# June 6,2014 a

EYE DETECTION AND GAZE ESTIMATION

Tóm tắt: Một kỹ thuật

nhận diện mắt và ước lượng đã được phát triển trong Matlab cho các trường hợp phổ biến và trong đó người sử dụng ngồi trước mặt ảnh và màn hình. Kỹ thuật như vậy có thể cho phép giao diện của máy tính xác định đầu vào của sự vận động mắt không hiệu quả thể hiện trạng thái tự nhiên của người sử dụng. Hơn nữa, nhiều ứng dụng như đánh giá sự mệt mỏi của người lái xe, giám sát, quảng cáo sẽ được hưởng lợi từ công nghệ đó. Vị trí mắt trong hình ảnh là một vấn đề đã được giải quyết trước đó; tuy nhiên, “hộp đen” công nghệ nhận diện mắt không cho phép chúng ta linh động trong việc hiểu được cốt lõi của chức năng được yêu cầu về giao diện hướng nhìn.

Để đánh giá hiệu quả của công nghệ này, 2 thí nghiệm đã được thực hiện. Đầu tiên, nhận diện mắt ( định vị mống mắt và hốc mắt ) được thực hiện hiện trên những bức ảnh từ việc tập huấn với tỷ lên thành công lên đến 50% trong việc tìm kiếm cả 2 mắt với số không tuyệt đối, 94% thành công trong việc tìm với ít nhất một mắt của chủ thể, và 17% thành công trong việc tìm cả 2 mắt của chủ thể không mang kính. Thứ hai, việc ước lượng đánh giá được thực hiện trên chủ thể đang quan sát một đối tượng trong phòng.

IGiới thiệu

Giá trị đáng kể có thể thu được từ việc xác định nơi trên màn hình người sử dụng đang tìm kiếm. Động lực của chúng tôi cho việc phát triển công nghệ phát hiện mắt và ước lương đánh giá chủ yếu xuất phát từ việc quan tâm đến một giao diện mới cho máy tính xách tay. Giao diện này sẽ cho phép người sử dụng điều khiển máy tính thông qua công việc đơn giản là nhìn vào các điểm quan tâm, và có lẽ là nháy mặt để thực hiện nó.

Các ứng dụng như đánh giá sự mệt mỏi của tài xế, giám sát, quảng cáo, nghiên cứu xã hội học, etc sẽ được hưởng lợi từ điều này. Ví dụ: một chiếc oto có thể đến nơi an toàn và dừng lại và họ có thể trải qua một giấc ngủ ngắn. Những nhân viên thực thi pháp luật có thể chú ý đến các đối tượng nhìn vào một vật trong một thời gian dài. Một công ty quảng cáo có thể nắm mục đích dựa trên những gì thu hút người nhìn. Một bản đồ về ngụ ý tương tác xã hội có thể hiện ra từ ý tưởng mà mọi người trong cùng một phòng nhìn vào nhau, nhưng không liên lạc thực.

Mấu chốt của một giao diện hướng nhìn mới lạ là, không đáng ngạc nhiên, định vị mắt và đánh giá ước lượng, Kỹ thuật đã giải thích về việc theo dõi hướng nhìn của một người sau khi hiệu chỉnh với một hình ảnh của đối tượng nhìn thằng về phía trước. Các đối tượng không phải là thành phần của hình ảnh thì được sử dụng để tạo module nhận diện mắt.

II) Phương pháp luận

Chúng tôi đã làm giả định sau về vấn đề:

+) Đối tượng thì trực tiếp quay camera và gần trung tâm bức ảnh.

+) CHúng tôi có một bức ảnh “hiệu chỉnh” của đối tượng đang nhìn thẳng về phía trước.

Chúng tôi hiển thi trọng tâm của mắt trong hình hiệu chuẩn như là “mắt chuẩn”, và các trọng tâm của mắt trong kiểm tra hình ảnh như là “kiểm tra mắt”. Cốt lõi của thuật toán đánh giá ánh nhìn là phát hiện mắt. Một khi vị trí điểm ảnh của việc kiểm tra mắt được phát hiện ra, nó đơn giản là việc tính toán các vector giữa mắt hiệu chỉnh và mắt kiểm tra, do đó xác định một vector tỉ lệ thuận với ////

A.NHận diện mắt:

Chiến lược tổng thể của chúng tôi là sử dụng việc phát hiện MAP dựa trên pixel màu để nhận ra mống mắt. Phương pháp màu đã được sử dụng thành công trong quá khứ để tìm mống mắt [1]. Chúng tôi làm một sửa đổi đối với MAP, tuy nhiên, trong đó chúng tôi bình thường hoá mọi pixel của vector RGB có chiều dài đơn vị, làm cho máy dò không thay đổi với điều kiện ánh sang

Chúng tôi đã xây dựng một bộ đào tạo dựa trên 50 bộ hình ảnh chân dung chuyên nghiệp bằng cách chỉ tay vào các khu vực mắt sử dụng MS Paint cho mỗi bức ảnh nhu

Sử dụng việc tập huấn này, chúng tôi đã thiết lập huấn luyện một máy dò Map với 16 thùng trong mỗi kích thước màu: thuật toán của chúng ta chỉ đơn giản đếm số điểm ảnh tích cực và tiêu cực mà rơi vào thùng, và tổng hợp các số liệu này trên tất cả các ảnh huấn luyện .

Kể từ khi mắt đại diện cho một tỷ lệ nhỏ của các điểm ảnh , điều này làm thiết dữ liệu sai lệch. Hơn nữa, kể từ khi mỗi hình ảnh đào tạo đại diện cho một màu mắt khác nhau, ngưỡng tích cực cần phát hiện là rất thấp. Chúng tôi đã sản xuát đường cong ROC ( xem hình 1) và được vẽ chính xác đối với ngưỡng để xác định một ngưỡng hợp lý để sử dụng.

Chúng ta lưu ý rằng trục của đường cong ROC không bằng nhau, và đường rải rác gần cuối đại diện cho đường của độ dốc đơn vị. Xem hình 2.

Chúng tôi đã tìm ra trong khi thử nghiệm bằng cách sử dụng các ngưỡng khác nhau ( cả trên và dưới ngưỡng cuối cùng được lựa chọn) rằng **false positives** thì khó giải quyết, phổ biến hơn, và hơn **false negatives**.. CHúng tôi đã chọn 1 ngưỡng tương ứng với một khoảng chính xác là 1/3. Việc tăng ngưỡng vượt xa ngưỡng của chúng tôi tương ứng với sự giảm độ chính xác, vì vậy đây là ngưỡng lớn nhất có vẻ khả thi.

**( FIG. 3)**

Ngay cả khi sử dụng ngưỡng này, chúng tôi cũng tìm thấy được Map quá nhiều **false positive.**

Chúng tôi đã tìm thấy, tuy nhiên hầu hết các **false positive**  thì liên quan đến tóc, quần áo, và các khu vực khác bên ngoài khuôn mặt, như thể hiện trong hình hình 3. Quan sát này đã thúc đẩy bước tiếp theo trong thuật toán của chúng tôi., trong đó chúng tôi đã loại trừ từ những xem xét MAP đã phát hiện các điểm bên ngoài khu vucej được phát hiện như khuôn mặt( xem “Nhận diện khuôn mặt” bên dưới). Kết quả của việc loại trừ các điểm bên ngoài khuôn mặt được thể hiện trong hình 4.

Bước cuối cùng là loại bỏ tất cả các điểm khác, những điểm mà không đại diện cho mắt. Chúng tôi giả sử các điểm ảnh được phát hiện tương ứng để mắt đại diện cho 2 khu vực kết nối lớn nhất. Chúng tôi liên tục loại bỏ những khu vực kết nối nhỏ hơn ngưỡng, để tăng gá trị của ngưỡng, cho đến khi chỉ còn 2 khu vực được kết nối lại.Hình 5 cho ta thấy một đồ thị của những con số về vùng được kết nối như là một thuật của ngưỡng về kích thước của vùng được kết nối. Hinh 6 trình bày về kết quả của việc xử lý ngưỡng này.

(HÌnh 5)

(HÌnh 6)

Cuối cùng thì chúng tính các **centroids** của 2 khu vực

Đây là những phát hiện

Tóm tắt:

1. Đào tạo MAP
2. Phát hiện MAP
3. Khu vực ngoài của mặt nạ
4. Xóa các vùng kết nối nhỏ đến khi chỉ còn 2 điểm
5. Tính các  **centroids** của vùng được kết nối này. Gồm 2 mắt.

B.) Nhận diện khuôn mặt

Chúng tôi giả định bổ sung là khuôn mặt được căn giữa trung bình trong hình ảnh, và chiếm phần lớn trung tâm hình vuông của lưới 3 × 3 phủ lên hình ảnh. Thuật toán của chúng tôi liên quan đến các phương pháp được trình bày trong [2], và như trong bài báo này, mục tiêu chính của chúng tôi là tốc độ vì chúng ta không đòi hỏi độ chính xác cao hoặc theo dõi đối tượng.

Chúng tôi quan sát thấy rằng gương mặt có màu sắc tương đối đồng nhất. Do đó gương mặt có thể được xem như là tập hợp các điểm ảnh được đóng trong không gian màu sắc đó là trung tâm hình vuông. Thuật toán để nhận diện khuôn mặt, như sau:

1) Cắt hình ảnh khoảng 1/3 ở trung tâm ở cả hai chiều.

2) Tính màu trung bình

3) Đánh dấu tất cả các điểm ảnh nằm trong một khoảng cách nhất định trong không gian màu sắc trung bình của hình ảnh.

4) Xóa hình ảnh với một điểm ảnh để loại bỏ điểm giả mạo.

5) Tính phần thân nồi của các điểm còn lại. Điều này bao quanh mặt. (chị google dịch vậy)

Một lợi thế của phương pháp này là nó bỏ qua các đặc điểm bên ngoài của vùng mặt như được xác định bởi các đặc trưng màu sắc. Có nghĩa là tóc và quần áo sẽ không xuất hiện trong khu vực được xem xét

C.) Ước lượng tầm nhìn

Do hình ảnh thử nghiệm có thể cho thấy khuôn mặt đã được di chuyển, quay hoặc được chỉnh tỷ lệ với hình ảnh hiệu chỉnh, chúng ta cần tính toán sơ đồ từ hình ảnh được dùng thử nghiệm so với hình hiệu chuẩn. Để làm việc này, chúng tôi chọn SIFT kết hợp với RANSAC và một mô hình homography. Vì một số webcam có khả năng xem góc rộng (và do đó làm méo mó) và khuôn mặt có thể dịch đi xung quanh đáng kể trong phạm vi tầm nhìn, chúng ta phải tính đến những mặt nghiêng của kết quả (dịch không ổn câu này) hoặc rõ ràng có thể là kết quả của biến dạng đó, do đó lựa chọn của chúng tôi là mô hình homography. Hơn nữa, như đã lưu ý trong [3], nhiều phần của khuôn mặt (ngoài mũi) có thể được coi là nằm trong mặt phẳng. Do đó một homography là một mô hình hợp lý để áp dụng cho mắt dưới sự xoay vòng mặt

**( FIG. 9)**

SIFT tính toán sự tương ứng giữa các điểm chính trên khuôn mặt, mà chúng tôi đưa lên xử lí để lấy ra một homography. Sử dụng homography này, chúng ta có thể ánh xạ vị trí mắt trong hình ảnh thử nghiệm đến các vị trí mắt tương ứng trong hình hiệu chuẩn.Nhận diện mắt trên hình ảnh hiệu chuẩn cho phép mắt hiệu chuẩn, và chúng ta có thể tính nhanh các vectơ giữa đôi mắt hiệu chuẩn và mắt thử nghiệm

**( FIG. 10)**

Tóm tắt:

1) Xác định cặp mắt thử nghiệm và cặp mắt chuẩn.

2) Sử dụng mô hình SIFT + RANSAC + homography để tính toán

một homography từ hình ảnh thử nghiệm tới hiệu chuẩn hình ảnh

3) Sử dụng hormography này để ánh xạ mắt thử nghiệm vào hình hiệu chuẩn.

4) Đối với mỗi mắt chuẩn, tính toán véc tơ đến mắt thử nghiệm gần nhất (từ chối nếu không có đủ gần mắt thử nghiệm). Điều này liên quan chặt chẽ đến góc nhìn. Nếu [ Δx Δy ]T là vector trung bình từ cặp mắt chuẩn đến mắt thử nghiệm, thì Δx ∞ sin θx và Δy ∞ sin θy

III. Kết quả

Chúng tôi đã đánh giá độ chính xác của việc ước lượng tầm nhìn bằng cách chụp ảnh một đối tượng được hướng dẫn để xem các điểm cụ thể trong buồng, cả ở các góc độ tương đối nhỏ và lớn từ đường giữa máy ảnh và khuôn mặt của đối tượng. Những điểm này được đo bằng máy ảnh - điều này cho phép so sánh giữa hướng nhìn thực tế và góc nhìn bằng thuật toán. Kết quả của kỹ thuật được thể hiện trong hình 11

**( FIG. 11)**

Các mô tả của MATLAB đã được sử dụng để xác định thời gian hao phí trong công đoạn của việc nhận dạng mắt và công việc ước tính tầm nhìn:

23,8% SIFT (để tính đồng bộ)

21.7% Nhận dạng mắt (trong hình ảnh thử nghiệm)

20.7% nhận dạng mắt (trong hình ảnh hiệu chuẩn)

Việc canh chừng nhà từ xa 16.5% (nghĩa là hình ảnh đã đọc)

Những thứ khác, tính toán góc nhìn, kết nối tiêu chuẩn

Để phát hiện chính xác cả hai mắt trong một hình ảnh 1944x1296 của một đối tượng, cần phải có kỹ thuật nhận dạng mắt được yêu cầu tầm khoảng 6 giây trên một lõi đơn của bộ xử lý i7. Đáng chú ý là việc tính toán ước lượng tầm nhìn chiếm một phần nhỏ trong các hướng dẫn sau khi mắt đã nhận dạng.

Chúng ta không thể kiểm tra thuật toán của chúng ta với một bộ dữ liệu lớn mà không có thiết lập trước vì chúng ta không có thời gian để biên soạn một tập hợp lớn bao gồm nhiều khuôn mặt, mỗi khuôn mặt có nhiều góc nhìn khác nhau,dựa vào điều kiện ánh sáng và chất lượng mà theo đó thuật toán của chúng ta được thiết kế.

Bởi vì việc nhận dạng mắt nằm ở cốt lõi của thuật toán, chúng ta tiến hành thử nghiệm để xác định độ chính xác của nó. Để làm điều này, chúng tôi chạy thuật toán nhận dạng mắt trên mỗi tấm hình trong số 50 hình ảnh và kiểm tra trực quan các kết quả, được lập bảng như sau:

***[vẽ bảng]***

Một vài cải tiến nhanh có thể được thực hiện với ít thời gian phát triển bổ sung tương đối

• Sử dụng homography từ hình ảnh hiệu chuẩn để kiểm tra hình ảnh để chỉ tìm kiếm đôi mắt trong vùng trên khuôn mặt. Điều này sẽ cải thiện tốc độ của thuật toán và có khả năng cải thiện đáng kể độ chính xác.

• Thử nghiệm với lớp mặt nạ trên khuôn mặt bị xói mòn bằng nhiều mức khác nhau để loại bỏ các lỗi sai hay phát sinh từ một vùng trên khuôn mặt bị phát hiện quá cỡ.

• Tính đường viền của khuôn mặt nên hình ảnh hiệu chuẩn không được yêu cầu.

Kỹ thuật được phát triển trong dự án cuối cùng này được thúc đẩy bởi các công cụ được giảng dạy trong EE368. Chúng tôi tin rằng những cải tiến sau đây liên quan đến các kỹ thuật không được đề cập đến trong lớp học, được đề xuất bởi những người tham dự Phiên Họp EE368 sẽ là những bước tiếp theo để cải tiến kỹ thuật của chúng ta:

• Các mẫu nhị phân cục bộ cho các bộ mô tả khuôn mặt [4].

• Định lượng lượng tử hóa bao gồm thông tin màu sắc trong kí hiệu mô tả keypoint [5].

Chúng tôi nhận thấy rằng ước lượng tầm nhìn theo chiều ngang so với chiều dọc sẽ có kết quả tốt hơn. Chúng tôi nghĩ rằng hình dạng theo chiều ngang của mắt góp phần cho hiệu ứng này.