

نموذج معتمد على المظهر ومستقل عن الإيماءات لاكتشاف اليد البشرية المجردة وملاحظتها*

م. مدحت الصوص**

د. عمار جوخدار***

الملخص

تتعدد البحوث والدراسات العلمية التي تعمل على تحسين طريقة تفاعل الإنسان مع الآلة وجعلها أقرب إلى الأسلوب البشري في التفاعل سواء من خلال محاولة فهم الكلام أو الإيماءات أو حتى إشارات الدماغ، ولكن اعتماد الإنسان اعتماداً كبيراً على يديه في التفاعل مع بيئته المحيطة أعطى أهمية خاصة للبحوث التي تهتم بتعرف اليد وحركاتها وملاحظتها. يقدم هذا البحث أسلوباً جديداً لتعرف يد بشرية تتحرك بشكل طبيعي أمام كاميرا رقمية عادية وملاحظتها ضمن بيئة غير معروفة مسبقاً، بغض النظر عن وضعية الأصابع أو شكل اليد ودون الحاجة إلى ارتداء قفازات أو معلمات إضافية. تقوم الطريقة المقترحة بالتمييز بين اليد والأجسام المتحركة الأخرى وخاصة الوجه عن طريق تحديد نقاط الفيديو المتحركة من لون الجلد البشري، وتطبيق مجموعة من المعايير المقترحة المعتمدة على معالجة صور الفيديو لتعرف الجسم الممثل لليد. قمنا بإجراء دراسة عملية لتقييم الأداء فضلاً عن تصميم تلفاز افتراضي يجري التفاعل معه باستخدام اليد المجردة كتطبيق مباشر على الطريقة المقترحة لمعرفة مدى فعاليتها وسهولة استعمالها بنظر المستخدمين. أظهرت النتائج إمكانية تعرف اليد بالزمن الحقيقي وبدقة تقارب 99% ضمن ظروف إضاءة الغرفة الطبيعية.

الكلمات المفتاحية: التفاعل بين الإنسان والآلة، تعرف اليد البشرية، ملاحظة اليد البشرية.

* أعد هذا البحث في سياق رسالة الماجستير للمهندس مدحت الصوص في قسم الذكاء الصناعي - كلية الهندسة المعلوماتية - جامعة دمشق بإشراف الدكتور عمار جوخدار

** قسم الذكاء الصناعي - كلية الهندسة المعلوماتية - جامعة دمشق

*** أستاذ مساعد في قسم الذكاء الصناعي - كلية الهندسة المعلوماتية - جامعة دمشق

1- المقدمة:

العالي لهذه المسألة بسبب طبيعة البنية الفيزيولوجية لليد البشرية، حيث تتميز مفاصل اليد البشرية باستقلالية عالية فيما بينها مما يمنح اليد البشرية 27 درجة حرية (21 درجة حرية للمفاصل و6 درجات حرية لموضع ودوران اليد) [2] وهذا ما يفسر العدد الكبير للوضعيات الممكنة لليد. تزداد مسألة استخدام اليد في التفاعل تعقيداً عند عدم وجود معرفة مسبقة بالخلفية التي تتحرك أمامها اليد، أو عند وجود أجسام أخرى متحركة ضمن المشهد إذ يصبح من الضروري تمييز اليد من بين مجموعة من الأجسام المتحركة. يمكن تصنيف الطرائق التي اتبعتها البحوث الخاصة بالتفاعل باستخدام اليد البشرية ضمن ثلاث طرائق (approaches) أساسية هي: الطريقة التقليدية، وطريقة المَعْلَمَات (markers)، وطريقة اليد المجردة وفيما يأتي فكرة موجزة عن كل منها.

1. **الطريقة التقليدية:** تعتمد هذه الطريقة على استخدام عتاد (hardware) معين لتحويل مواقع أصابع اليد وحركاتها إلى معطيات رقمية مثل قفاز (glove) مجهز بحساسات دقيقة تقوم بنقل معلومات رقمية عن مواقع الأصابع باستخدام أسلاك موصولة إلى الحاسب [3] أو جهاز phantom الذي يمكن استخدامه لتحريك نقطة بشكل ثلاثي الأبعاد في الفراغ. تعمل هذه الطريقة بالزمن الحقيقي وتتميز بالدقة في الحصول على موضع اليد، ولكنها تُعدُّ مكلفة نسبياً، كما أنها تحدّ من إمكانية حركة اليد بسبب العتاد الواجب استخدامه.

2. **طريقة المَعْلَمَات:** لا تتطلب هذه الطريقة عتاداً معيناً، ولكنها تتطلب ارتداء مَعْلَمَات مخصصة مثل لفافات ملونة من الأقمشة يتم ارتداؤها في رؤوس الأصابع ليجري تعرّفها عليها بسهولة [4] أو ارتداء قفاز ملون بعدة ألوان يسمح بتعرّف وضعية اليد بالكامل من خلال مقارنة صورة القفاز بمجموعة من الصور المخزنة

لم تشهد طريقة تعاملنا مع الأجهزة الالكترونية والحواسب الشخصية تطورات جوهرية تُذكر منذ بدء ظهورها؛ وذلك مع التطور الهائل التي شهدته هذه الأجهزة وتوسّع مجالات استخدامها في حياتنا اليومية. فمازلنا حتّى اليوم نعتمد على الأزرار أو شاشات اللمس أو أجهزة التحكم عن بعد للتعامل مع غالبية الأجهزة الالكترونية على الرغم من المشكلات العديدة التي ظهرت في أثناء استخدام وسائل الإدخال هذه، فضياع جهاز التحكم الخاص بالآلة أو تعطل أحد أزرارها (نتيجة التفاعل المباشر مع اليد البشرية) كقيل بتحويل هذه الآلة مهما كان ثمنها إلى قطعة من المعدن عديمة النفع حتى يتم إصلاحها من جديد.

لذلك بدأ الباحثون بالتفكير بجيل جديد من الآلات التي يمكن للإنسان أن يتواصل معها بطريقة فطرية ودون تماس مباشر، تماماً كما يتواصل مع بني جنسه من البشر الذين يستطيعون معرفة ما يريد من خلال كلامه أو حركات جسده أو حتّى قسماوات وجهه، فظهرت أنظمة الاستعلام والحجز الهاتفية الآلية، وكذلك الروبوتات (Robots) القادرة على فهم الأوامر الصوتية. كما تمّ التوجّه نحو وسائل جديدة للتفاعل يمكن أن تقيّد ذوي الاحتياجات الخاصة باستخدام حركة العينين لتحريك فأرة الحاسب أو تحريك رُبوّط عبر إشارات الدماغ [1] فضلاً عن العديد من البحوث الأخرى التي سعت إلى ردم الهوة بين أسلوب الإنسان في التفاعل وطريقة الآلة في تلقي الأوامر.

لكنّ طبيعة اليد الفريدة التي جعلتها الوسيلة الأكثر استخداماً من قبل الإنسان للتواصل مع محيطه الخارجي، حثّت العديد من الباحثين على الاستفادة من إمكانات اليد البشرية في التفاعل مع الآلات على الرغم من التعقيد

مفصل من مفاصل اليد فضلاً عن موضع اليد، ويمكن ضمن هذه الطريقة اعتماد أسلوب التحليل بالتركيب (analysis by synthesis)، أو أسلوب النمذجة الحركية (kinematics). يعتمد أسلوب التحليل بالتركيب على نمذجة اليد البشرية على شكل مجموعة من الأسطوانات والكرات المرتبطة معاً بمجموعة من المفاصل، تُعدّل زوايا هذه المفاصل بشكل مستمر حتى الحصول على قيم المعاملات المثلى التي تحقق أكبر نسبة تطابق بين مسقط النموذج الثلاثي الأبعاد على مستوي الكاميرا وبين الصورة الحقيقية لليد [10]. في حين يعتمد أسلوب النمذجة الحركية على نمذجة كل إصبع باستخدام سلسلة من المفاصل واعتبار أن رأس الإصبع يمثل نهاية السلسلة، تُحسب زوايا المفاصل المحددة للوضعية الحالية باستخدام النمذجة الحركية العكسية (inverse kinematics) [11]. تُوضَع قيود على حركة هذه المفاصل بشكل مقابل للقيود الفيزيولوجية على مفاصل اليد الحقيقية؛ وذلك لتخفيض حجم فضاء العينة، ولكن ذلك لا يستطيع ضمان الحصول على حل وحيد لقيم زوايا المفاصل دوماً.

تعدّ الطرائق المعتمدة على النماذج بشكل عام مكلفة جداً من جهة العمليات الحسابية؛ مما يجعلها غير قابلة للاستخدام في تطبيقات الزمن الحقيقي.

(2) **الطرائق المعتمدة على المظهر:** تعتمد هذه الطرائق على استخلاص مجموعة من السمات (features) انطلاقاً من الصورة وتعرّف اليد من خلالها، فمثلاً لتعرّف مجموعة من الحالات الممكنة لليد يتم إيجاد تقابل بين متجه السمات وبين فضاء الحالات الممكنة [9, 12]، بحيث يقابل كل متجه من قيم السمات وضعية من وضعيات اليد المختلفة. تمثّل السمات مجموعة من النقاط أو الخطوط أو أية خصائص مميزة ضمن الصورة. تكمن مشكلة هذه الطرائق بأن التقابل بين متجه قيم

مسبقاً لكل وضعية من الوضعيات المراد تعرّفها [5]. تُلتقط في هذه الطريقة حركات اليد باستخدام كاميرا عادية، من ثم تُطبّق مجموعة من الخوارزميات على الصور الملتقطة لتعرّف مواضع المعلمات أو شكل اليد. تعدّ هذه الطريقة أرخص نسبياً من الطريقة التقليدية بسبب عدم اعتمادها على عتاد خاص، كما أنها أصبحت تعمل بالزمن الحقيقي باستخدام حواسيب اليوم، إلا أن هذه الطريقة مازالت تتطلب وجود تجهيزات خاصة يجب على المستخدم حملها معه وارتداؤها في أثناء عملية التفاعل؛ مما يجعلها غير عملية للاستخدام اليومي.

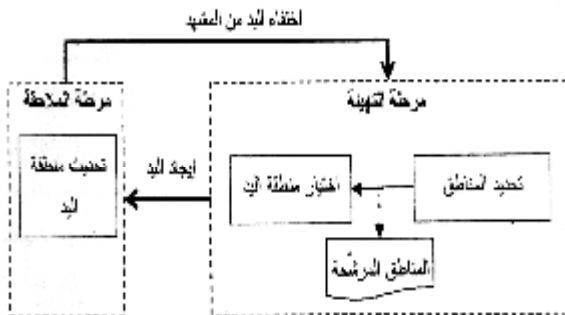
3. طريقة اليد المجردة: تعتمد على السماح للإنسان بالتفاعل عن طريق يده مباشرة ودون وجود أية معلمات أو أسلاك؛ وذلك عن طريق كاميرا عادية تراقب حركة اليد أمامها [6] أو حركة إصبع فوق سطح محدد للكتابة مثلاً [7]، كما يمكن استخدام كاميرا مجسّمة (stereo) للحصول على موضع اليد ضمن فضاء ثلاثي الأبعاد مما يتيح إمكانية التفاعل مع الأغراض في بيئة افتراضية ثلاثية الأبعاد [8]. تعدّ طريقة اليد المجردة الأكثر مرونة بالنسبة إلى المستخدم نظراً إلى قربها من طريقتيه الفطرية في التفاعل مع محيطه الخارجي وعدم حاجتها إلى ارتداء أي تجهيزات، ولكنها تعدّ الأكثر كلفة من جهة العمليات الحسابية بسبب درجة الحرية العالية لليد فضلاً عن المشكلات الناجمة عن الإضاءة وحجب الأصابع لبعضها بعضاً. يمكن تقسيم الطرائق المتبعة ضمن طريقة اليد البشرية المجردة إلى مجموعتين رئيسيتين [9] هما: الطرائق المعتمدة على النماذج (models) والطرائق المعتمدة على المظهر (appearance).

(1) **الطرائق المعتمدة على النماذج:** تعتمد هذه الطرائق على وضع نموذج رياضي لليد البشرية، ومن ثم استنتاج معاملات هذا النموذج التي عادة ما تمثّل زوايا كل

الكف ضمن المشهد وتسمح مرحلة الملاحقة بمتابعتها ضمن فيديو الدخل. تتألف مرحلة التهيئة من خطوتين هما تحديد المناطق المرشحة واختيار المنطقة الممتلئة لليد. يجري تحديد المناطق المرشحة عن طريق عزل عناصر الصورة المتحركة من لون الجلد البشري وتجزئتها بناءً على حركتها إلى مجموعة من المناطق، بينما يجري اختيار المنطقة الممتلئة لليد من خلال تطبيق مجموعة من سمات التشكل (morphological features) المعتمدة على مميزات اليد البشرية. بعد تحديد منطقة اليد تقوم مرحلة الملاحقة بمتابعتها خلال الصور التالية، وعند خروج اليد من المشهد تُنفذ مرحلة التهيئة من جديد لإيجاد منطقة اليد عند عودتها للمشهد. يظهر الشكل (1) مرحلتي النظام وفيما يأتي شرح مفصل لكل منهما.

2-1- مرحلة التهيئة:

تقوم هذه المرحلة بتحديد منطقة اليد البشرية ضمن صورة الدخل، وذلك عن طريق إيجاد النقاط اللونية القريبة من لون الجلد ومقاطعها مع النقاط المتحركة للحصول على نقاط الجلد المتحركة، ومن ثم تجميع هذه النقاط ضمن مناطق مرشحة بناءً على جهة حركتها وسرعتها وقربها من بعضها بعضاً، وأخيراً إيجاد منطقة اليد من ضمن هذه المناطق المرشحة. تتألف هذه المرحلة من خطوتين هما: (أ) تحديد المناطق المرشحة (ب) اختيار منطقة اليد.



الشكل (1) بنية النظام المقترح لتعرف اليد البشرية المجردة وملاحقتها

السمات ووضع اليد عادة ما يكون غير وحيد بسبب اختفاء الأصابع خلف بعضها بعضاً أو دوران اليد، كما أن تعلم التقابل بشكل صحيح يحتاج إلى عدد كبير من أمثلة التدريب. لذلك ونظراً إلى ضخامة فضاء الحالات الممكنة فإن معظم أنظمة التعرف وملاحقة اليد البشرية تكتفي بملاحقة موضع (position) وتوجّه (orientation) اليد فضلاً عن درجتين أو ثلاث درجات حرية إضافية بحسب التطبيق المقترح.

قمنا في هذا البحث بتطوير نموذج معتمد على المظهر يعمل بالزمن الحقيقي (real-time) لتعرف اليد البشرية المجردة وملاحقتها بمختلف وضعياتها. يسمح النظام للمستخدم باستعمال يديه للتفاعل بشكل طبيعي ودون الحاجة لأية أدوات أو معلّات إضافية، وذلك بالاعتماد على كاميرا عادية مثبتة على الجهاز الذي يجري التفاعل معه. لا يُشترط في خلفية المشهد أن تكون بسيطة أو ثابتة، كما لا يُشترط عدم وجود أجسام متحركة أخرى ضمن المشهد. يستطيع النظام المقترح التمييز بين اليد والأجسام المتحركة الأخرى ضمن المشهد وخاصة الوجه بوصفه أكثر الأجسام التي يجري الخلط بينها وبين اليد شيوعاً.

تتألف بقية هذه الورقة من المقاطع الآتية: المقطع 2 يبيّن بنية النظام المقترح لتعرف اليد البشرية وآلية عمله بالتفصيل، ويعرض المقطع 3 نتائج اختبار النظام. ويشرح المقطع 4 التطبيق العملي الذي بُني ويظهر آراء المستخدمين حوله، وأخيراً يلخّص المقطع 5 العمل المنجز وينتظر إلى الآفاق المستقبلية.

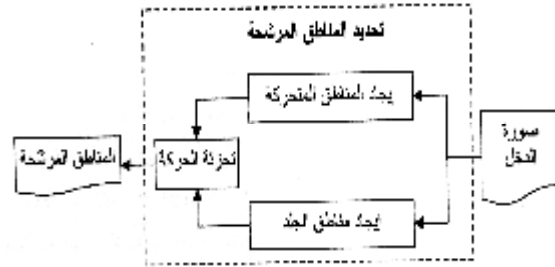
2- النظام المقترح:

يتكون دخل النظام المقترح من سلسلة الصور الملتقطة باستخدام كاميرا رقمية عادية في حين يتألف النظام بحد ذاته من مرحلتين أساسيتين هما: (1) مرحلة التهيئة (2) مرحلة الملاحقة. تقوم مرحلة التهيئة بتحديد منطقة

الصورة الحالية من صورة الخلفية، ومن ثم تعتيب الناتج للحصول على النقاط المتحركة، وهذا ما تعبّر عنه المعادلات الآتية:

$$FD_n(x, y) = |f_n(x, y) - f_{n-1}(x, y)|$$

$$FDM_n(x, y) = \begin{cases} 1, & \text{if } FD_n(x, y) \geq Th \\ 0, & \text{if } FD_n(x, y) < Th \end{cases}$$



الشكل (2) تحديد المناطق المرشحة

إذ $f_n(x,y)$ تمثل قيمة عنصر الصورة الموجود في الموضع (x,y) ضمن الصورة رقم n . إذا كان ناتج الطرح $FD_n(x,y)$ أصغر من عتبة معينة Th فإن عنصر الصورة يُصنّف على أنه عنصر ثابت، وإلا فإنه ينصف على أنه متحرك. ولكن هذا الحل يسبب كثيراً من الضجيج بسبب التغييرات الطفيفة بالصورة الناتجة عن تعديل مستوى الإضاءة بشكل تلقائي من قبل الكاميرات الحديثة قبل النقاط الصورة، وللتخلص من هذه المشكلة اقترحت بعض البحوث حساب قيمة العتبة Th بشكل ديناميكي انطلاقاً من معدل إضاءة الصورة [6]، مما ساعد على تخفيف الضجيج الناتج عن الإضاءة. عملياً لا يمكننا دائماً تمثيل الخلفية باستخدام صورة واحدة من الصورة الملتقطة بسبب إمكانية وجود أجسام متحركة في خلفية المشهد (أوراق شجر، مروحة...)، لذا ولمعالجة حالة الأجسام المتحركة في الخلفية لابد من استخدام نموذج رياضي لتعلم الألوان الممكنة لكل عنصر صورة ضمن الخلفية مثل نموذج MOG (Mixture of Gaussians) [13] الذي يسمح بنمذجة كل عنصر صورة على شكل K توزعاً غوسياً، يُمثّل كل توزع منها بقيمة الوسطى μ والانحراف المعياري σ الخاص به ويُنقل كل توزع بوزن w_k . يجري تعلم هذه الأوزان تباعاً من خلال مجموعة من الصور المتتالية عن طريق المعادلة:

$$w_{k,t} = (1-a)w_{k,t-1} + a(M_{k,t})$$

أ- تحديد المناطق المرشحة:

تقوم هذه الخطوة بإيجاد المناطق المرشحة للاحتواء على اليد البشرية؛ وذلك لأنها تحتوي على مناطق جلد متحركة. يُظهر الشكل (2) الخطوات الفرعية لتحديد المناطق المرشحة، وهي: (1) إيجاد المناطق المتحركة (2) إيجاد مناطق الجلد (3) تجزئة الحركة، وفيما يأتي شرح لكل منها.

(1) **إيجاد المناطق المتحركة:** يقوم الإنسان بشكل فطري باستخدام يديه للإمساك بالأغراض وتحريكها من مكانها، كما يقوم بإيصال أفكار معينة من خلال إشارات يديه، وهذا ما يبقي يديه بحالة حركة مستمرة في أثناء تفاعله مع البيئة المحيطة به، لذلك فإننا نستطيع تعرّف اليد البشرية من خلال تحديد الأجسام المتحركة ضمن المشهد، ومن ثمّ البحث عن اليد بين هذه الأجسام.

يمكن القول: إنّ النقاط المتحركة ضمن المشهد هي النقاط التي تتغير لونها بشكل كافٍ عن لون الخلفية، ولإيجادها لابد من دراسة سلسلة من الصور المتتالية للمشهد. عملياً لا يمكننا دائماً معرفة خلفية المشهد بشكل مسبق بسبب إمكانية تشغيل النظام في أماكن مختلفة، مما يجعل من الضروري إيجاد طريقة آلية لتحديد خلفية المشهد. للقيام بذلك عدت بعض البحوث أن إحدى صور المشهد (كالصورة الأولى مثلاً أو الصورة السابقة للصورة الحالية) يمكن أن تمثل الخلفية، ومن ثمّ يمكننا طرح

كلفة من جهة العمليات الحسابية. يظهر خطأ! لم يتم العثور على مصدر المرجع. (3) مقارنة بين طريقة طرح الصور وتعتيبيها وبين نموذجي MOG وFGD إذ يمكن بوضوح ملاحظة النقاط المتحركة الناتجة عن ضجيج الإضاءة عند استخدام طريقة طرح الصور، بينما نلاحظ زوال هذا الضجيج عند استخدام نموذجي MOG و FGD فضلاً عن الحصول على نتائج أفضل بكثير في حال وجود أجسام متحركة بخلفية المشهد بسبب قدرتهما على بناء نموذج للخلفية وعدم تمثيلها باستخدام صورة واحدة فقط. إن النتائج المتقاربة لنموذجي MOG وFGD وحاجة النظام للعمل بالزمن الحقيقي، جعلنا نعتمد على نموذج MOG عن طريق تمثيل كل عنصر صورة بثلاثة توزيعات غوسية.

(2) إيجاد مناطق الجلد: يعد لون اليد أحد أهم العوامل المميزة لليد المجردة، فإيجاد المناطق اللونية المشابهة للون الجلد ضمن الصورة يسمح لنا بحذف الأجسام جميعها من الألوان الأخرى؛ مما يسرع من عملية تعرف اليد ويزيد من دقتها.

إذ wk,t تعبر عن وزن التوزيع الغوسي رقم k في اللحظة الزمنية t بينما تمثل α معامل التعلم، وتعوّض Mk,t بالقيمة 1 من أجل التوزيع الغوسي المطابق لعنصر الصورة الحالي (وذلك بالاعتماد على μ و σ) في حين تعوّض بالقيمة 0 من أجل بقية التوزيعات. من أجل التوزيعات الغوسية غير المطابقة لعنصر الصورة الحالي تُحدّث قيمتا μ و σ في كل لحظة زمنية على الشكل:

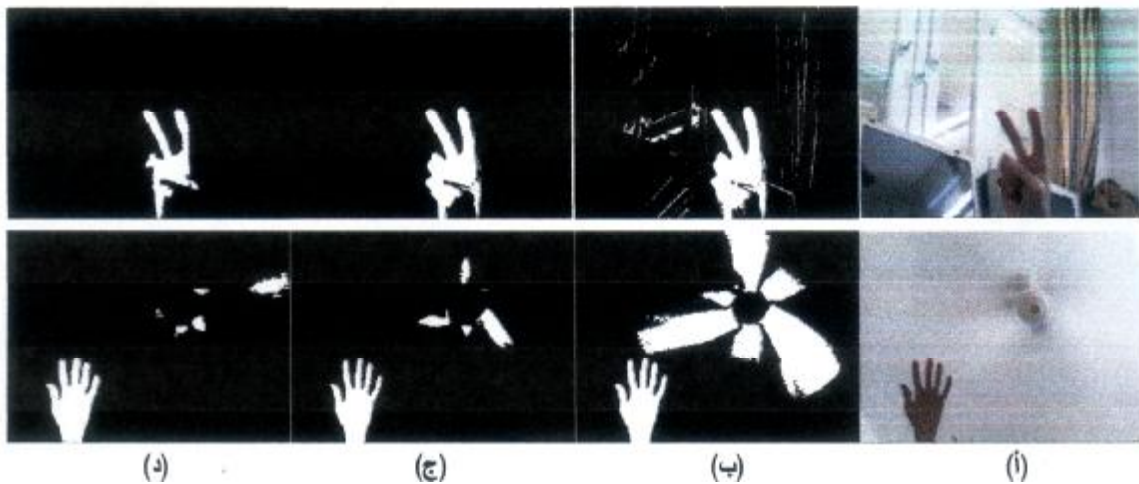
$$m_t = (1-r)m_{t-1} + r f(x, y)$$

$$s_t^2 = (1-r)s_{t-1}^2 + r(f(x, y) - m_t)^2$$

إذ:

$$r = \text{an} \left(\frac{f(x, y)}{m_k, s_k} \right)$$

يمثل ρ معامل تعلم قابل للتعديل بحسب قيمة عنصر الصورة؛ وذلك تبعاً لتابع الكثافة الاحتمالية الغوسي η . يُقارن لون كل عنصر صورة مع التوزيعات الغوسية الخاصة به وعده من الخلفية إذا انتمى لأحد هذه التوزيعات. يمكن استخدام طريقة FGD (Foreground Detection) [14] التي تعتمد على قواعد Bayes للتصنيف بناءً على متجه من السمات ولكنها تُعد أكثر



الشكل (3) مقارنة بين طرائق إيجاد المناطق المتحركة المختلفة
 يحتوي السطر الأول على خلفية ثابتة في حين يحتوي السطر الثاني على خلفية متحركة
 (أ) الصورة رقم 19 من سلسلة صور الدخل والجاري تحديد المناطق المتحركة فيها بناءً على الصور السابقة لها
 (ب) نتيجة تطبيق طريقة طرح الصور والتعتيب (ج) نتيجة تطبيق MOG (د) نتيجة تطبيق FGD

$$Cb = \frac{U}{2} + 0.5$$

$$Cr = \frac{V}{1.6} + 0.5$$

إن الاعتماد على النظام اللوني YCbCr يحد من أثر الإضاءة، ولكنه لا يساعد كثيراً على حل مشكلة اختلاف لون الجلد بين الأجناس البشرية؛ لذلك لجأ بعض الباحثين إلى إضافة مرحلة تهيئة للنظام يقوم المستخدم خلال هذه المرحلة بتحديد منطقة من الصورة تحتوي على جزء من اليد ليتم تعلم لون الجلد مباشرة من هذا الجزء [17]، إن إضافة مرحلة التهيئة هذه تزيد من دقة النظام ولكنها تتطلب من المستخدم إعادة تهيئة النظام كلما تغيرت ظروف الإضاءة المحيطة أو كلما شغل النظام ضمن بيئة عمل جديدة. للتخلص من هذه المشكلة يمكن القيام بعملية تعلم لون الجلد مسبقاً من خلال مجموعة كبيرة من الأمثلة الموجبة والسالبة (مناطق من الصورة تحتوي على يد بشرية ومناطق لا تحتوي عليها)، وهذا ما قامت به بعض الدراسات مثل الدراسة التي أجريت في [18] إذ قامت هذه الدراسة بإيجاد التوزيع اللوني لأكثر

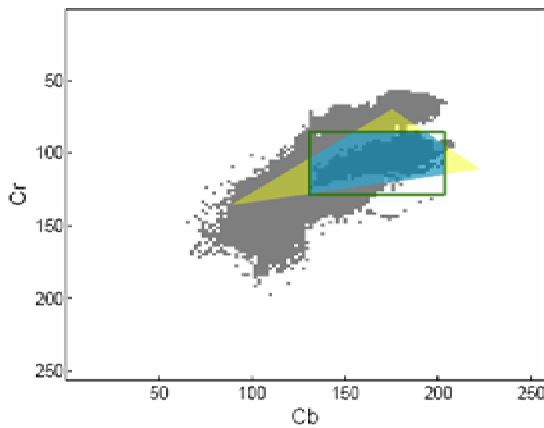
ولكن يجب الأخذ بالحسبان عدم ثبات لون الجلد بين الأجناس البشرية المختلفة، فضلاً عن تغيره من شخص إلى آخر ومن مكان إلى آخر تبعاً لظروف الإضاءة. معظم البحوث تخلت عن استخدام النظام اللوني RGB لتعرف مناطق الجلد بسبب تأثره بشدة بتغيرات الإضاءة ولجأت إلى استخدام أنظمة لونية أخرى تمكننا من عزل مركبة الإضاءة عن مركبات صبغة اللون مثل نظام [15] L^*a^*b أو نظام YUV [16] الذي تدعمه بعض الكاميرات مباشرة، ويحسب انطلاقاً من النظام اللوني RGB عبر العلاقات:

$$Y = 0.3R + 0.6G + 0.1B$$

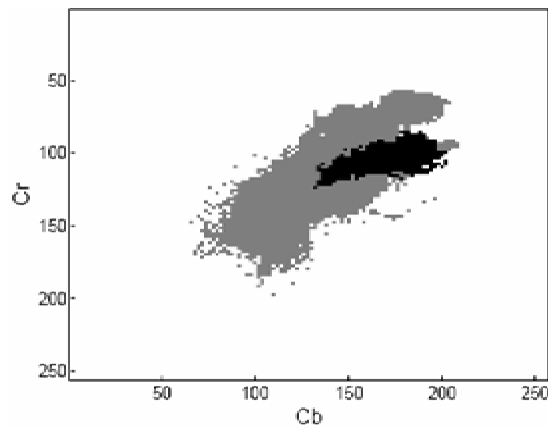
$$U = B - Y$$

$$V = R - Y$$

اعتمدنا في عملنا على نظام YCbCr الذي يُعد نسخة مزاحة من نظام YUV إذ يُحسب Cb و Cr انطلاقاً من U و V:



(ب)



(أ)

الشكل (4) التوزيع الإحصائي لعينات الجلد والخلفية

- (أ) التوزيع الإحصائي في الفضاء CbCr إذ يمثل اللون الأسود عينات الجلد واللون الرمادي عينات الخلفية
(ب) نمذجة عينات الجلد بالمساحة المشتركة بين أصغر مستطيل يحدد هذه العينات وبين المثلث الحاوي عليها

مناطق الجلد على صور دخل ملتقطة في ظروف إضاءة مختلفة.

(3) تجزئة الحركة: لتحديد منطقة اليد يمكننا الاكتفاء بدراسة النقاط المشتركة بين المناطق المتحركة ومناطق الجلد اللونية، ولندعو هذه النقاط بمناطق الجلد المتحركة. إن دراسة مناطق الجلد المتحركة فقط يسمح لنا بحذف المناطق الجلدية غير المتحركة (كالأجسام المشابهة للون الجلد في خلفية المشهد) وكذلك حذف الأجسام المتحركة من الألوان الأخرى ضمن المشهد. بسبب استقلالية مرحلتي إيجاد المناطق المتحركة وإيجاد مناطق الجلد عن بعضهما بعضاً وبغرض التخفيض من زمن التنفيذ الكلي، قمنا بتنفيذ هاتين المرحلتين معاً على التوازي، ومن ثم حساب المساحات المشتركة فيما بينهما التي تمثل مناطق الجلد المتحركة. بعد حساب مناطق الجلد المتحركة نقوم بحذف المناطق ذات المساحة الصغيرة الناتجة عن ضجيج الإضاءة أو عن التغييرات اللونية الطفيفة ومن ثم نقوم بدمج بقية المناطق ضمن مجموعة من الأجسام لنتمكن من دراسة كل جسم على حدة وتحديد الجسم الممثل لليد وملاحقتها.

من 0.5 مليون عينة موجبة (مناطق جلدية من اليدين والوجه لمجموعة من الأشخاص من أعراق مختلفة) ونحو 1.2 مليون عينة سالبة (عينات من خلفيات مختلفة)، وخُصت هذه الدراسة إلى ضرورة استخدام نظام لوني يمكن من عزل مركبة الإضاءة (مثل نظام YCbCr) لزيادة قابلية فصل عينات الجلد عن عينات الخلفية. كما أظهرت نتائج الدراسة توزع عينات الجلد في الفضاء اللوني CbCr بشكل إهليلجي، كما في الشكل (4) (أ).

قمنا بالاستفادة من نتائج هذه الدراسة لبناء نموذج إحصائي لتوزيع لون الجلد ضمن الفضاء CbCr، ولكن بدلاً من نمذجة هذا التوزيع على شكل قطع ناقص ولتخفيض الكلفة الحسابية للنموذج، قمنا بنمذجته على شكل تقاطع بين أصغر مجال حاوٍ عليه (المستطيل الموازي لجملة الإحداثيات) وبين المثلث الحاوي على عينات الجلد جميعها كما في الشكل (4) (ب). وبذلك أصبح بإمكاننا الحصول على مناطق الجلد ضمن الصورة عن طريق أخذ عناصر الصورة (pixels) التي تنتمي ألوانها إلى منطقة التقاطع وإهمال بقية عناصر الصورة، يُظهر الشكل (5) نتيجة تطبيق مرحلة إيجاد



(أ)



(ب)

الشكل (5) أمثلة عن نتيجة تطبيق مرحلة إيجاد مناطق الجلد

يمثل السطر (أ) مجموعة من صور الدخل الملتقطة في ظروف إضاءة مختلفة في حين يمثل السطر (ب) نتيجة تطبيق مرحلة إيجاد مناطق الجلد على هذه الصور، إذ يعبر اللون الأبيض عن مناطق الجلد في حين يعبر اللون الأسود عن المناطق الأخرى

تُحسبُ جهة الحركة ϕ من خلال صورة الحركة الزمنية عن طريق حساب المشتق الخاص بكل عنصر صورة من أجل الاتجاهين X و Y : $F_x(x,y)$ و $F_y(x,y)$ ومن ثم تُطبَّقُ العلاقة:

$$f(x, y) = \arctan \frac{F_y(x, y)}{F_x(x, y)}$$

بينما تُحسبُ جهة الحركة الكلية للصورة $\bar{\phi}$ بالاعتماد على جهات الحركة الموجودة ضمنها:

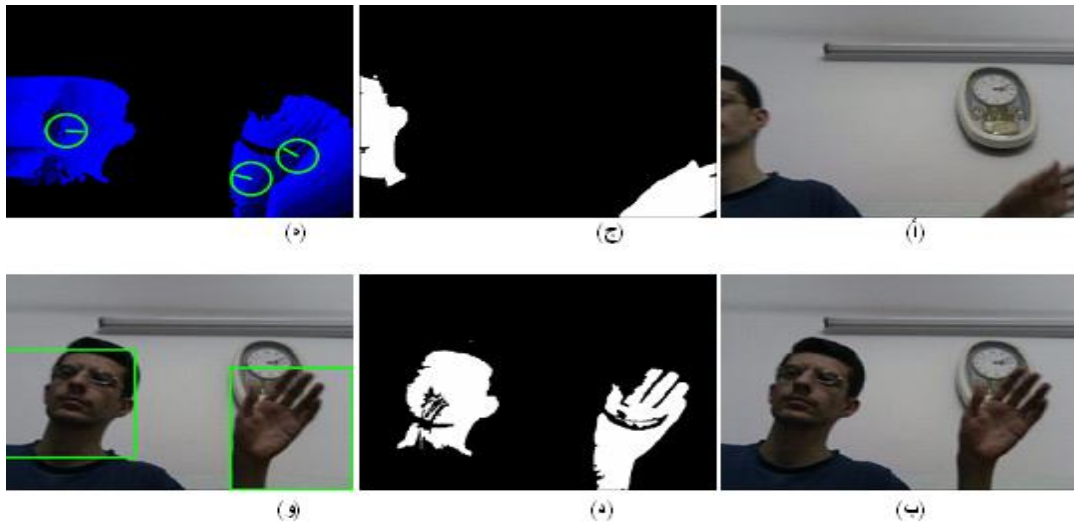
$$\bar{\phi} = \phi_{ref} + \frac{\sum_{x,y} Diff(\phi(x,y), \phi_{ref}) \times norm(\tau, \delta, tMHI_{\delta}(x,y))}{\sum_{x,y} norm(\tau, \delta, tMHI_{\delta}(x,y))}$$

تمثل ϕ_{ref} الزاوية المرجعية للحركة وتحتوي $norm(\tau, \delta, tMHI_{\delta}(x,y))$ على قيمة $tMHI(x,y)$ بعد تقييسها على المجال $[0,1]$ باستخدام اللحظة الحالية τ والمدة الزمنية القصوى δ وتعبّر $Diff(\phi(x,y), \phi_{ref})$ عن الفرق بين جهة الحركة $\phi(x,y)$ والزاوية المرجعية للحركة ϕ_{ref} .

للقيام بذلك قمنا بتطبيق طريقة [19] Motion Templates على مناطق الجلد المتحركة، تقوم هذه الطريقة بدمج المناطق المتحركة من مجموعة من الصور المتتالية ضمن صورة واحدة مع تقويل كل منها بحسب اللحظة الزمنية الخاصة بها، تدعى الصورة الناتجة بصورة الحركة الزمنية $tMHI$ (timed Motion History Image) وتحسب عن طريق العلاقة:

$$tMHI(x, y) = \begin{cases} t & \text{if there is movement at } (x, y) \\ 0 & \text{if } tMHI(x, y) < (t-d) \end{cases}$$

تمثل τ اللحظة الزمنية الحالية في حين تمثل δ المدة الزمنية القصوى التي تُحسبُ صورة الحركة الزمنية ضمنها (عادة ما تكون عدة ثوانٍ فقط). نلاحظ أن لون عنصر الصورة ضمن صورة الحركة الزمنية يعكس زمن حركته، فكلما كانت قيمة عنصر الصورة أعلى كانت حركته أقرب زمنياً، كما في الشكل (6) (هـ).



الشكل (6) تجزئة حركة مناطق الجلد المتحركة

(أ) و(ب) هما صورتا الدخل رقم 89 و 97 على التوالي (ج) و(د) يحتويان على مناطق الجلد المتحركة في صورتَي الدخل رقم 89 و 97 على التوالي (هـ) تمثل صورة الحركة الزمنية $tMHI$ لمناطق الجلد المتحركة من أجل كافة صور الدخل المحصورة بين الرقمين 89 و 97، تعبّر الدوائر الخضراء عن جهة الحركة حيث يمثل مركز الدائرة مركز الحركة والخط المرسوم بداخلها اتجاه الحركة (و) تمثل نتيجة تجزئة الحركة بناءً على جهة الحركة ومساحة المنطقة المتحركة

[17]؛ ممّا يسمح بملاحقة اليد بشكل أفضل، ولكنّ هذا الشرط يقيد حرية المستخدم في أثناء تفاعله مع البيئة المحيطة، كما لجأت العديد من البحوث الأخرى إلى فرض قيود على شكل اليد بدلاً من سرعتها؛ وذلك عن طريق التعرف فقط إلى وضعيات محددة سلفاً لليد من خلال مقارنة صورة المنطقة التي تحتوي على اليد بمجموعة من الصور المخزنة سلفاً لكل وضعية [9, 20] على اعتبار أن وضعيات اليد الأخرى جميعها غير مهمة، مما قيّد أيضاً من حرية المستخدم في أثناء التفاعل. نلاحظ أن الطرائق السابقة جميعها تعتمد في التعرف إلى وضع قيود على اليد البشرية سواء من جهة حركتها باعتبار أنها الجسم الأكبر حجماً أو الأسرع حركة، أو من جهة شكلها من خلال تعرّف اليد في وضعيات محددة سلفاً فقط.

قمنا في هذا البحث بتطوير طريقة جديدة لتعرّف اليد البشرية تسمح بتعرّف اليد دون قيود على سرعة حركتها أو شكلها. تعتمد الطريقة المقّدمة على تطبيق مجموعة من المعايير المستوحاة من شكل اليد البشرية على مناطق الجلد المتحركة ضمن المنطقة المرشّحة بعد القيام بعملية تقييس (normalization) عليها لتصبح بالأبعاد 256x256 بهدف التخلص من مشكلة اختلاف حجم اليد بحسب بعدها عن الكاميرا، ومن ثمّ تطبيق خوارزمية تصنيف (classification) لمعرفة هل توجد يد ضمن المنطقة المرشّحة أم لا انطلاقاً من قيم هذه المعايير؟

قمنا في أثناء البحث بتجربة مجموعة من المعايير المقترحة، وتوصلنا إلى 3 معايير كافية لتمييز اليد البشرية عن الأجسام المتحركة الأخرى وخاصة الوجه لأنه أكثر الأجسام المشابهة لليد وجوداً ضمن المشهد، وهذه المعايير هي:

ممّا سبق نجد أن صورة الحركة الزمنية تسمح بتحديد جهة الحركة من خلال صورة واحدة بدلاً من مجموعة صور متتالية، ممّا يسهل عملية تجزئة الصورة إلى مجموعة من مناطق الحركة المنفصلة بناءً على جهة الحركة بحيث تضم كل منطقة جسماً متحركاً واحداً أو مجموعة من الأجسام القريبة التي تتحرك معاً بالجهة نفسها. نسمي مناطق الحركة المنفصلة هذه بالمناطق المرشّحة، لأنّ كلاً منها مرشّح لأن يحتوي على اليد ولا بدّ من تطبيق بعض المعايير الإضافية لتحديد المنطقة الحقيقية لليد من بين هذه المناطق. يظهر الشكل (6) مثلاً على نتيجة تجزئة الحركة حيث فصلت حركة اليد عن حركة الرأس واستبعدت حركة الجسم؛ وذلك بسبب دراسة مناطق الجلد المتحركة فقط.

ب- اختيار منطقة اليد:

بعد إيجاد المناطق المرشّحة لا بدّ من تحديد المنطقة أو المناطق التي تحتوي فعلاً على اليد بناءً على الخصائص المميزة لليد البشرية والقيام بملاحقتها اعتماداً على موضعها في اللحظات السابقة.

تفترض معظم البحوث أن اليد هي الجسم المتحرك الأكبر حجماً في المشهد [15] أو صاحب أكبر محيط [7]، ولكنّ هذه الفرضيات غير صحيحة دوماً خاصة عندما يظهر الرأس ضمن المشهد الذي يكون عادة مقارباً لحجم اليد أو أكبر منها، أو عندما تكون اليد مغلقة؛ ممّا يجعل محيطها أصغر من محيط الرأس. للتخفيف من هذه المشكلة لجأت بعض البحوث للاعتماد على مكان اليد في اللحظة السابقة وعدّ أقرب منطقة له هي منطقة اليد، ولكنّ هذه الطريقة وحدها قد تؤدي إلى ملاحقة الأجسام المتحركة القريبة من اليد عند خروج اليد من المشهد. قام بعض الباحثين بإضافة قيود إلى سرعة حركة اليد بحيث تتحرك اليد بسرعة شبه ثابتة دوماً [6]

به، هذا ويمكن حساب قيمة الاختلاف المركزي للقطع الناقص الذي معادلته:

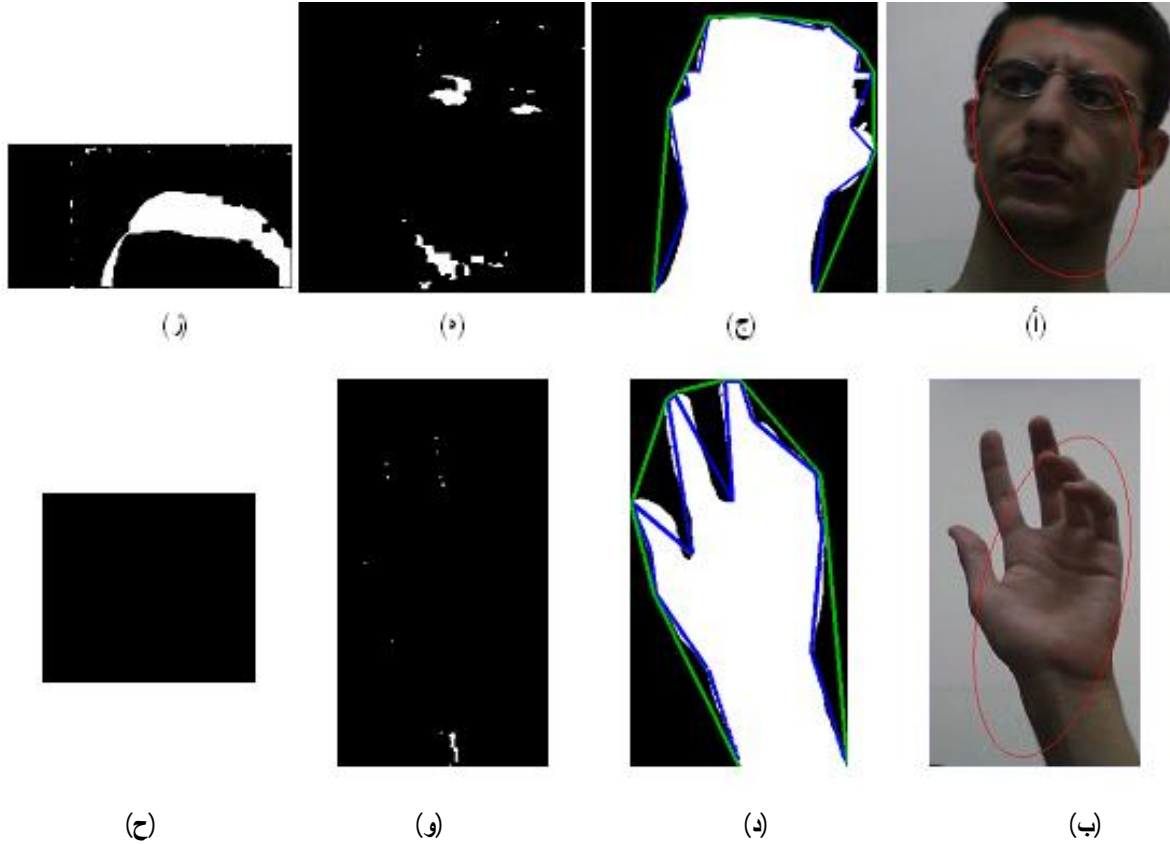
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

من خلال العلاقة:

$$Eccentricity = \frac{\sqrt{|a^2 - b^2|}}{a}$$

تختلف القيمة Eccentricity من جسم إلى آخر، كما أنها تختلف بحسب توضع أصابع اليد.

(1) الشكل الإهليلجي: تتميز اليد البشرية عموماً بشكلها المفلطح الذي تحافظ عليه بغض النظر عن وضعية الأصابع، ففي حالة اليد المغلقة مثلاً تحافظ اليد على تفلطحها بسبب ارتباطها بالساعد. ولقياس مدى التفلطح قمنا بإيجاد القطع الناقص الأمثل الممثل لمجموعة نقاط الجلد المتحركة الشكل (7) (ب) ومن ثم قياس الاختلاف المركزي (Eccentricity) له الذي يساوي النسبة بين المسافة بين محرتي القطع وطول القطر الكبير الخاص



الشكل (7) نتيجة تطبيق معايير تعرف اليد على المنطقة المرشحة

يظهر السطر الأول نتيجة تطبيق المعايير على الوجه في حين يظهر السطر الثاني نتيجة تطبيقها على اليد

(أ) و(ب) تظهران صورة الدخل مرسوماً عليها القطع الناقص الناتج عن معيار الشكل الإهليلجي

(ج) و(د) تظهران مناطق خلل التحبب لمنطقة الجلد المتحرك بعد تعبئة الثغرات وهي المثلثات المحصورة بين الخطين الأزرق والأخضر

(هـ) و(و) تظهران المساحات الداخلية المغايرة للون اليد

(ز) و(ح) تظهران المساحات المتحركة المغايرة للون الجلد والملاصقة للجزء العلوي من منطقة الجلد المتحركة

جسماً متحركاً واحداً مع الوجه). لذا فإن الجسم المتحرك الممثل لليد يجب أن يتألف فقط من مساحات لونية مطابقة للون الجلد، ويجب ألا تكون هناك مساحات من لون مختلف عن الجلد موجودة داخل حدود منطقة الجلد المتحركة أو تتحرك معها.

لإيجاد المساحات اللونية المختلفة عن الجلد الموجودة داخل المنطقة المرشحة المراد التأكد من احتوائها على اليد (وإنطلق على هذه القيمة اسم CbCrCount)، نقوم بتعبئة الثغرات الموجودة ضمن منطقة الجلد المتحركة، ومن ثم نحذف مناطق الجلد اللونية فتبقى المساحات المختلفة عن لون الجلد كما في خطأ! لم يتم العثور على مصدر المرجع. (ه)، نقوم بقياس توزع الألوان ضمن المناطق الناتجة عن طريق تحويلها إلى الفضاء YCbCr وحساب المساحة التي تشغلها ضمن المستوي CbCr حيث تزداد هذه المساحة مع زيادة عدد الألوان.

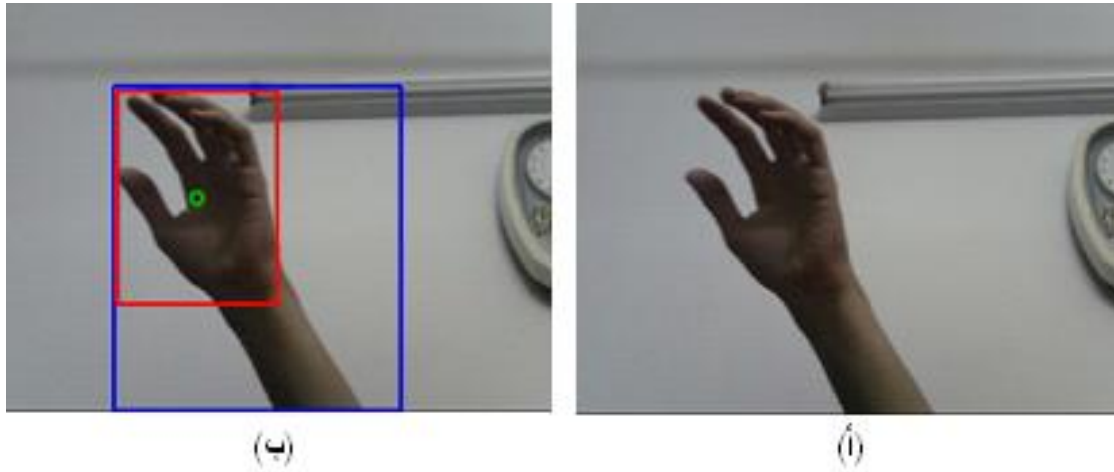
بينما لإيجاد المساحات المتحركة المغايرة للون الجلد والملاصقة لمنطقة الجلد المتحركة (وإنطلق عليها اسم FGNearCount) نقوم بطرح منطقة الجلد المتحركة من المناطق المتحركة الموجودة ضمن نافذة منطقة الجلد المتحركة كما يظهر الشكل (7) (ز) ومن ثم نقوم بحساب نسبة مساحة المنطقة الناتجة مقارنة بأبعاد النافذة. ولتجنب احتساب المساحات اللونية المتحركة الناتجة عن الملابس المغطية للساعد في حال وجود جزء منه ضمن المنطقة المرشحة، نقوم بحساب المساحات المتحركة المغايرة للون الجلد القريبة من الجزء العلوي من المنطقة المرشحة فقط؛ وذلك على افتراض أن جهة اليد لن تكون باتجاه الأسفل. إذاً يمكننا التعبير عن معيار اللون الموحد من خلال القيمتين CbCrCount و FGNearCount.

(2) مناطق خلل التحذب: غالباً ما يكون شكل اليد البشرية غير محدب بسبب تباعد الأصابع عن بعضها بعضاً؛ وهذا ما يترك مساحات فارغة بين جسم اليد وبين المضلع المحدب المحيط بها. لإيجاد هذه المساحات قمنا أولاً بتعبئة الثغرات (holes) ضمن منطقة الجلد المتحركة، ومن ثم استخدمنا خوارزمية convexity defects [21] التي تعمل على تقريب هذه المساحات الفارغة إلى مجموعة من المثلثات، للقيام بذلك تقوم الخوارزمية من أجل كل رأسين متتاليين $P1(x1,y1)$, $P2(x2,y2)$ من رؤوس المضلع المحدب بالبحث عن النقطة $Pm(xm,ym)$ ضمن نقاط المضلع المحدب بحيث يكون لها أبعد مسافة D عن القطعة المستقيمة $[P1,P2]$ وتحسب المسافة D من خلال العلاقة:

$$D = \frac{\frac{(y_2 - y_1)}{(x_2 - x_1)} \cdot (x_1 - x_m) + y_m - y_1}{\frac{(y_2 - y_1)^2}{(x_2 - x_1)^2} + 1}$$

للتعبير عن معيار مناطق خلل التحذب قمنا بحساب القيمة DefectsCount، وهي عبارة عن عدد المثلثات التي تكون قيمة D الخاصة بكل منها كبيرة نسبياً. يوضّح الشكل (7) (د) مساحات خلل التحذب بعد تقريبها إلى مثلثات في حالة صورة اليد، إذ يبدو بوضوح أن المثلثين الواقعيين بين الأصابع لهما ارتفاع أكبر من بقية المثلثات؛ مما يجعل DefectsCount=2 في حالة الصورة (د) في حين تكون DefectsCount=0 في حالة الصورة (ج) بسبب صغر ارتفاعات المثلثات.

(3) اللون الموحد: على الرغم من إمكانية اختلاف الدرجة اللونية للجلد من منطقة إلى أخرى ضمن اليد بسبب ظروف الإضاءة، إلا أننا لا يمكن أن نجد بقاءً لونية مغايرة للون الجلد داخل منطقة اليد أو تشكل معها جسماً متحركاً واحداً (كما في حالة شعر الرأس الذي يشكل



الشكل (8) نتيجة تطبيق مرحلة التعرف وملاحقة اليد

(أ) صورة الدخل, (ب) نتيجة مرحلة التعرف وملاحقة اليد إذ يعبر المستطيل الأزرق الكبير عن منطقة اليد, في حين يعبر المستطيل الأحمر الصغير عن منطقة الكف وتمثل الدائرة الخضراء مركز نقاط الحافات.

الطرائق تخفق عندما يتغير شكل اليد نتيجة تغير وضعية أصابع اليد.

في هذا البحث وبعد إيجاد منطقة اليد بالكامل قمنا بخطوة إضافية لتحديد المنطقة الحاوية على الكف ضمن منطقة اليد التي تم إيجادها؛ وذلك من خلال إيجاد حافات (edges) منطقة الجلد المتحركة ضمن منطقة اليد وحساب مركز النقاط التي تنتمي إلى هذه الحافات P:

$$p = \frac{1}{G} \sum_{(x,y) \in edges} (x, y)$$

إذ C هو عدد النقاط التي تنتمي إلى الحافات. إن النقطة P تكون أقرب إلى أصابع اليد منها إلى الساعد بسبب احتواء منطقة أصابع اليد على حافات أكثر من منطقة الساعد. لتحديد منطقة الكف يتم إيجاد المستطيل الأكبر المتمركز في مركز نقاط الحافات P الواقع ضمن منطقة اليد كما في الشكل (8) (ب). تتجلى أهمية هذه الخطوة عند ظهور جزء كبير وعارٍ من

إن أي معامل من المعاملات الأربعة السابقة لا يكفي وحده لتعرف اليد البشرية ضمن المشهد لأن أيًا منها لا يملك مجالاً محدداً من القيم التي تخص اليد البشرية دون سواها من الأجسام الأخرى، لذلك قمنا بتجربة مجموعة من خوارزميات التعلم التلقائي المخصصة للتصنيف بهدف تعلم العلاقة بين قيم هذه المعاملات وبين وجود اليد البشرية ضمن المشهد، وستشرح ضمن المقطع 3 المتعلق بالنتائج.

يهتم الباحثون بتحديد منطقة الكف ضمن المشهد وليس منطقة اليد بالكامل؛ وذلك نظراً إلى أن الكف وحده يكفي عادةً لمعرفة الإيماءة (gesture) التي يريد المستخدم إيصالها، ومن ثم التفاعل مع الآلة. تقوم الطرائق التقليدية بالتعرف مباشرة إلى منطقة الكف دون تعرف منطقة اليد بالكامل من خلال مجموعة من السمات المتعلقة بشكل الكف المراد التعرف إليه، أو من خلال تقسيم صورة الدخل إلى مجموعة من المناطق ومقارنتها بمجموعة مخزنة سلفاً من الأوضاع المراد تعرفها، ولكن هذه

المتبني به واعتماد المنطقة الواقعة ضمن عتبة معينة عنه مع الأخذ بالحسبان تقارب حجم نافذة اليد ضمن اللحظتين الحالية والتالية [23]، عند عدم وجود نافذة تحقق هذه الشروط نعدُّ أنّ اليد قد خرجت من المشهد ونعود لتطبيق مرحلة التهيئة؛ مما يسمح بإعادة تحديد منطقة اليد من جديد ضمن صورة الدخل.

3- النتائج:

نفذنا (implement) النظام المقترح باستخدام Visual C++ وبالاعتماد على مكتبة OpenCV لمعالجة الصور، إذ استطاع النظام معالجة فيديو بالأبعاد 640×480 بسرعة نحو 14 صورة بالثانية على معالج Intel i5 2.5GHz.

إن دقة تعرف اليد في الحل المقدم في هذا البحث تتعلق بخوارزمية التصنيف المستخدمة لتعلم العلاقة بين قيم معاملات المعايير المقترحة وبين وجود اليد البشرية ضمن المشهد؛ لذلك قمنا بتجربة مجموعة من هذه الخوارزميات وقياس دقة كل منها واعتماد الخوارزمية الأعلى دقة. حُسبت دقة الخوارزمية باستخدام مقياس F-Measure الذي يعتمد على قيمة كل من Precision و Recall، تُحسب هذه القيم عن طريق المعادلات:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - Measure = \frac{2Precision \cdot Recall}{Precision + Recall}$$

إذ:

TP: عدد الصور التي تحتوي يداً والتي صنفها النظام كصور تحتوي يداً.

FP: عدد الصور التي لا تحتوي يداً والتي صنفها النظام كصور تحتوي يداً.

FN: عدد الصور التي تحتوي يداً والتي صنفها النظام كصور لا تحتوي يداً.

الساعد ضمن الصورة، وهذا ما يسبب زيادة كبيرة في حجم منطقة اليد مقارنة بحجم منطقة الكف. بعد ذلك نقوم بتطبيق مرحلة الملاحقة على منطقة الكف بدلاً من منطقة اليد بالكامل.

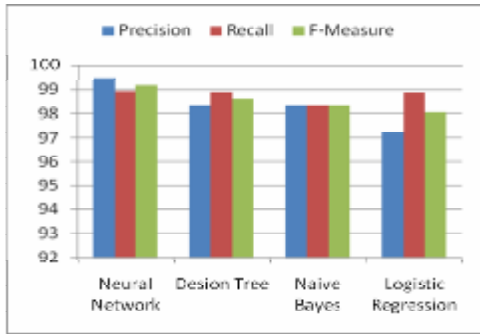
2-2- مرحلة الملاحقة: يمكن أن يوجد ضمن المشهد أكثر من يد بشرية واحدة، فقد تظهر في أثناء عملية التفاعل يد لمستخدم آخر أو جزء من يد المستخدم الأخرى؛ مما يتطلب إيجاد طريقة للحفاظ على تتبع يد بشرية واحدة في مثل هذه الحالات. يمكن تحديد الأجسام المتحركة وملاحقتها باستخدام خوارزمية واحدة مثل خوارزمية optical flow التي تقوم بمقارنة عناصر الصورة فيما بينها وتحديد المكان الجديد لكل منها، ومن ثم حساب متجه السرعة لكل عنصر صورة على حدة [22]، هذه الطريقة ليست عملية في حالة حركة اليد لأن عناصر الصورة الواقعة داخل اليد تحافظ على اللون نفسه خلال صورتين متتاليتين في حين يتغير لون عناصر الصورة الواقعة على حافات اليد فقط؛ مما يعطي انطباعاً بأن الإطار الخارجي لليد فقط هو الذي يتحرك. للتخلص من هذه المشكلة يمكن دراسة بعض النقاط المميزة ضمن اليد مثل النقاط التي تمثل التقاء الخطوط أو حافات اليد، ولكن هذه الطريقة لا تعطي نتائج جيدة عند اختفاء هذه النقاط بسبب تغيير شكل اليد في أثناء حركتها، كما أنّ استخدام خوارزمية واحدة لتحديد الأجسام المتحركة وملاحقتها قد يسبب نتائج غير صحيحة في حالة وجود أجسام متحركة ضمن خلفية المشهد؛ لذلك لا بدّ من استخدام خوارزمية تعتمد على موضع الجسم ككل وليس على موضع نقاط مميزة ضمنه. في هذا البحث اعتمدنا على حساب المشتق الأول والثاني لموضع اليد لحساب سرعة حركة اليد وتسارعها والاعتماد عليها للتنبؤ بموضع اليد في اللحظة التالية، ومن ثمّ مقارنة مواضع المناطق المرشحة بالموضع

TN: عدد الصور التي لا تحتوي يدًا والتي صنفها النظام كصور لا تحتوي يدًا.

الجدول (1) دقة خوارزميات التعلّم التلقائي المستخدمة

Algorithm	Precision	Recall	F-Measure	Error Rate
Neural Network	99.46	98.92	99.19	0.008929
Desion Tree	98.34	98.89	98.62	0.01148
Naive Bayes	98.33	98.33	98.33	0.012755
Logistic Regression	97.24	98.88	98.05	0.021684

قمتنا بتجربة 4 خوارزميات تصنيف دخل كل منها هو قيم المعاملات الأربع المقترحة (Eccentricity و DefectsCount و CbCrCount و FGNearCount) وخرجه قيمة منطقية true أو false تدلُّ على وجود يد أو عدم وجودها، وهذه الخوارزميات هي: Neural Network و Desion Tree و Naive Bayes و Logistic Regression. لتعلّم العلاقة بين المعاملات وبين وجود اليد البشرية حُصِّلت 1120 صورة من 4 ملفات فيديو لمستخدمين مختلفين في أثناء استخدامهم للنظام ضمن ظروف إضاءة غرفة طبيعية، ومن تمَّ صنَّفَت المناطق المرشحة ضمن هذه الصور بشكل يدوي إلى قسمين: مناطق تحتوي على يد بشرية، ومناطق لا تحتوي على يد بشرية، وبعد ذلك جرى حساب قيم المعاملات الأربع لهذه المناطق. جرى تقسيم هذه البيانات إلى قسمين: قسم للتعلّم يحتوي على 70% من البيانات مختارة عشوائياً، وقسم للاختبار يحتوي على باقي البيانات، حُسِبَت نسبة خطأ التعلّم (Error Rate) لكل خوارزمية باستخدام طريقة 10-fold cross validation التي تعتمد على تقسيم العينات المستخدمة في التعلّم إلى 10 أقسام وبناء 10 نماذج مقابلة لهذه الأقسام، يُختَبَرُ كل نموذج على القسم المقابل له بعد تدريبه باستخدام بقية الأقسام، ثم يُؤخَذُ متوسط أخطاء هذه النماذج جميعها للحصول على خطأ التعلّم. يبين دقة كل من خوارزميات التصنيف المذكورة.



الشكل (9) مقارنة بين خوارزميات التعلّم التلقائي

مقارنة بالبحوث الأخرى التي استطاعت تعرّف اليد البشرية وملاحقتها بغض النظر عن شكلها، نلاحظ أن تلك البحوث اعتمدت على النقاط فيديو للإنسان بالكامل يظهر فيه الرأس واليدان والرجلان، ومن تمَّ حُدِّدَت القمم الموجودة في حافات جسم الإنسان واختيرت القمم الممثلة لليدين [24-25]، أي إنَّ هذه البحوث افترضت أن جسم الإنسان ظاهر بشكل كامل طول عملية التفاعل. عملياً لا يمكننا دوماً ضمان ظهور كامل جسم الإنسان في فيديو الدخل، وخاصة في تطبيقات التفاعل ضمن المباني (indoors) بسبب الأغراض الموجودة ضمن البيئة التي قد تحجب أجزاء كبيرة من جسم الإنسان مثل المكتب أو الطاولة التي عادة ما توجد أمام المستخدم، وهذا ما يميّز

قمتنا بتجربة 4 خوارزميات تصنيف دخل كل منها هو قيم المعاملات الأربع المقترحة (Eccentricity و DefectsCount و CbCrCount و FGNearCount) وخرجه قيمة منطقية true أو false تدلُّ على وجود يد أو عدم وجودها، وهذه الخوارزميات هي: Neural Network و Desion Tree و Naive Bayes و Logistic Regression. لتعلّم العلاقة بين المعاملات وبين وجود اليد البشرية حُصِّلت 1120 صورة من 4 ملفات فيديو لمستخدمين مختلفين في أثناء استخدامهم للنظام ضمن ظروف إضاءة غرفة طبيعية، ومن تمَّ صنَّفَت المناطق المرشحة ضمن هذه الصور بشكل يدوي إلى قسمين: مناطق تحتوي على يد بشرية، ومناطق لا تحتوي على يد بشرية، وبعد ذلك جرى حساب قيم المعاملات الأربع لهذه المناطق. جرى تقسيم هذه البيانات إلى قسمين: قسم للتعلّم يحتوي على 70% من البيانات مختارة عشوائياً، وقسم للاختبار يحتوي على باقي البيانات، حُسِبَت نسبة خطأ التعلّم (Error Rate) لكل خوارزمية باستخدام طريقة 10-fold cross validation التي تعتمد على تقسيم العينات المستخدمة في التعلّم إلى 10 أقسام وبناء 10 نماذج مقابلة لهذه الأقسام، يُختَبَرُ كل نموذج على القسم المقابل له بعد تدريبه باستخدام بقية الأقسام، ثم يُؤخَذُ متوسط أخطاء هذه النماذج جميعها للحصول على خطأ التعلّم. يبين دقة كل من خوارزميات التصنيف المذكورة.

يقارن الشكل (9) بيانياً بين خوارزميات التصنيف ويُظهر أن طريقة الشبكات العصبونية (Neural

الطريقة المقدمّة عن البحوث الأخرى لأنها لا تتطلب ظهور نسبة معينة من جسم الإنسان ضمن المشهد.



الشكل (11) بعض القوائم الخاصة بنظام التلفاز التفاعلي الافتراضي

4- التطبيق العملي:

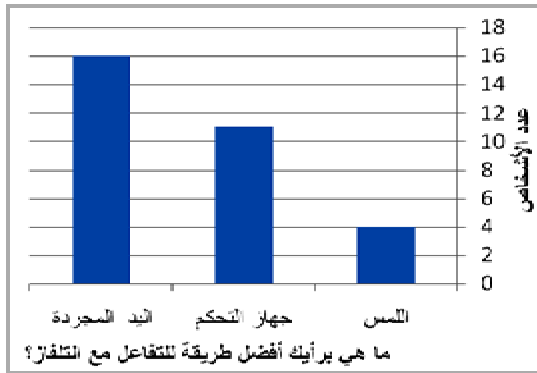
لمعرفة آراء المستخدمين بشكل عملي عن الطريقة المقدمّة في هذا البحث قمنا بتطوير برنامج تلفاز تفاعلي افتراضي يحاكي الأنظمة المستخدمة في أجهزة التلفاز الحقيقية، يمكن المستخدم عن طريق تحريك يده المجردة أمام الكاميرا من خفض صوت التلفاز ورفعها فضلاً عن تغيير المحطة الحالية باختيار إحدى المحطات المخزنة سلفاً ضمن مجموعة من قوائم المفضلات. يستطيع المستخدم محاكاة الضغط على أزرار الأسهم الأربعة في جهاز التحكم (أعلى، أسفل، يمين، يسار) عن طريق تحريك يده بأحد هذه الاتجاهات، وذلك للتنقل ضمن القوائم التي يعرضها نظام التلفاز التفاعلي الافتراضي، كما يستطيع اختيار أحد عناصر هذه القوائم عن طريق فتح أصابع يده وإغلاقها بدلاً من الضغط على زر "OK" الموجود في جهاز التحكم. يُظهر الشكل (10) بعضاً من أشكال القوائم الخاصة بالتلفاز التفاعلي الافتراضي.

إنّ تحويل جهاز التلفاز الافتراضي المقترح إلى منتج واقعي يُستخدم في المنزل يوفر على المستخدمين عناء البحث عن جهاز التحكم في أنحاء المنزل، كما يغنيهم عن اقتناء جهاز تحكم خاص لكل جهاز يُوصَل على التلفاز وتُغيّر بطارياته باستمرار إذ تكفي كاميرا واحدة

مثبتة على جهاز التلفاز للقيام بمهام أجهزة التحكم مجتمعة.

لتقييم التلفاز التفاعلي الافتراضي الذي قمنا بتطويره لجأنا إلى إجراء دراسة عملية على 31 مستخدماً (21 من الذكور و 10 من الإناث) تراوح أعمارهم بين 21 و 55 عاماً، إذ طُلبَ من كل منهم تجربة عملية التفاعل مع التلفاز الافتراضي مدة دقيقتين، ومن ثم قُيِّمَت تجربتهم من خلال الإجابة عن مجموعة من الأسئلة. طُلبَ من المستخدمين تقييم تجربتهم في التفاعل مع التلفاز باستخدام اليد المجردة مقارنة بالطريقة التقليدية التي تعتمد على جهاز التحكم عن بعد، وذلك بمقياس من 3 وحتى 3- (الأرقام الموجبة تدلُّ على ربحان كفة اليد المجردة في حين تدلُّ الأرقام السالبة على ربحان كفة جهاز التحكم) إذ يُعطى تقييم لكل من المتعة والدقة والراحة، يظهر الشكل (11) (أ) نتيجة هذا التقييم حيث يظهر بوضوح تفوق طريق التفاعل باستخدام اليد المجردة من حيث متعة الاستخدام على الرغم من انخفاض دقتها بشكل طفيف مقارنة بجهاز التحكم، في حين تتساوى الطريقتان تقريباً من حيث راحة المستخدم. وأخيراً طُلبَ من المستخدمين اختيار الطريقة الأكثر ملاءمة للتفاعل مع التلفاز من ضمن مجموعة الطرائق الآتية: التفاعل باللمس، أو التفاعل باستخدام جهاز

وملاحقة اليد البشرية المتحركة أمام كاميرا عادية في بيئة غير معلومة مسبقاً، وتحتوي عن خلفية غير ثابتة بالضرورة بسرعة تزيد على 14 صورة بالثانية وبدقة تعرّف تقارب 99% ضمن ظروف إضاءة الغرفة. كما دلت آراء المستخدمين حول التطبيق العملي المتمثل بالتلفاز التفاعلي الافتراضي على تفضيلهم طريقة التفاعل باليد على جهاز التحكم من جهة متعة الاستخدام على الرغم من حصولهم على دقة أداء أقل.

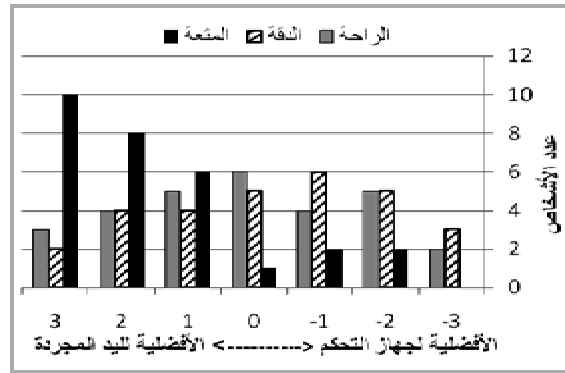


(ب)

التحكم، أو التفاعل عن طريق اليد المجردة، فكانت اليد المجردة الأكثر تفضيلاً كما يظهر الشكل (11)(ب).

5- الخلاصة والآفاق المستقبلية:

قدّمنا في هذا البحث طريقة جديدة معتمدة على المظهر للتعرف إلى يد بشرية مجردة تتحرك أمام كاميرا رقمية بشكل مستقل عن شكلها، وذلك بالاعتماد على مجموعة من المعايير المستوحاة من خصائص اليد. أظهرت النتائج العملية أنّ الطريقة المقترحة قادرة على التعرف



(أ)

الشكل (11) نتيجة تقييم المستخدمين لطريقة التفاعل باليد المجردة

العديد من التطبيقات الأخرى مثل تطبيقات فهم لغة الصم والبكم أو تعرّف الإيماءات، وهذا ما يسهم في زيادة فعالية هذه التطبيقات بسبب معالجتها مباشرة لمنطقة اليد بدلاً من كامل الصورة.

يمكن تطبيق الطريقة المقترحة في التفاعل المباشر مع جهاز التلفاز العادي بدلاً من التلفاز الافتراضي عن طريق ربط جهاز الحاسب بجهاز تحكم عن بعد، أو دارة للأشعة تحت الحمراء تقوم بإرسال الأوامر إلى التلفاز، كما يمكن استخدامها لتحديد منطقة اليد كخطوة أولى في

12. D. Huan and E. Charbon, "A virtual keyboard system based on Multi-Level Feature Matching," in Human System Interactions, 2008 Conference on, 2008, pp. 176-181.
13. C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, pp. 747-757, 2000.
14. L. Li, et al., "Foreground object detection from videos containing complex background," presented at the Proceedings of the eleventh ACM international conference on Multimedia, Berkeley, CA, USA, 2003.
15. M. Yuan, et al., "Robust Hand Tracking Using a Simple Color Classification Technique," The International Journal of Virtual Reality, vol. 8, pp. 7-12, 2009.
16. Y. Huang, et al., "Real time face detection based on skin tone detector," International Journal of Computer Science and Network Security, vol. 9, p. 71, 2009.
17. C. Manresa-Yee, et al., "Hand Tracking and Gesture Recognition for Human-Computer Interaction," Electronic Letters on Computer Vision and Image Analysis, vol. 5, pp. 96-104, 2007.
18. T. E. de Campos, "3D Visual Tracking of Articulated Objects and Hands," Department of Engineering Science, University of Oxford Trinity Term 2006 This thesis is submitted to the Department of Engineering Science, University of Oxford, 2006.
19. G. R. Bradski and J. Davis, "Motion segmentation and pose recognition with motion history gradients," in Applications of Computer Vision, 2000, Fifth IEEE Workshop on., 2000, pp. 238-244.
20. S. M. Nadgeri, et al., "Hand Gesture Recognition Using CAMSHIFT Algorithm," in 3rd International Conference of Emerging Trends in Engineering and Technology (ICETET), 2010, pp. 37-41.
21. K. Homma and E. Takenaka, "An image processing method for feature extraction of space-occupying lesions," Journal of Nuclear Medicine, vol. 26, p. 1472, 1985.
22. B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artificial intelligence, vol. 17, pp. 185-203, 1981.
23. N. Soontranon, et al., "Improved face and hand tracking for sign language recognition," in Information Technology: Coding and Computing, 2005. ITCC 2005. International

المراجع:

1. J. d. R. Millán, et al., "Brain-actuated interaction," Artificial intelligence, vol. 159, pp. 241-259, 2004.
2. J. M. Rehg and T. Kanade, "Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking," presented at the Proceedings of the Third European Conference-Volume II on Computer Vision - Volume II, 1994.
3. K. Dorfmüller-Ulhaas and D. Schmalstieg, "Finger tracking for interaction in augmented environments," in Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on, 2001, pp. 55-64.
4. P. Mistry and P. Maes, "SixthSense: a wearable gestural interface," presented at the ACM SIGGRAPH ASIA 2009 Sketches, Yokohama, Japan.
5. R. Y. Wang, et al., "Real-time hand-tracking with a color glove," ACM Trans. Graph., vol. 28, pp. 1-8, 2009.
6. F. S. Chen, et al., "Hand gesture recognition using a real-time tracking method and hidden Markov models," Image and Vision Computing, vol. 21, pp. 745-758, 2003.
7. Y. Duan-Duan, et al., "An effective robust fingertip detection method for finger writing character recognition system," in Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on, 2005, pp. 4991-4996 Vol. 8.
8. P. Song, et al., "Vision-based 3D Finger Interactions for Mixed Reality Games with Physics Simulation," The International Journal of Virtual Reality, vol. 8, pp. 1-6, 2009.
9. G. Haiying, et al., "The isometric self-organizing map for 3D hand pose estimation," in Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, 2006, pp. 263-268.
10. B. Stenger, et al., "Model-based 3D tracking of an articulated hand," in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001, pp. II-310-II-315 vol.2.
11. C.-C. Lien and C.-L. Huang, "Model-based articulated hand motion tracking for gesture recognition," Image and Vision Computing, vol. 16, pp. 121-134, 1998.

- Conference on, 2005, pp. 141-146 Vol. 2.
24. J. Chia-Feng, et al., "Computer Vision-Based Human Body Segmentation and Posture Estimation," Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions, vol. 39, pp. 119-133, 2009.
 25. I. Haritaoglu, et al., "W4: real-time surveillance of people and their activities," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, pp. 809-830, 2000.

مسرد المصطلحات:

analysis by synthesis	التحليل بالتركيب
appearance	المظهر
approach	طريقة
classification	تصنيف
edges	حافات
features	السمات
gesture	إيماءة
glove	قفاز
hardware	عتاد
hidden layer	طبقة مخفية
holes	ثغرات
implement	تنفيذ
indoors	ضمن المباني
inverse kinematics	الحركية العكسية
kinematics	الحركية
markers	المعلّقات
models	نماذج
normalization	تقييس
neural network	شبكة عصبونية
orientation	توجه
pixel	عنصر صورة
position	موضع
real-time	الزمن الحقيقي
robots	الروبوتات
stereo	مجسّم
tMHI	صورة الحركة الزمنية

تاريخ ورود البحث إلى مجلة جامعة دمشق 2011/12/11