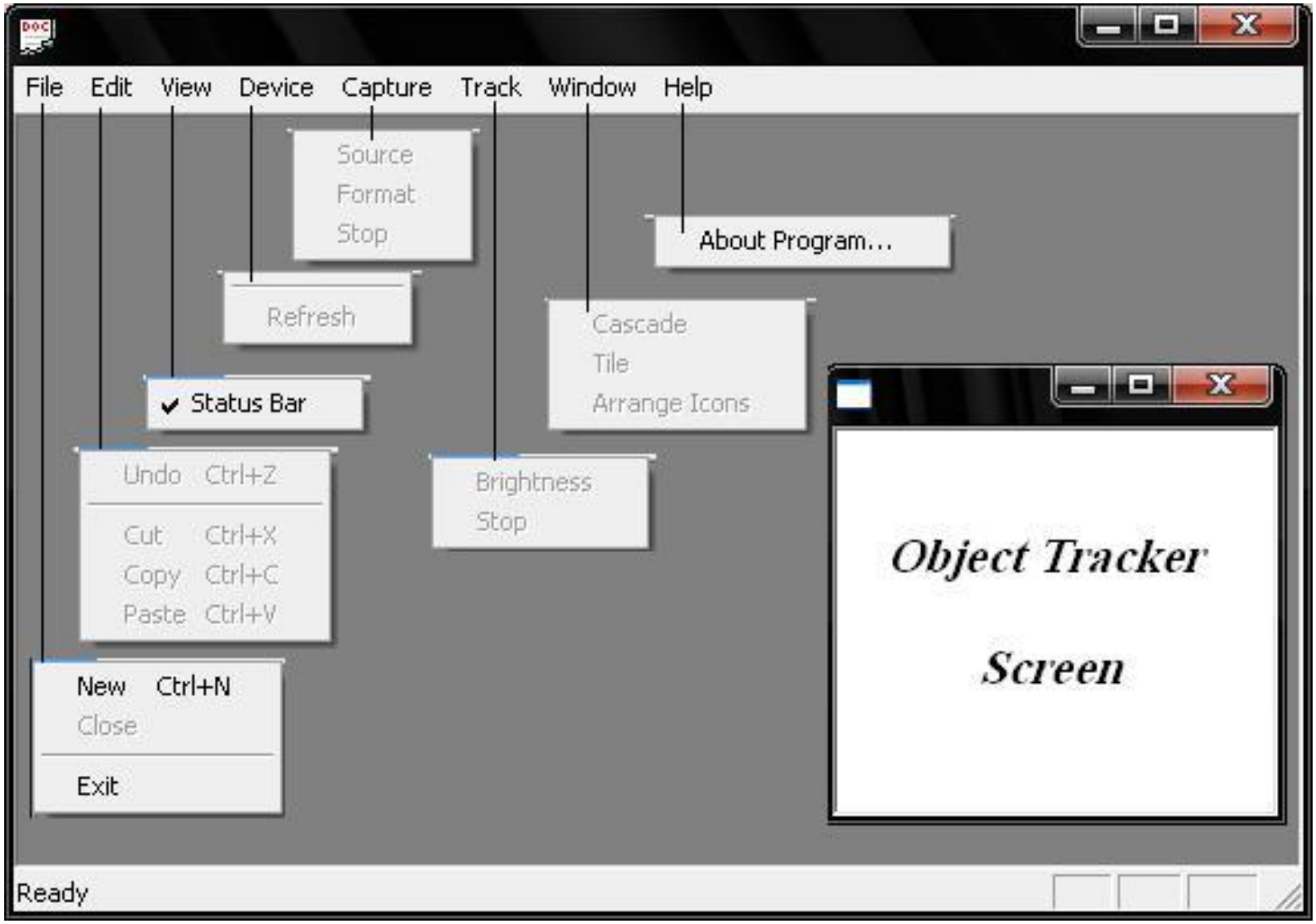
تمتاز خوارزمية التعقب الحي ببضعة محاسن:

- القدرة على تعقب اكثر من Object.
- القدرة على ان تتعقب شكل هذه Objects.

يمكن حساب عدد Pixel التي تكون Object على الشاشة . مع هذ هـ المعلومات والمسافة الصحيحة المحددة لموقع Object في أبعاده الثلاثة سيكون بالإمكان تحديده.

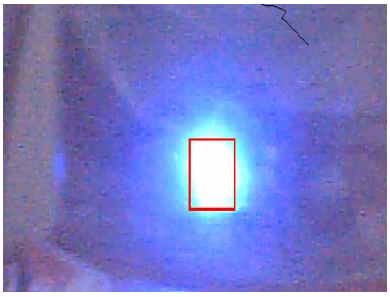


شكل رقم (6): مخطط يوضح متابعة هدف من ملف فيديوي حي

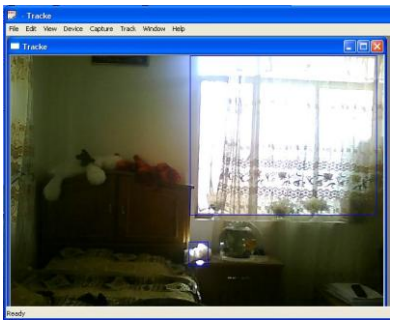


شكل رقم (7): الشاشة الرئيسية والقوائم الفرعية لبرنامج التعقب الحي

(a)



(b)



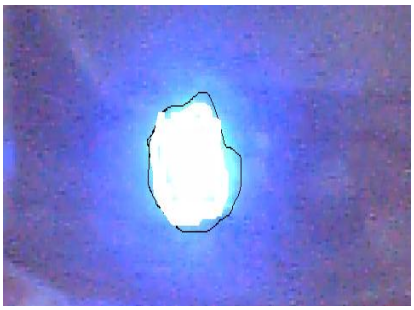
(c)

الشكل رقم (8): (c،b،a)

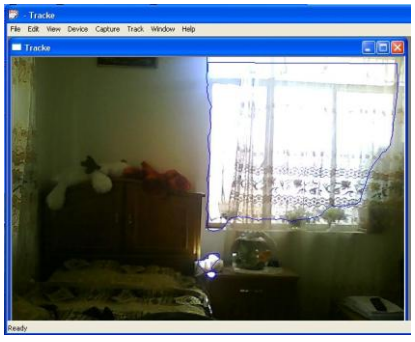
يوضح نموذج object محاط

بمستطيل او مربع ضمن فيديو حي

(a)



(b)



(c)

الشكل رقم (9): (c،b،a)

يوضح نموذج object محاط بنقاط منتظمة

او غير منتظمة ضمن فيديو حي

8. المصادر

١) باسل الخياط و عبير عبد الخالق (2005)، "خوارزمية مقترحة للتعقب الفيديوي وتطبيقها على كرة السلة "، مجلة الرافدين لعلوم الحاسبات والرياضيات المجلد (2) العدد (2).

2) Yilmaz, A., et. al., (2006), "Object Tracking: A Survey", ACM Computing Surveys, Vol. 38, No. 4, Article 13, December.

3) Cecilia, C. M. B., et. al., (2010), "A Parallel Method For Object Tracking", 17th International Conference on Systems, Signals and Image Processing.

4) Gonzalez, R. C. and Woods, R. E., (2002), "Digital Image Processing", Prentice Hall, Inc., 2nd edition.

5) Acharya, T., and Ray, A. K., (2005), "Image Processing Principles and Applications". A John Wiley & Sons, Inc., Publication.

٦) حرز الله، د .نائل، و الضامن، د .ديما، (2008)، "الوسائط المتعددة "، دار وائل للنشر، الأردن.

7) Nori, A. Sami, (2006), "An Investigation for Steganography in Moving Pictures", Ph.D., Thesis submitted to the Computer Science Dept., College of Computer Science & Mathematics Sciences, University of Mosul.

8) Utkarsh, K. S., (2009), "Fragment Based Object Tracking", Dept of Computer Science and Engineering IIT Kanpur –INDIA.

9) Zheng, Y., and Meng, Y., (2008), "Swarm Intelligence Based Dynamic Object Tracking", 978-1-4244-1823-7/08 IEEE.

10) "Object Tracking Toolbox", (2009), The European Commission within the Sixth Framework Program.