Tạp chí Tin học và Điều khiển học, T.29, S.2 (2013), 173–185

XÁC ĐỊNH VỊ TRÍ MẮT NGƯỜI TRONG VIDEO BẰNG CÁCH KẾT HỢP DÒ TÌM VÀ THEO VẾT

CÁP PHẠM ĐÌNH THĂNG[1], DƯƠNG CHÍ NHÂN[2], NGÔ ĐỨC THÀNH[3], LÊ ĐÌNH DUY[1], DƯƠNG ANH ĐỨC[1]

[1] *Trường Đại học Công nghệ Thông tin, ĐHQG TP HCM*
[2] *Trường Đại Học Khoa Học Tự Nhiên, ĐHQG TPHCM*
[3] *The Graduate University for Advanced Studies (Sokendai), Japan*

Tóm tắt. Bài báo trình bày một phương pháp xác định vị trí mắt người dựa trên việc kết hợp một bộ dò tìm mắt người (eye detector) và một bộ theo vết mắt người (eye tracker). Phương pháp này giúp cải tiến kết quả xác định vị trí mắt người nhờ bộ dò tìm cung cấp những ước lượng tốt nhất cho các vị trí ứng viên của mắt người, trong khi đó bộ theo vết sẽ tìm ra vị trí tốt nhất trong các vị trí ứng viên đó bằng việc sử dụng thêm thông tin về thời gian. Thực nghiệm được tiến hành trên video từ cơ sở dữ liệu TRECVID 2009, cơ sở dữ liệu "Tư Thế Đầu Người" (HEAD POSE DATASET) của trường đại học Boston và video từ Đài truyền hình Việt Nam cho thấy kết quả của phương pháp kết hợp này đem lại hiệu quả cao hơn so với việc chỉ sử dụng bộ dò tìm hoặc theo vết đơn lẻ.

Từ khóa. Xác định vị trí mắt người, dò tìm mắt người, theo vết mắt người.

Abstract. In this paper, we propose a method to combine an eye tracker and an eye detector for robust eye localization in video. Instead of sequential intergration of the two systems, we use eye locations suggested by an eye detector for initialization and measurement of updating steps of particlesused inan eye tracker. This combination helps to improve the localization performance since the detector provides good estimation of eye location candidates, meanwhile the tracker helps to find the best eye location by using temporal information. Experiments were conducted on two benchmark video databases (TRECVID and Boston University Headpose datasets) and videos from Vietnamese Television. The results show that our method achieves a remarkable improvement compared to the state-of-the-art eye detector and eye tracker.

Key words. Eye localization, human eye detection, human eye tracking.

1. GIỚI THIỆU

Dò tìm đặc trưng của mặt người là nhiệm vụ chính yếu trong nhiều ứng dụng liên quan đến ảnh mặt người như: nhận dạng mặt người, xác định những biểu hiện cảm xúc trên mặt người, điều khiển tương tác giữa người và máy... Những đặc trưng của khuôn mặt người nổi bật là mắt, lông mày, mũi, miệng, cằm. Giữa những đặc trưng này, mắt người có vai trò rất quan trọng trong việc chuyển giao những tín hiệu tương tác, ý định hoặc chỉ dẫn của người dùng cho máy tính. Thông tin về vị trí mắt người trên mặt người ổn định nên việc xác định vị

trí mắt người là bước cần thiết trong nhiều phương pháp phân loại ảnh mặt người, căn chỉnh và chuẩn hoá ảnh mặt người. Chính vì vậy đã có rất nhiều những nghiên cứu chuyên sâu về nhận dạng mắt người trong ảnh mặt người hoặc video [1, 3, 4, 5]. Tuy nhiên việc dò tìm vị trí mắt người gặp phải nhiều khó khăn như sự thay đổi của tư thế đầu người, mắt nhắm hoặc mở, điều kiện ánh sáng thay đổi, bị che khuất một phần bởi tóc, đeo kính... nên việc xác định vị trí mắt người một cách chính xác vẫn đang là một thách thức. Hình 1cho thấy một số ví dụ về các trường hợp khó khăn gặp phải, dữ liệu được lấy từ cơ sở dữ liệu TRECVID 2009.

Hình 1. Ví dụ về các trường hợp khó khăn gặp phải. a) Biểu hiện mặt người thay đổi, b) tư thế đầu người thay đổi, c) nhắm mắt và che khuất, d) Người có đeo kính.

Hiện nay, việc xác định vị trí của mắt người trên video bằng cách áp dụng các kỹ thuật xử lý ảnh thông thường được tiến hành qua 2 bước chính: (1) xác định vị trí mặt người; (2) xác định vị trí mắt người trên ảnh mặt người. Tiến trình như sau: đầu tiên một bộ dò tìm mặt người được sử dụng để xác định vị trí mặt người tại khung hình đầu tiên, sau đó việc xác định vị trí mắt người dựa vào một bộ dò tìm hoặc một bộ theo vết mắt người. Đối với phương pháp dựa trên bộ dò tìm, ý tưởng chính là sử dụng bộ dò tìm trên mỗi khung hình của video. Bộ dò tìm mắt người hiện nay rất mạnh đối với ảnh mặt người nhìn thẳng và mắt người đang mở. Tuy vậy, phương pháp này bị hạn chế đối với các biểu hiện cảm xúc của mặt người thay đổi nhiều (tư thế đầu người, nhắm mắt, cười làm vùng mắt bị nhỏ lại). Mặt khác, phương pháp sử dụng bộ theo vết mắt người [12, 13] trong một số trường hợp có thể đáp ứng được những hạn chế của bộ dò tìm mắt người, nó có thể ước lượng được những vị trí của mắt người mà tại đó đang nhắm mắt hoặc bị ảnh hưởng bởi tư thế đầu người thay đổi. Tuy nhiên, độ chính xác của bộ theo vết thì phụ thuộc khá nhiều vào bước khởi tạo ban đầu. Hơn nữa, bộ theo vết thường không ổn định và dễ bị sai đối với những chuyển động quá nhanh tại một thời điểm nào đó dẫn đến các ước lượng ở những khung ảnh tiếp theo sẽ không chính xác.

Trong bài viết này, chúng tôi phát triển một phương pháp kết hợp các kỹ thuật tiên tiến trước đây để xác định vị trí mắt người trong video. Phương pháp này kết hợp kết quả của một bộ dò tìm mắt người và một bộ theo vết mắt người, bộ theo vết mắt người sử dụng mô hình "particle filter". Cụ thể là thông tin có được từ bộ theo vết giúp xác định vị trí của mắt người ngay cả trong những khung ảnh mà tại đó bộ dò tìm bị lỗi. Mặt khác thông tin của bộ dò tìm được tích hợp vào trong bộ theo vết tại mỗi thời điểm nên việc tích luỹ lỗi của bộ theo vết sẽ được giảm xuống theo thời gian. Đối với việc theo vết đối tượng trong video hoặc chuỗi ảnh, "particle filter" đã chứng tỏ được lợi thế của nó đối với các ước lượng không tuyến tính và không phải phân bố Gauss. Trong particle filter, thông tin quan trọng nhất của mỗi particle là trọng số của nó, nếu trọng số này được ước lượng càng chính xác thì độ chính xác của bộ theo vết lại càng được cải thiện. Chính vì vậy, phương pháp của chúng tôi tập trung vào việc tính toán và cập nhật lại trọng số này bằng cách kết hợp thông tin được cung cấp bởi cả bộ dò tìm và bộ theo vết mắt người. Thực nghiệm được tiến hành trên video từ cơ sở

dữ liệu video TRECVID 2009, "Tư Thế Đầu Người" của đại học Boston Hoa Kỳ và video từ đài truyền hình Việt Nam. Kết quả thực nghiệm cho thấy độ chính xác của phương pháp kết hợp này cao hơn phương pháp chỉ sử dụng bộ dò tìm hoặc bộ theo vết riêng lẻ.

2. CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, có một số lượng lớn các phương pháp xác định vị trí mắt người trong ảnh và video được công bố, các phương pháp này được chia làm 2 hướng riêng biệt như sau [3, 16]: (1) sử dụng các thiết bị đo xung điện để ghi lại thông tin xung điện ở các vùng da xung quanh hốc mắt hoặc sử dụng các thiết bị đặc biệt gắn trước ống kính máy quay phim[17, 18], và (2) sử dụng các kỹ thuật xử lý ảnh để xác định vị trí mắt người [1, 3, 4, 5].

2.1. Các phương pháp sử dụng thiết bị phụ trợ

Đối với hướng (1), các phương pháp dò tìm đều phải tốn chi phí cao cho các thiết bị đo xung điện. Các phương pháp như [17, 18] xác định vị trí mắt người rất mạnh và nhanh bằng cách dùng một số thiết bị phần cứng đeo trực tiếp vào mắt người, thiết bị này sẽ chiếu đèn hồng ngoại vào mắt người làm cho đồng tử sáng hơn và phân biệt với các vùng khác, từ đó xác định ra vị trí của mắt người. Tuy nhiên các phương pháp này còn gặp phải hạn chế như sau: video phải được quay ở khoảng cách rất gần với mắt người, hơn nữa độ chính xác còn phụ thuộc nhiều vào mắt nhắm, kích thước của mắt, và video chỉ thu trong phòng thí nghiệm.

2.2. Các phương pháp sử dụng kỹ thuật xử lý ảnh trong xác định vị trí mắt người

2.2.1. Dò tìm mắt người

Như đã trình bày ở phần giới thiệu, ý tưởng chính của phương pháp này là sử dụng bộ dò tìm trên mỗi khung hình của video. Các kỹ thuật dò tìm trên từng khung hình này dựa trên thông tin về hình học hoặc dựa trên đặc trưng. Dựa trên thông tin về hình học, phương pháp này xây dựng một mẫu hình học của mắt người và xác định vị trí mắt người trên các khung hình dựa trên việc so khớp mẫu thông qua một độ đo tương đồng. Yuille và các cộng sự [5] phân vùng trên khuôn mặt và tìm ra vùng nào giống mắt người nhất ước lượng vị trí của mắt. Một mở rộng phương pháp của Yuille, K. Lam và các cộng sự [19] ước lượng vị trí gần đúng của mắt người bằng việc tính trung bình và sử dụng các góc mắt đã được xác định để giảm số lần lặp trong việc tối ưu hoá mẫu. Cũng nằm trong hướng tiếp cận này, vào năm 2008, Valenti và các cộng sự [3] công bố một phương pháp xác định vị ví mắt người bằng cách kết hợp một bộ dò tìm mắt người và kỹ thuật "isophote voting" để xác định ra các mẫu vòng tròn của mắt người. Thực nghiệm cho thấy phương pháp này vẫn có kết quả tốt trong các trường hợp ánh sáng môi trường bị thay đổi hay tư thế đầu người thay đổi. Các phương pháp dựa trên thông tin hình học đều đạt độ chính xác cao với các ảnh mặt người nhìn thẳng, tuy nhiên ảnh đầu vào lại đòi hỏi phải có độ tương phản cao, mô hình hình học bước khởi tạo phải chính xác, và vẫn chưa đáp ứng được với các trường hợp chuyển động của đầu người thay đổi lớn.

Mặt khác, một số hướng tiếp cận dựa vào rút trích các đặc trưng liên quan đến mắt người thì thường trải qua 2 giai đoạn chính: (1) rút trích đặc trưng, (2) dùng các kỹ thuật phân lớp

xác suất để xác định vị trí của mắt người. Ở giai đoạn (1), một số phương pháp được công bố như dựa trên thông tin về đặc trưng cạnh [6], đặc trưng dạng sóng [10]. Và ở giai đoạn (2), các kỹ thuật như SVM [7, 9], Adaboost hoặc mạng nơ ron [2, 8, 9] được sử dụng. Dựa trên mô hình như vậy, [2, 9] trình bày phương pháp sử dụng mạng nơ ron đa lớp: bộ dò tìm vị trí mắt người được huấn luyện thông qua mạng nơ ron có thể xác định được vị trí của mắt người trong các trường hợp mắt xoay, co dãn và có thể hoạt động tốt với các điều kiện ánh sáng môi trường thay đổi. Tuy nhiên, các phương pháp này chỉ huấn luyện với ảnh mặt người nhìn thẳng.

2.2.2. Theo vết mắt người

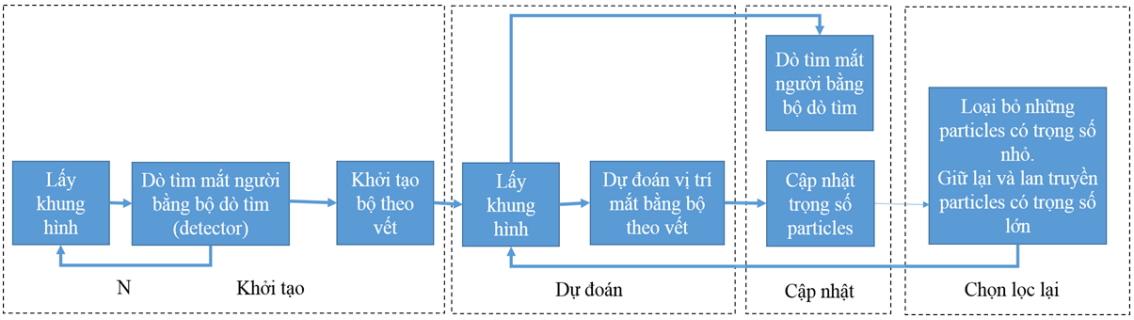
Với phương pháp sử dụng bộ theo vết mắt người, sử dụng một bộ theo vết để xác định vị trí của mắt người qua một chuỗi ảnh mặt người hoặc một video. Phương pháp này thông thường được tiến hành thông qua 2 bước: (1) bước khởi tạo xác định vị trí của mắt người ở khung hình đầu tiên, (2) bước tiếp theo dùng một bộ theo vết để xác định vị trí mắt người trong các khung hình tiếp theo. Wu Junwe và các cộng sự [13] xác định vị trí mắt người ở khung hình đầu tiên tự động bằng cách dựa trên một cây nhị phân xác suất, sau đó thông tin ban đầu sẽ được cung cấp cho bộ theo vết để xác định vị trí mắt người trong các khung hình tiếp theo. Phương pháp này cũng cho thấy ngoài việc xác định vị trí mắt người còn dò tìm được cái nháy mắt trong video. Các phương pháp ở hướng này cho thấy có thể đáp ứng được các trường hợp tỉ lệ hay kích thước của mắt người bị thay đổi trong video và với mặt người nhìn thẳng.

3. PHƯƠNG PHÁP TIẾP CẬN ĐỀ XUẤT

Như đã đề cập trong phần trước, ý tưởng chính và quan trọng nhất của mô hình đề xuất là nhằm khai thác được các thông tin của video như thông tin thời gian, thông tin chuyển động để có thể nâng cao hiệu quả cho toàn bộ hệ thống xác định vị trí mắt người. Khác với những cách tiếp cận trước đây, chúng tôi không chỉ đơn thuần sử dụng bộ dò tìm và bộ theo vết mắt người một cách riêng lẻ hoặc tuần tự mà là tích hợp 2 bộ này thành một hệ thống duy nhất. Nhờ vậy, hệ thống đề xuất sẽ có được những ưu điểm sau: (1) nhờ vào kết quả từ bộ dò tìm, bộ theo vết mắt người sẽ được khởi tạo một cách tự động; (2) nhờ vào vị trí mắt người ở những khung hình trước, bộ theo vết sẽ có thể xác định được vị trí của mắt người trong khung hình tiếp theo một cách liên tục ngay cả trong trường hợp bộ dò tìm không thể xác định được vị trí của mắt; (3) ứng với từng thời điểm nhất định, kết quả của bộ dò tìm sẽ được sử dụng để hiệu chỉnh kết quả dò tìm của bộ theo vết và nhờ vậy, đảm bảo được kết quả dò tìm chính xác hơn. Chính nhờ những cải tiến trên, kết quả dò tìm của toàn bộ hệ thống sẽ được nâng cao đáng kể. Hình 2 mô tả mô hình hệ thống kết hợp giữa bộ dò tìm mắt người và bộ theo vết mắt người.

3.1. Dò tìm mắt người sử dụng kỹ thuật isophote

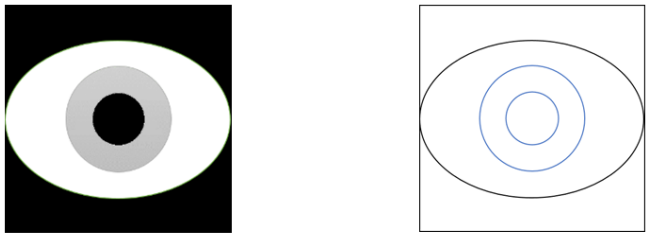
Một trong những phương pháp dò tìm vị trí mắt người đạt được hiệu quả cao nhất hiện nay có thể kể đến là phương pháp sử dụng "isophote voting" [3]. Dựa trên đặc điểm về dạng hình học của mắt là các đường cong và đối xứng, phương pháp này tìm cách xác định các



Hình 2. Mô hình hệ thống kết hợp bộ dò tìm và bộ theo vết

vùng có nhiều dạng đường cong đối xứng trong ảnh nhằm tìm được vùng biên của mắt. Sau đó, kĩ thuật bầu chọn (voting) có trọng số sẽ được áp dụng để tìm ra vị trí tâm mắt sao cho chính xác nhất.

Trong phương pháp này, tác giả sử dụng một khái niệm là "isophote". Đây chính là các đường cong kết nối các điểm ảnh có cùng độ sáng. Hình 3 minh họa cho các isophote của ảnh.



(a) **(b)**

Hình 3. Minh hoạ Isophote của ảnh mắt người. a) Ảnh nguyên bản; b) Các đường cong isophote tương ứng

Nhờ các isophote này độc lập với các phép xoay và thay đổi tuyến tính của ánh sáng, phương pháp của Valenti đã thể hiện được nhiều ưu điểm trong quá trình xác định vị trí của mắt. Các thực nghiệm cũng đã cho thấy phương pháp này đạt hiệu năng khá cao ngay cả với các điều kiện ánh sáng khác nhau hay tư thế đầu người thay đổi ít.

Với những ưu điểm trên, bộ dò tìm này sẽ được tích hợp vào hệ thống xác định vị trí mắt người. Và kết quả thực nghiệm sẽ chứng minh việc kết hợp với bộ theo vết, độ chính xác sẽ được nâng cao và có thể xác định tốt khi điều kiện ngoại cảnh thay đổi (hướng của khuôn mặt thay đổi nhiều, mắt nhắm,...)

3.2. Theo vết mắt người sử dụng particle filter framework

Bộ theo vết mắt người được sử dụng là mô hình theo vết sử dụng particle filter kết hợp với biểu đồ đặc trưng màu do Perez và các cộng sự [14] đề xuất. Mỗi đối tượng được theo vết (mắt người) sẽ được biểu diễn bởi N particle, mỗi particle tại thời điểm t được mô tả bởi vector $S_t = (x_t, y_t, s_t)$ với (x_t, y_t) là vị trí của particle (vị trí của mắt) và s_t là tỉ lệ (kích thước của mắt) của particle. Như vậy, trạng thái của mắt người được xác định như một dãy

các trạng thái của particle như sau: $X_t = \{(S_t^i, \pi_t^i)|i = 1..N\}$ và π là trọng số của particle, trong đó $\Sigma_(i = 1)^N \pi_t^i = 1$.

Ý tưởng chính của mô hình particle filter là xấp xỉ xác xuất hậu nghiệm $p(X_t, Z_t)$ bởi một tập mẫu được gán trọng số. Trong đó Z_t là trạng thái của mắt người được quan sát được tại thời điểm t. Và trọng số của mỗi particle được tính dựa trên khoảng cách Bhattacharyya giữa biểu đồ đặc trưng màu của particle tương ứng với biểu đồ đặc trưng màu mẫu được tính tại thời điểm khởi tạo.

3.3. Xác định vị trí mắt người bằng cách kết hợp bộ dò tìm và bộ theo vết mắt người

Để có thể duy trì quá trình theo vết với độ chính xác cao, bộ theo vết mắt người cần phải đảm bảo trọng số của các particle được tính toán hợp lý. Nếu các trọng số của particle không được tính toán hợp lý sẽ làm cho hệ thống dễ dàng mất dấu của đối tượng và rất khó có thể phục hồi trong các khung hình tiếp theo. Do đó, trọng số của particle là một trong những yếu tố quan trọng nhất quyết định hiệu năng của bộ theo vết.

Trong mô hình đề xuất, ngoài việc sử dụng bộ dò tìm để khởi tạo tự động cho bộ theo vết mắt người, kết quả của bộ dò tìm còn được tích hợp vào bước tính toán và cập nhật trọng số cho các particle trong các giai đoạn theo vết nhằm đảm bảo từng bước trong hệ thống được tính toán một cách hợp lý nhất và nhờ vậy nâng cao đáng kể độ chính xác của toàn bộ hệ thống.Mô hình đề xuất gồm có 4 bước xử lý chính như sau:

(a) Khởi tạo. Nhằm tránh việc khởi tạo thủ công cho bộ theo vết, ngay tại thời điểm đầu tiên bộ dò tìm xác định vị trí của mắt người, các tham số cho bộ theo vết sẽ được khởi tạo. Các tham số này gồm có:

- Biểu đồ đặc trưng màu tại tọa độ mắt trái và mắt phải (x_1, y_1), (x_2, y_2) có được từ bộ dò tìm. Biểu đồ màu này sẽ được dùng làm biểu đồ tham chiếu cho vùng mắt người ở các khung hình tiếp theo.

- N: số lượng particles xung quanh 2 điểm (x_1, y_1), (x_2, y_2). Trong thực nghiệm, chúng tôi tạo ra 300 particles ngẫu nhiên xung quanh vị trí mắt người để dự đoán vị trí các điểm tiếp theo. Do vùng mắt người trên ảnh là khá nhỏ, nên số lượng particles quá lớn sẽ dẫn đến nhiều particles được sinh ra sẽ ra cách xa vị trí thực của mắt. Ngược lại nếu số lượng particles quá nhỏ sẽ không phủ hết được các vị trí cần thiết dùng cho dự đoán.

(b) Dự đoán. Một mô hình chuyển động được áp dụng để ước lượng vị trí mới của n particle trong khung hình tiếp theo. Ta định nghĩa vector trạng thái như sau: $S_t = (x_t, y_t, s_t)$ trong đó (x_t, y_t) là toạ độ vị trí mắt người và s_t là kích thước của mắt người tại thời điểm t. Cho lan truyền tập mẫu này và ước lượng sự chuyển động của mắt người thông qua một mô hình hồi qui động cấp 2 (a second order autoregressive dynamic model). Công thức chuyển động cụ thể như sau

$$x_{t+1} = Ax_t + Bx_{t-1} + Cv_t,\ v_t \sim N(0, \sigma_t)$$

$$y_{t+1} = Ay_t + By_{t-1} + Cv_t,\ v_t \sim N(0, \sigma_t)$$

$$s_{t+1} = As_t + Bs_{t-1} + Cv_t,\ v_t \sim N(0, \sigma_t)$$

trong đó ma trận A, B đại diện cho thành phần bất biến và ma trận C đại diện cho thành phần ngẫu nhiên và $\sigma_t$là phương sai của phân phối chuẩn Gauss. Như vậy v_t sẽ được tỉ lệ với một phân phối Gaussian như sau

$$v_t = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-z_t^2/2\sigma_t^2} \tag{4}$$

trong đó z_t là giá trị phát sinh ngẫu nhiên trong $[-1, 1]$.

Trong thực nghiệm các giá trị $a[i,j],\ b[i,j],\ c[i,j]$ lần lượt là 2, -1, 1 để mô phỏng phương trình chuyển động không đều.

(c) Cập nhật trọng số. Để tính toán trọng số w_t^i cho mỗi particle của bộ theo vết, chúng tôi ước lượng một xác suất điều kiện được mô tả như sau

$$w_t^i \propto p(y_t|x_t^i) \tag{5}$$

và được chuẩn hoá vào đoạn $[0, 1]$

$$w_t^i = \frac{w_t^i}{\Sigma_{i=1}^N w_t^i}. \tag{6}$$

Đây là bước quan trọng nhất trong việc cải tiến mô hình kết hợp đề ra, trong Mục 3.4 sẽ trình bày chi tiết về cách tính trọng số và cách kết hợp trọng số của bộ dò tìm và bộ theo vết.

(d) Chọn lọc mẫu. Những particle nào có trọng số thấp sẽ bị loại bỏ và những particle có trọng số cao sẽ được giữ lại, tạo ra bộ mẫu tốt cho xác định vị trí mắt.

Các bước 2, 3, 4 tiếp tục lặp lại cho đến khung hình cuối cùng của video.

3.4. Cập nhật trọng số tích hợp với thông tin từ bộ dò tìm mắt người

Để gán trọng số cho tập mẫu chúng tôi tính hệ số Bhattacharyya giữa biểu đồ tham chiếu (reference histogram) và biểu đồ mục tiêu (target histogram). Sau đó sử dụng một phân phối xác suất Gauss cho hệ số Bhattacharyya này. Khoảng cách Bhattacharyya ngắn nhất thể hiện trọng số cao nhất. Cụ thể như sau

$$\pi_{iB} = \frac{1}{\sqrt{2\pi\sigma_B^2}} e^{-(d_B^2)/(2\sigma_B^2)} \tag{7}$$

trong đó d_B là khoảng cách Bhattacharyya giữa biểu đồ tham chiếu và biểu đồ của particle thứ i, π_{iB} là trọng số chưa chuẩn hoá và σ_B^2 là phương sai của phân phối chuẩn Gauss, trong thực nghiệm chọn bằng 10 để cho kết quả tối ưu.

Tiếp theo, thông tin của bộ dò tìm mắt người được sử dụng ngay thời điểm này để kết hợp nó với thành phần Gauss Bhattacharyya. Trong trường hợp này ta tính phân phối chuẩn Gauss cho khoảng cách Euclidean của mỗi vị trí particle với vị trí cung cấp bởi bộ dò tìm mắt người, cụ thể như sau

$$\pi_{iE} = \frac{1}{\sqrt{2\pi\sigma_E^2}} e^{-(d_E^2)/(2\sigma_E^2)} \tag{8}$$

trong đó d_E là khoảng cách Euclidean giữa vị trí của mỗi particle và vị trí được cung cấp bởi bộ dò tìm mất người, π_{iE} là trọng số thu được từ tính toán của bộ dò tìm và σ_E^2 là phương sai của phân phối chuẩn Gauss, trong thực nghiệm ta sử dụng phân phối chuẩn hoá $\sigma_E^2 = 1$. Sau đó, các trọng số này cũng được chuẩn hoá vào đoạn [0,1]

$$w_i = \frac{\pi_i}{\Sigma_{i=1}^{N} \pi_i} \tag{9}$$

trong đó π_i là trọng số thu được từ bộ dò tìm hoặc bộ theo vết.

Và cuối cùng thông tin của bộ dò tìm và bộ theo vết được kết hợp như sau

$$w_i = \propto * w_{iB} + (1 - \propto) * w_{iE}$$

trong đó $\propto$ là trọng số kết hợp của bộ theo vết và bộ dò tìm. $\propto$ được gán gần bằng 1 cho video có nhiều chuyển động của khuôn mặt, ngược lại $\propto$ gán gần 0 cho trường hợp sử dụng độ chính xác của bộ dò tìm và ít chuyển động của mặt người. Trong bài báo này, việc sử dụng kết hợp trực tiếp hai trọng số mang tính tổng quát và đảm bảo cho chi phí tính toán thấp do đó có thể đáp ứng được cho các hệ thống thực thi thời gian thực. Hơn nữa, kết quả thực nghiệm ở Mục 4.2 chứng minh được hiệu quả của mô hình đề xuất so với các phương pháp nhận dạng và theo vết riêng lẻ.

4. THỰC NGHIỆM

4.1. Cơ sở dữ liệu và đánh giá

Tiến hành thực nghiệm trên một video được chọn ngẫu nhiên từ cơ sở dữ liệu TRECVID 2009. Các khuôn mặt người trong video xuất hiện với các hướng nhìn khác nhau, kích cỡ khác nhau, nhiều thay đổi biểu hiện trên khuôn mặt khác nhau và với môi trường hậu cảnh khác nhau.

Cơ sở dữ liệu của đại học Boston bao gồm 45 video của 5 người thực hiện 9 động tác thay đổi tư thế đầu người trong điều kiện ánh sáng phòng thí nghiệm. Cơ sở dữ liệu này có mặt người và đầu người luôn xuất hiện trừ những trường hợp đặc biệt do cá nhân đối tượng tự mình làm che khuất một phần.

Ngoài ra, còn thực nghiệm trên video thu từ đài truyền hình Việt Nam. Mỗi video gồm khoảng 1000 khung hình chứa ảnh mặt người với các động tác và tư thế của đầu người khác nhau.

Để đánh giá độ chính xác chúng tôi sử dụng độ đo cho vị trí mắt người được công bố bởi Jesorsky và các cộng sự [15], độ đo này đã được sử dụng trong nhiều công trình đã công bố [3, 11]. Công thức đo lường tỉ lệ lỗi chuẩn hoá (normalize error) như sau

$$e = \frac{\max(d_{left}, d_{right})}{d} \tag{11}$$

trong đó d_{left} và d_{right} là khoảng cách Euclidean giữa vị trí mắt người được xác định bởi mô hình và vị trí của chính mắt người đó trong bảng xác thực dữ liệu (ground truth), d là khoảng cách giữa mắt trái và mắt phải trong bản xác thực cơ sở dữ liệu.

Trong phần kết hợp giữa bộ dò tìm và bộ theo vết, tham số $\propto$ là trọng số kết hợp. Do vậy trước khi đánh giá kết quả của toàn bộ hệ thống thì ta thực nghiệm đánh giá độ ảnh hưởng của trọng số kết hợp này khi thay đổi giá trị của nó. Hình 4c cho thấy kết quả của hệ thống khi trọng số $\propto$ thay đổi với cơ sở dữ liệu của đại học Boston. Kết quả này cho thấy trong tất cả các trường hợp độ chính xác của hệ thống gần như là tương đương nhau. Do đó trong thực nghiệm, chúng tôi chọn $\propto$=0.5 cho cân bằng giữa bộ dò tìm và bộ theo vết. Bảng 1 thể hiện kết quả của hệ thống của chúng tôi trên cơ sở dữ liệu TRECVID 2009, Boston và VTV. Thông số kích thước của vùng mắt được tính dựa vào nhân trắc học và độ chính xác được đánh giá với tỉ lệ lỗi chuẩn hoá trong khoảng từ [0, 0.5].

Bảng 1. Thực nghiệm trên cơ sở dữ liệu TRECVID2009 và BOSTON

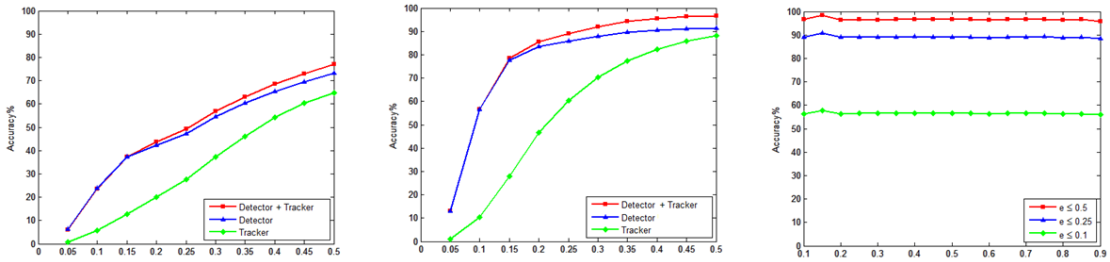
Phương pháp	**Độ chính xác với tỉ lệ lỗi e**								
	TRECVID2009			**BOSTON**			**VTV**		
	e<0.1	**e<0.25**	**e<0.5**	**e<0.1**	**e<0.25**	**e<0.5**	**e<0.1**	**e<0.25**	**e<0.5**
Bộ theo vết (Tracker)	5.82	27.65	64.77	10.45	60.48	88.12	72.00	93.08	94.08
Bộ dò tìm (Detector)	23.90	47.22	73.10	56.65	85.79	91.32	71.99	93.44	95.70
Kết hợp dò tìm và theo vết (Detector+tracker)	**23.66**	**49.17**	**76.92**	**56.50**	**89.06**	**96.66**	**76.28**	**97.73**	**98.33**

Hình 4a thể hiện biểu đồ độ chính xác của hệ thống kết hợp đề xuất và độ chính xác của bộ dò tìm và bộ theo vết được sử dụng riêng lẻ với cơ sở dữ liệu TRECVID. Từ kết quả cho thấy rằng, khi bộ theo vết (tracker) bị lỗi hay hội tụ cục bộ thì trọng số của các particle gần vị trí mà bộ dò tìm (detector) trả về sẽ lớn hơn các particle ở xa, do đó giảm thiểu được lỗi của bộ theo vết và tăng độ chính xác của bộ theo vết. Chính vì vậy khi so sánh giữa hệ thống kết hợp đề xuất (detector + tracker) và bộ theo vết (tracker) sử dụng riêng lẻ, độ chính xác ở $e < 0.1$ tăng 17% và ở $e < 0.5$ tăng 12%. Ngược lại, nếu bộ dò tìm (detector) bị lỗi thì bộ theo vết vẫn có thể ước lượng được vị trí của mắt người trong những khung hình tiếp theo. Kết quả so sánh độ chính xác của hệ thống đề xuất và bộ dò tìm riêng lẻ (detector) ở $e < 0.25$ tăng 2% và ở $e < 0.5$ tăng 3%.

Đối với cơ sở dữ liệu của đại học Boston, các thông số thử nghiệm đánh giá giống với các thông số sử dụng đánh giá trên cơ sở dữ liệu Trecvid. Kết quả thử nghiệm đánh giá trình bày ở Bảng 1 và biểu đồ 4b, kết quả cũng cho thấy rằng khi sử dụng kết hợp cả bộ dò tìm và bộ theo vết thì độ chính xác của cả hệ thống tăng lên đáng kể so với việc chỉ sử dụng riêng lẻ.

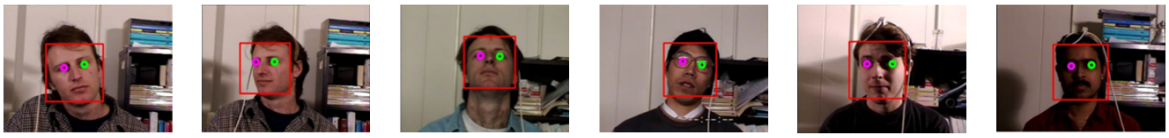
Ngoài ra, với kết quả thực nghiệm trên video thu được từ VTV đài truyền hình Việt Nam cũng cho thấy được kết quả ổn định của hệ thống kết hợp đề xuất.

Hình 5 thể hiện kết quả của hệ thống kết hợp trên cơ sở dữ liệu của đại học Boston với các trường hợp khác nhau như mặt người có đeo kính, ánh sáng môi trường thay đổi và các tư thế của đầu người thay đổi (ngước lên, ngước xuống, quay trái, phải).

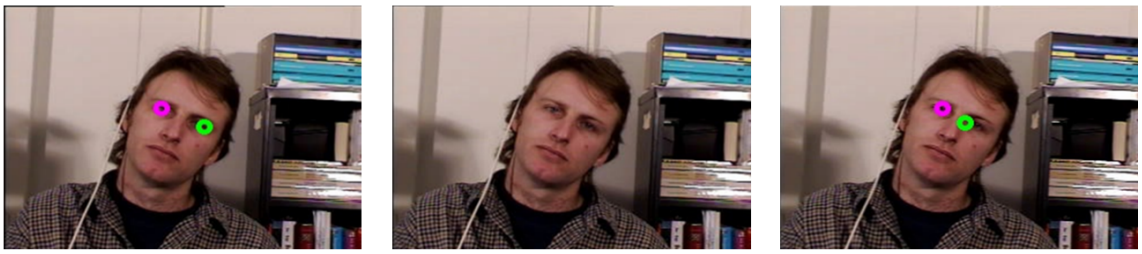


(a) (b) (c)

Hình 4. (a) Kết quả thực nghiệm trên cơ sở dữ liệu TRECVID 2009; (b) Kết quả thực nghiệm trên cơ sở dữ liệu của Đại Học Boston; (c) Kết quả về ảnh hưởng của trọng số kết hợp khi thay đổi giá trị



Hình 5. Kết quả thực nghiệm trên cơ sở dữ liệu của đại học Boston với các trường hợp khác nhau

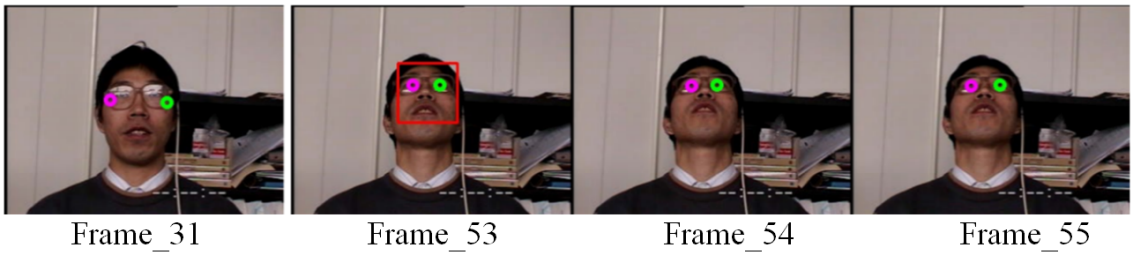


(a) (b) (c)

Hình 6. Kết quả so sánh: a) Hệ thống kết hợp, b) bộ dò tìm, c) bộ theo vết

Thực nghiệm ở Hình 6 so sánh kết quả của hệ thống kết hợp bộ theo vết và bộ dò tìm so với kết quả của bộ dò tìm và bộ theo vết riêng lẻ. Khung hình thứ 155 được trích từ video của cơ sở dữ liệu Boston. Trong trường hợp này, đầu người trong tư thế nghiêng sang phải và bộ dò tìm bị lỗi do không dò tìm được mặt người trong khung hình (Hình 6b), trong khi đó bộ theo vết và hệ thống kết hợp vẫn xác định được vị trí của mắt người. Mặt khác khi so sánh giữa bộ theo vết và hệ thống kết hợp thì kết quả thu được từ hệ thống kết hợp chính xác hơn do hệ thống kết hợp đã giảm được tỉ lệ lỗi từ các bước trước đó nhờ kết quả của bộ dò tìm. Một trường hợp khác, ở thí nghiệm như Hình 7: tại khung hình 31 hệ thống bị lỗi nhưng nhờ bộ dò tìm (khung hình 53) giúp cho bộ theo vết giảm lỗi tích luỹ và cho kết quả chính xác hơn (khung hình 54, 55).

Thực nghiệm tích hợp vị trí mắt người vào hệ thống chuẩn hoá và nhận dạng ảnh mặt người Chuẩn hoá ảnh mặt người là một giai đoạn đầu tiên và quan trọng trong hệ thống nhận dạng mặt người. Việc xác định tự động vị trí mắt sẽ giúp cho hệ thống nhận dạng mặt người không chỉ giảm thiểu chi phí và thời gian mà còn nâng cao độ chính xác cho hệ thống. Bước



Hình 7. Bộ dò tìm giúp cho bộ theo vết giảm thiểu lỗi tích luỹ

đầu tiên, bộ dò tìm mặt người sẽ xác định vùng chứa mặt người. Sau đó khoản cách giữa hai vị trí trung tâm của mắt trái và phải sẽ là đầu vào quan trọng cho công đoạn chuẩn hoá ảnh mặt người.

Trong phần thực nghiệm này sẽ tích hợp kết quả xác định vị trí của mắt người vào hệ thống chuẩn hoá và nhận dạng mặt người của trường đại học Colorado State của Hoa Kỳ. Trong đó, các trường hợp ảnh mặt người với tư thế đầu nhìn nghiêng và xoay được chuẩn hoá lại. Hình 8 là ví dụ về các trường hợp sử dụng vị trí mắt người để chuẩn hoá ảnh mặt người dạng nhìn nghiêng và trong tư thế xoay. Và kết quả là các ảnh mặt người sẽ được chuẩn hoá thành dạng nhìn thẳng dựa vào khoảng cách hai mắt.

Hình 8. Kết quả sử dụng vị trí mắt người trong chuẩn hoá ảnh mặt người

Đối với việc nhận dạng mặt người, cũng tiến hành thực nghiệm trên cơ sở dữ liệu Boston. Số ảnh mặt người của cơ sở dữ liệu Boston là 8955 ảnh của 5 người khác nhau. Thực nghiệm trên 2 trường hợp: có sử dụng vị trí mắt người và không sử dụng. Khi không sử dụng vị trí mắt người thu được 8158 ảnh mặt người cho kết quả đúng. Và khi sử dụng vị trí mắt thì kết quả là 8591 ảnh đúng. Đối với cơ sở dữ liệu Boston nhờ kết quả của vị trí mắt được tích hợp nên đã giúp cho hệ thống nhận dạng được những khuôn mặt trong tư thế nhìn nghiêng và xoay (những trường hợp này bộ dò tìm mặt người bị lỗi) nhờ đó nâng cao độ chính xác của toàn hệ thống. Hơn nữa, việc xác định vị trí mắt tự động giúp cho giảm thiểu nhiều thời gian và chi phí cho việc gán nhãn mắt cho công đoạn chuẩn hoá ảnh mặt người.

5. KẾT LUẬN

Bài báo đã đề xuất một phương pháp xác định vị trí mắt người dựa trên việc kết hợp một bộ theo vết mắt người sử dụng particle filter với một bộ dò tìm mắt người. Qua đó cho thấy những lợi thế của việc sử dụng các thông tin về thời gian và chuyển động của video. Cả bộ dò tìm và bộ theo vết sử dụng particle filter đều cho thấy hiệu quả của nó trong việc kết hợp để tăng độ chính xác của cả hệ thống. Thực nghiệm cho thấy được ưu thế của việc kết hợp này đối với các trường hợp tư thế đầu người thay đổi, mắt nhắm hoặc mắt người bị che khuất một phần. Kết quả thực nghiệm cũng cho thấy độ chính xác của hệ thống kết hợp tăng từ $3\% \pm 5\%$ so với bộ dò tìm riêng lẻ và tăng từ $12\% \pm 17\%$ so với bộ theo vết riêng lẻ trên cả cơ sở dữ liệu TRECVID 2009 và đại học Boston. Ngoài ra thực nghiệm còn cho thấy kết quả của vị trí mắt còn đạt được nhiều lợi thế khi được tích hợp vào trong hệ thống chuẩn hoá và nhận dạng ảnh mặt người.

TÀI LIỆU THAM KHẢO

[1] P. Campadelli, R. Lanzarotti, and G. Lipori, Eye localization: a survey, *The Fundamentals of Verbal and Non-verbal Communication and the Biometrical Issue*, NATO Science Series, 2007.

[2] W. Peng, MB. Green, J. Qiang, J. Wayman, Automatic eye detection and its validation, *Proc. 2005 IEEE CS Conf. Computer Vision and Pattern Recognition* **3** (2005) 164–164.

[3] R. Valenti and T. Gevers, Accurate eye center location and tracking using isophote curvature, *Proc. 2008 IEEE CS Conf. Computer Vision and Pattern Recognition* **0** (2008) 1–8.

[4] L. Bai, L. Shen, and Y. Wang, A novel eye location algorithm based on radial symmetry transform, *International Conf. on Pattern Recognition* **3** (2006) 511—514.

[5] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, Feature extraction from faces using deformable templates, *International Journal of Computer Vision* **8** (2) (1992) 99–111.

[6] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, An eye detec-tion algorithm using pixel to edge information, *IEEE CS Conf. on Control, Communications, and Signal Processing* **2** (2006) 1–4.

[7] P. Campadelli, R. Lanzarotti, and G. Lipori, Precise eye and mouth lo-calization, *International Journal of Pattern Recognition and Artificial Intelligencece* **23** (3) (2009) 359–379.

[8] C. Garcia and M. Delakis, Convolutional face finder: A neural architecture for fast and robust face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** (11) (2004) 1408–1423.

[9] X. Tang, Z. Ou, T. Su, H. Sun, and P. Zhao, Robust precise eye location by adaBoost and SVM techniques, *Advances in Neural Networks* **3497** (2) (2005) 93—98.

[10] J. Huang and H. Wechsler, Eye detection using optimal wavelet packets and radial basis functions, *Journal of Pattern Recognition and Artificial Intelligence* **13** (7) (1999) 1009–1026.

[11] F. Yang, J. Huang, P. Yang, D. Metaxas, Eye localization through multiscale sparse dictionaries, *IEEE Conf. on Automatic Face and Gesture Recognition* **9** (0) (2011) 514–518.

[12] K. Grauman, M. Betke, J. Gips and G. R. Bradski, Communication via eye blinks detection and duration analysis in real time, *Proc. 2008 IEEE CS Conf.Computer Vision and Pattern Recognition* **1** (0) (2001) 1010–1017.

[13] Wu Junwen and Trivedi Mohan M., An eye localization, tracking and blink pattern recognition system: Algorithm and evaluation, *ACM Transactions on Multimedia Computing, Communications, and Applications* **6** (2) (2010).

[14] Prez Patrick, Hue Carine, Vermaak Jaco, and Gangnet Michel, Color-based probabilistic tracking, *European Conference on Computer Vision* **1** (3) (2002) 661–675.

[15] J. Oliver, K. J. Kirchberg, and F. Robert, Robust face detection using the hausdorff distance, *International Conf. on Audio- and Video-Based Biometric Person Authentication* **2091** (2) (2001) 90–95.

[16] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice,* Springer, 2007.

[17] Z. Zhu and Q. Ji, Robust real-time eye detection and tracking under variable lighting conditions and various face orientations, *Journal of Computer Vision and Image Understanding* **98** (1) (2005) 124–154.

[18] c. Morimoto, D. Koons, A. Amir, and Flickner, Pupil detection and tracking using multiple light sources, *Journal of Image and Vision Computing* **18** (4) (2000) 331–335.

[19] K. Lam and H. Yan, Locating and extracting the eye in human face images, *Journal of Pattern Recognition* **29** (5) (1996) 771—779.

Ngày nhận bài 10 - 1 - 2013

Nhận lại sau sửa ngày 04 - 6 - 2013