***Viola–Jones object detection***

المشكلة التي يجب حلها هي اكتشاف الوجوه في الصورة. يمكن للإنسان القيام بذلك بسهولة ، لكن الكمبيوتر يحتاج إلى تعليمات وقيود دقيقة. لجعل المهمة أكثر قابلية للإدارة ، تتطلب Viola-Jones رؤية كاملة لوجوه أمامية منتصبة. وبالتالي ، لكي يتم الكشف عنها ، يجب أن يشير الوجه بأكمله نحو الكاميرا ولا يجب إمالته إلى أي جانب. على الرغم من أنه يبدو أن هذه القيود قد تقلل من فائدة الخوارزمية إلى حد ما ، لأن خطوة الكشف غالبًا ما تتبعها خطوة التعرف ، إلا أن هذه القيود على الوضع مقبولة تمامًا من الناحية العملية.

***أنواع الميزات وتقييمها***

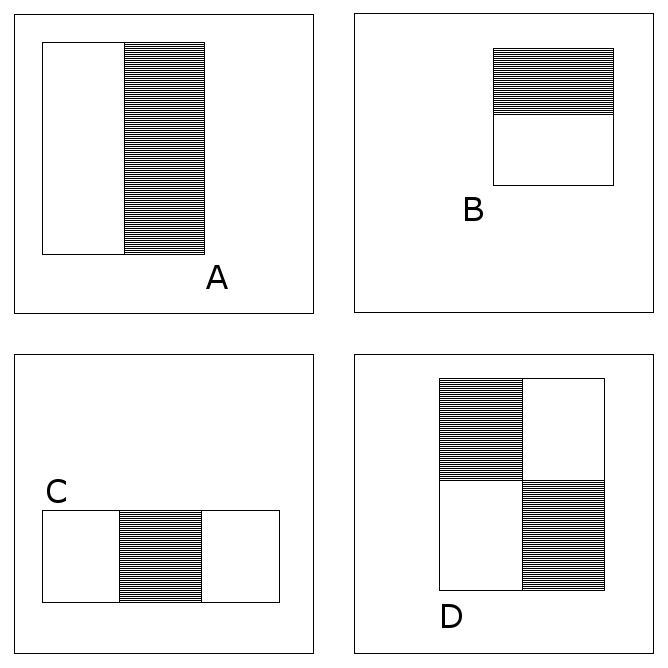
خصائص خوارزمية فيولا-جونز والتي تجعلها خوارزمية كشف جيدة هي:

1. متين - معدل اكتشاف مرتفع جدًا (معدل إيجابي حقيقي) ومعدل إيجابي كاذب منخفض جدًا دائمًا.
2. الوقت الفعلي - للتطبيقات العملية ، يجب معالجة إطارين على الأقل في الثانية.
3. اكتشاف الوجه فقط (وليس التعرف) - الهدف هو التمييز بين الوجوه وغير الوجوه (الاكتشاف هو الخطوة الأولى في عملية التعرف).

***تتكون الخوارزمية من أربع مراحل:***

1. اختيار ميزة Haar
2. تكوين صورة متكاملة
3. تدريب Adaboost
4. المصنفات المتتالية

تتضمن الميزات التي يبحث عنها إطار عمل الاكتشاف عالميًا مجاميع بكسلات الصورة داخل مناطق مستطيلة. على هذا النحو ، فإنها تحمل بعض التشابه مع وظائف أساس Haar ، والتي تم استخدامها سابقًا في مجال اكتشاف الكائنات القائمة على الصور. ومع ذلك ، نظرًا لأن الميزات المستخدمة من قبل Viola and Jones تعتمد جميعها على أكثر من منطقة مستطيلة واحدة ، فهي عمومًا أكثر مركب. يوضح الشكل الموجود على اليمين الأنواع الأربعة المختلفة من الميزات المستخدمة في إطار العمل. قيمة أي ميزة معينة هي مجموع البكسل داخل مستطيلات واضحة مطروحًا من مجموع البكسل داخل المستطيلات المظللة. تعتبر الميزات المستطيلة من هذا النوع بدائية عند مقارنتها بالبدائل مثل المرشحات القابلة للتوجيه. على الرغم من أنها حساسة للسمات الرأسية والأفقية ، إلا أن ملاحظاتها تكون أكثر خشونة.



***كيف تعمل***

هناك مرحلتان في خوارزمية Viola-Jones:

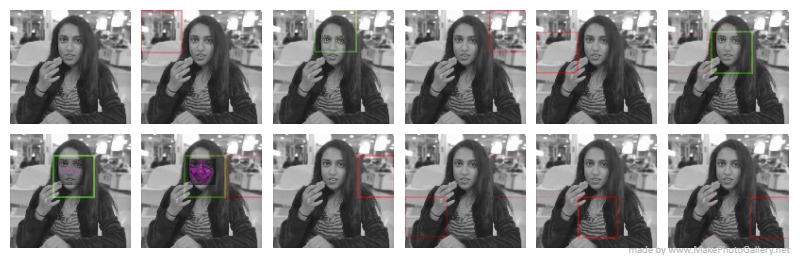
1. تدريب
2. كشف

يأتي التدريب قبل الاكتشاف ، ولكن من أجل التوضيح ، سأناقش الاكتشاف أولاً.

***كشف***

تم تصميم Viola-Jones للوجوه الأمامية ، لذا فهي قادرة على اكتشاف الوجه بشكل أفضل بدلاً من الوجوه التي تبدو جانبية ، لأعلى أو لأسفل. قبل اكتشاف الوجه ، يتم تحويل الصورة إلى درجات رمادية ، لأنه من الأسهل التعامل معها ووجود بيانات أقل للمعالجة. تقوم خوارزمية Viola-Jones أولاً باكتشاف الوجه على الصورة ذات التدرج الرمادي ثم البحث عن الموقع على الصورة الملونة.

تحدد Viola-Jones الخطوط العريضة لمربع وتبحث عن وجه داخل المربع. إنه يبحث بشكل أساسي عن هذه الميزات الشبيهة بhaar ، والتي سيتم شرحها لاحقًا. يتحرك المربع خطوة إلى اليمين بعد المرور بكل قطعة في الصورة. في هذه الحالة ، لقد استخدمت حجم صندوق كبير واتخذت خطوات كبيرة للشرح ، ولكن بشكل عام ، يمكنك تغيير حجم الصندوق وحجم الخطوة وفقًا لاحتياجاتك.



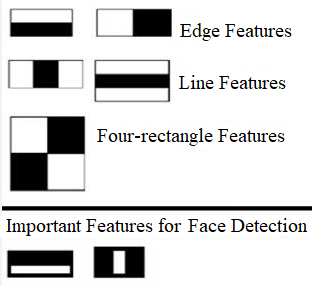
مع خطوات أصغر ، يكتشف عدد من المربعات ميزات تشبه الوجه (ميزات تشبه Haar) وبيانات كل هذه المربعات مجتمعة ، تساعد الخوارزمية على تحديد مكان الوجه.

***ميزات تشبه هار***

تمت تسمية الميزات التي تشبه هار على اسم ألفريد هار ، عالم الرياضيات الهنغاري في القرن التاسع عشر الذي طور مفهوم موجات هار (نوعًا ما يشبه سلف السمات الشبيهة بالهار). تُظهر الميزات أدناه مربعًا به جانب فاتح وجانب مظلم ، وهي الطريقة التي يحدد بها الجهاز ماهية الميزة. في بعض الأحيان يكون أحد الجانبين أفتح من الآخر ، كما هو الحال في حافة الحاجب. في بعض الأحيان ، قد يكون الجزء الأوسط أكثر لمعانًا من الصناديق المحيطة ، والتي يمكن تفسيرها على أنها أنف.

هناك 3 أنواع من الميزات المشابهة لـ Haar التي حددتها فيولا وجونز في بحثهما:

* ميزات الحافة
* ميزات الخط
* ميزات رباعية الجوانب

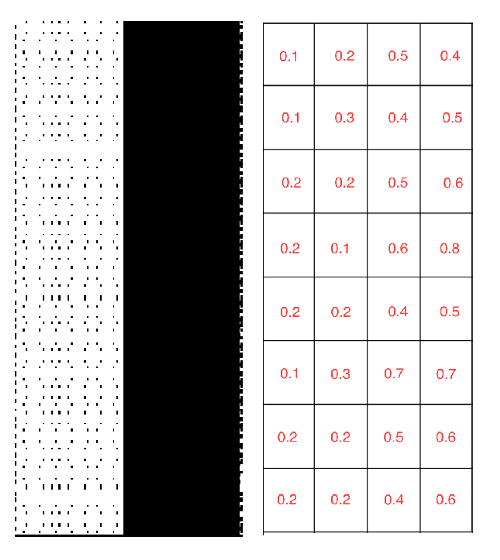




تساعد هذه الميزات الجهاز في فهم ماهية الصورة. تخيل كيف ستبدو حافة الجدول في صورة أبيض وأسود. سيكون أحد الجانبين أفتح من الآخر ، مما يؤدي إلى إنشاء تلك الحافة مثل ميزة الأبيض والأسود كما ترون في الصورة أعلاه.

في ميزتين مهمتين لاكتشاف الوجه ، تصف الميزات الأفقية والعمودية شكل الحاجبين والأنف ، على التوالي ، للجهاز.

بالإضافة إلى ذلك ، عند فحص الصور ، يكون لكل ميزة قيمة خاصة بها. الحساب سهل للغاية: اطرح المنطقة البيضاء من المنطقة السوداء. على سبيل المثال ، انظر إلى الصورة أدناه.



تخيل أن الميزة الشبيهة بالهار تم تحويلها إلى شبكة. يمثل كل مربع بكسل. للتوضيح ، اخترت شبكة 4 × 8 ، ولكن في الواقع ، سيكون هناك عدد أكبر من البكسل وبالتالي شبكة أكبر بكثير لميزة معينة. تمثل الأرقام الموجودة في الصناديق ظلام الميزات. كلما زاد ارتفاعه ، زاد قتامة البكسل. وهكذا ، يمكنك أن ترى أن الأرقام أعلى في الجانب الأيمن منها في الجانب الأيسر. الآن ، إذا قمت بجمع الأرقام على العمودين الأيسر (الأبيض) ، وطرحها من مجموع الأعمدة الموجودة على الجانب الأيمن ، فستحصل على قيمة الميزة المعينة.

إذن في هذه الحالة ، قيمة الميزة لدينا هي →

(0.5 + 0.4 + 0.5 +0.6 + 0.4 + 0.7 + 0.5 + 0.4 +

0.4 + 0.5 + 0.6 + 0.8 + 0.5 + 0.7 + 0.6 + 0.6) -

(0.1 + 0.1 + 0.2 + 0. 2+ 0.2 + 0.1 + 0.2 + 0.2 +

0.2 + 0.3 + 0.2 + 0.1 + 0.2 + 0.3 + 0.2 + 0.2)

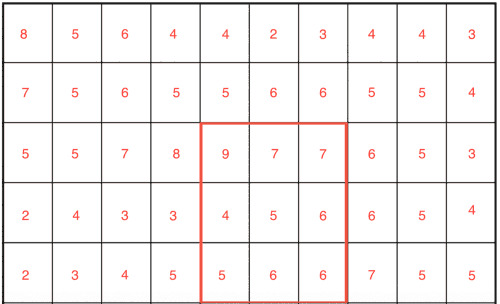
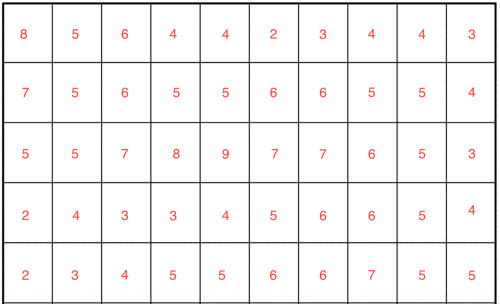
B-W = 8.7 - 3

= 5.7

***صورة متكاملة***

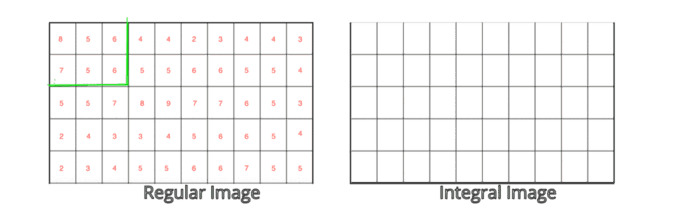
لذلك في القسم الأخير ، قمنا بحساب قيمة الميزة. في الواقع ، يمكن أن تكون هذه الحسابات مكثفة للغاية لأن عدد وحدات البكسل سيكون أكبر بكثير داخل ميزة كبيرة.

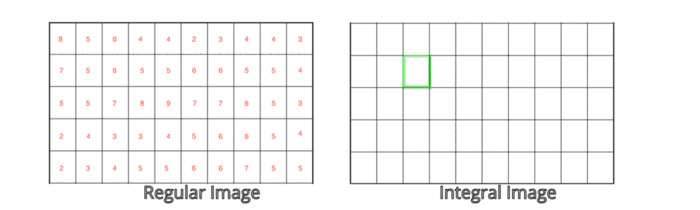
تلعب الصورة المتكاملة دورها في السماح لنا بإجراء هذه الحسابات المكثفة بسرعة حتى نتمكن من فهم ما إذا كانت سمة لعدد من الميزات تتناسب مع المعايير.



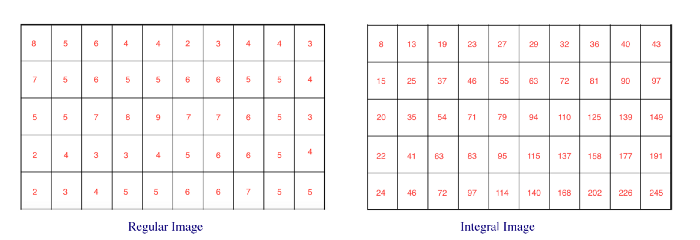
تخيل الآن أن الذي تم تمييزه باللون الأحمر هو شبكتنا لميزة معينة ، ونحن نحاول حساب قيمة هذه الميزة. عادةً ما نضيف المربعات فقط ، ولكن نظرًا لأن ذلك يمكن أن يكون مكثفًا من الناحية الحسابية ، فسننشئ صورة متكاملة.

لحساب قيمة مربع واحد في صورة متكاملة ، نأخذ مجموع كل المربعات على يساره. توضح الصورة أدناه مثالاً:

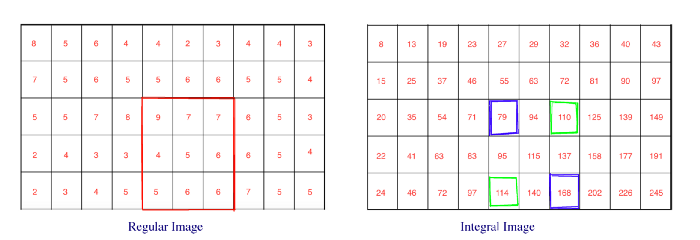




يتم حساب المربع الأخضر في الصورة المتكاملة كمجموع المنطقة المميزة في الصورة العادية. إذا فعلنا هذا لكل صندوق ، فسيكون لدينا تسلسل يمر عبر الشبكة وقد يبدو مثل الصورة أدناه.



لنلقِ نظرة على قيمة المثال الذي اخترته سابقًا:



كل ما علينا القيام به هو إلقاء نظرة على الزوايا الأربع لميزتنا ، وإضافة اللون الأرجواني ، وطرح اللون الأخضر.

→ 168–114 + 79–110 = 23

***فلماذا نستخدم الصورة المتكاملة؟***

نظرًا لأن الميزات التي تشبه Haar هي في الواقع مستطيلة ، وتسمح لنا عملية الصورة المتكاملة بالعثور على ميزة داخل صورة ما بسهولة بالغة لأننا نعرف بالفعل قيمة مجموع مربع معين ولإيجاد الفرق بين مستطيلين في الصورة العادية ، تحتاج فقط إلى طرح مربعين في الصورة المتكاملة. لذا ، حتى لو كان لديك 1000 × 1000 بكسل في شبكتك ، فإن طريقة الصورة المتكاملة تجعل العمليات الحسابية أقل كثافة ويمكن أن توفر الكثير من الوقت لأي نموذج للكشف عن الوجه.

***مصنفات التدريب***

الآن بعد أن غطينا الاكتشاف ، دعنا ننتقل إلى التدريب.

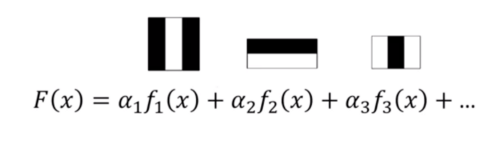
إذن ماذا نعني بالضبط نقول التدريب؟ نحن ندرب الآلة على تحديد هذه الميزات. نقوم بتزويدها بالمعلومات ، ومن ثم ندربها على التعلم من المعلومات للتنبؤ. لذلك في النهاية ، تحدد الخوارزمية حدًا أدنى لتحديد ما إذا كان يمكن تصنيف شيء ما كميزة أم لا.

تقوم الخوارزمية بتقليص الصورة إلى 24 × 24 وتبحث عن الميزات المدربة داخل الصورة. تحتاج إلى الكثير من بيانات صورة الوجه لتتمكن من رؤية الميزات بأشكال مختلفة ومتنوعة. لهذا السبب نحتاج إلى توفير الكثير من بيانات صورة الوجه للخوارزمية حتى يمكن تدريبها. قام فيولا وجونز بتغذية الخوارزمية الخاصة بهم بـ 4960 صورة (تم تسمية كل صورة يدويًا). بالنسبة لبعض الصور ، يمكنك تغذية صورة معكوسة لصورة معينة ، والتي قد تكون معلومات جديدة للكمبيوتر.

ستحتاج أيضًا إلى توفير خوارزمية الصور غير الوجهية حتى تتمكن من التمييز بين الفئتين. قدمت فيولا وجونز الخوارزمية 9544 صورة غير وجهية. ضمن هذه الصور ، قد تبدو بعض الصور مشابهة لميزات الوجه ، لكن الخوارزمية ستفهم الميزات التي من المرجح أن تكون على الوجه والميزات التي من الواضح أنها لن تكون على الوجه.

***تعزيز التكيف (AdaBoost)***

تتعلم الخوارزمية من الصور التي نقدمها لها وهي قادرة على تحديد الإيجابيات الخاطئة والسلبيات الحقيقية في البيانات ، مما يتيح لها أن تكون أكثر دقة. سنحصل على نموذج عالي الدقة بمجرد النظر في جميع المواضع والتركيبات الممكنة لهذه الميزات. يمكن أن يكون التدريب مكثفًا للغاية نظرًا لجميع الاحتمالات والمجموعات المختلفة التي يجب عليك التحقق منها لكل إطار أو صورة.



لنفترض أن لدينا معادلة لميزاتنا تحدد معدل النجاح (كما هو موضح في الصورة) ، مع f1 و f2 و f3 كميزات و a1 و a2 و a3 كأوزان خاصة بالميزات. تُعرف كل ميزة بمصنف ضعيف. يسمى الجانب الأيسر من المعادلة F (x) المصنف القوي. نظرًا لأن المصنف الضعيف قد لا يكون جيدًا ، فإننا نحصل على مصنف قوي عندما يكون لدينا مزيج من اثنين أو ثلاثة من المصنفات الضعيفة. بينما تستمر في الإضافة ، تصبح أقوى وأقوى. هذا يسمى الفرقة. تريد التأكد من أن لديك أهم الميزات في المقدمة ، ولكن السؤال هو كيف تجد أهم الميزات أو "الأفضل"؟ هذا هو المكان الذي يلعب فيه Adaptive Boosting.

على سبيل المثال ، لنفترض أن لديك 10 صور: 5 صور للوجه و 5 صور غير وجه.

لذلك تجد ميزة مهمة ، أفضل ميزة وفقًا لك وتستخدمها لعمل تنبؤ.

يعطينا هذا النموذج 3 من 5 إيجابيات حقيقية و 2 من 5 سلبيات حقيقية.



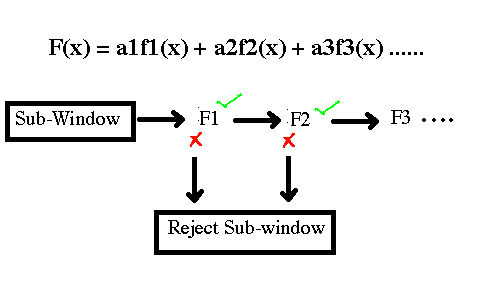
تتنبأ بشكل صحيح لهذه الصور ولكن هناك أيضًا بعض الأخطاء: 3 إيجابيات خاطئة و 2 سلبيات زائفة. لذلك لم يتم العثور على الميزة في هاتين الصورتين ، فهما في الواقع وجوه. حيث وجدت الميزة في 3 صور غير للوجه.



في الخطوة التالية ، يستخدم التعزيز التكيفي ميزة أخرى ، وهي الميزة التي تكمّل أفضل ميزة لدينا حاليًا. لذلك فهي لا تبحث عن ثاني أفضل ميزة ، ولكنها ميزة تكمل أفضل ميزة حالية. لذلك فهي تزيد من أهمية الصور التي أخطأت فيها باعتبارها سلبيات خاطئة ، وتجد الميزة التالية الأفضل التي تناسب هذه الصور ، بطريقة ما ، تزيد من وزن هذه الصور على الخوارزمية الكلية. لذلك ، مع إضافة ميزات جديدة ، سننزل إلى صورة واحدة في النهاية ستعطى وزنًا أكبر. بمجرد أن يتم تحسين الخوارزمية وتكون قادرة على حساب جميع الإيجابيات والسلبيات بشكل صحيح ، ننتقل إلى الخطوة التالية: التعاقب.

***التتالي***

التتالي هو نوع آخر من "الاختراق" لتعزيز سرعة ودقة نموذجنا. لذلك نبدأ بأخذ إطار فرعي وداخل هذه النافذة الفرعية ، نأخذ أهم ميزة أو أفضل ميزة لدينا ونرى ما إذا كانت موجودة في الصورة داخل النافذة الفرعية. إذا لم يكن في النافذة الفرعية ، فلن ننظر حتى إلى النافذة الفرعية ، بل نتجاهلها فقط. ثم إذا كانت موجودة ، فإننا ننظر إلى الميزة الثانية في النافذة الفرعية. إذا لم يكن موجودًا ، فنحن نرفض النافذة الفرعية. ننتقل إلى عدد الميزات التي لديها ، ونرفض النوافذ الفرعية بدون الميزة. قد تستغرق التقييمات بضع ثوانٍ ولكن نظرًا لأنه يتعين عليك القيام بذلك لكل ميزة ، فقد يستغرق الأمر الكثير من الوقت. يؤدي التتابع إلى تسريع هذه العملية كثيرًا ، ويمكن للآلة تقديم نتائج أسرع بكثير.



***خاتمة***

فقط لإضافة بعض الملاحظات الختامية حول خوارزمية فيولا جونز:

- تم تطوير الخوارزمية في عام 2001 من قبل بول فيولا ومايكل جونز ، وهي الأولى من نوعها ، وكانت تستخدم في المقام الأول لتطبيقات اكتشاف الوجه.

- هناك خطوتان للخوارزميات: هناك تدريب على صور الوجه وغير الوجه ثم هناك الاكتشاف الفعلي.

- لدينا خطوتين للتدريب: تدريب المصنفات و Adaboost

- لدينا خطوتان للكشف: اكتشاف الميزات الشبيهة بالعار وإنشاء صورة متكاملة

- تعد Viola-Jones واحدة من أقوى الخوارزميات في عصرها ، وعلى الرغم من وجود نماذج أفضل اليوم ، فقد وضعت Viola-Jones الأساس لها في مجال اكتشاف الوجه.

***المصادر***

الكشف السريع عن الأشياء باستخدام سلسلة معززة من الميزات البسيطة ؛ جونز وفيولا.

https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf

الإطار العام لاكتشاف الأشياء ؛ CP Papageorgiou et al. https://www.researchgate.net/publication/3766402\_General\_framework\_for\_object\_detection

تعزيز استرجاع الصور ؛ فيولا وتيو http://www.ee.columbia.edu/~sfchang/course/spr/papers/boosting-image-retrieval.pdf