

Face Processing in Video

Tác giả thứ nhất^{1,3}

Tác giả thứ 2²

Tác giả thứ 3^{3,1} and Tác giả thứ 4^{3,1}

¹ Trường ĐH.....

² University of Science
HCMC, Vietnam

³ National Institute of Informatics

What ?

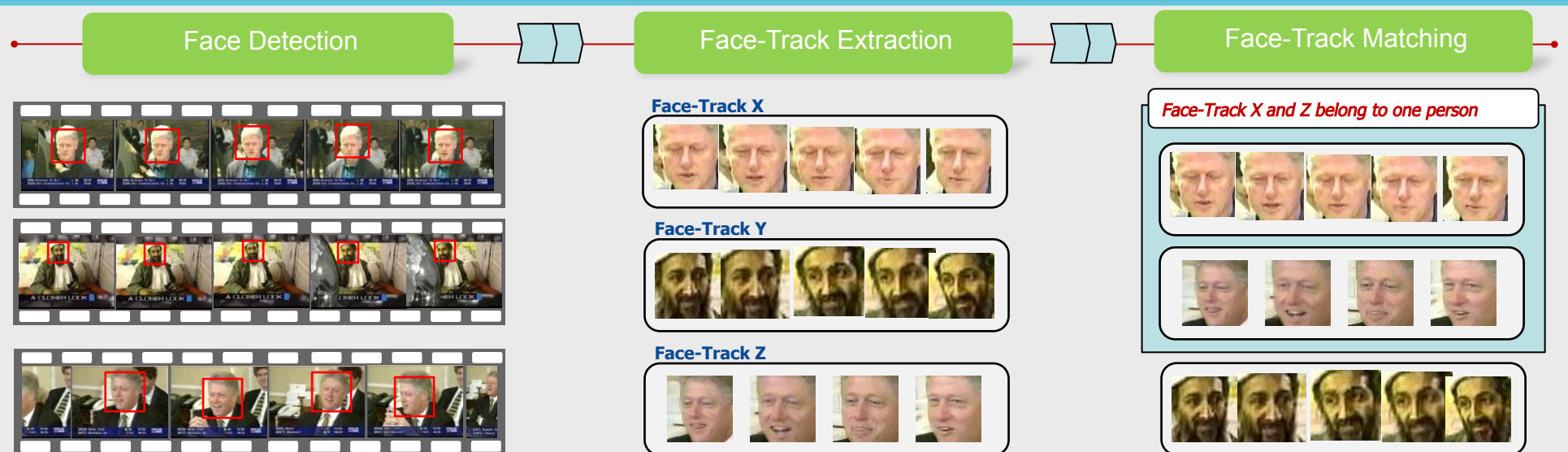
We introduce a framework to process and retrieve faces in video, in which we have:

- Proposed a robust method to extract face-tracks in news video.
- Built the largest face database compare to current popular worldwide face databases.
- Evaluated several face-track matching methods.

Why ?

- The human face is one of the most important objects in video since it provides rich information for spotting people of interest and is the basis for interpreting facts. Therefore, detecting and recognizing faces appearing in video are essential tasks of video indexing and retrieval applications.
- Most studies have focused on static images rather than **large-scale** and **real video dataset**.

Overview



Description

1. Face Detection

- The face detector implemented in OpenCV based on Viola method was used for detecting frontal views of faces in every frame of our video sequences.
- A high threshold was used to reduce the number of false positives.

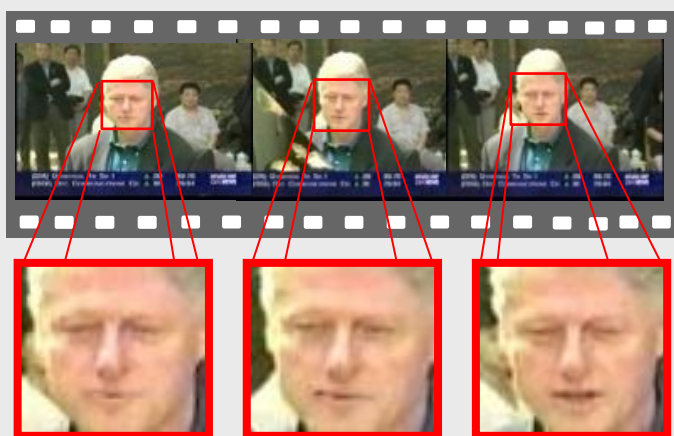


Figure 1. Detection results with a high threshold.

3. Face-Track Matching

- Face-track matching is done by applying similarity estimation methods (e.g. Min-Min distance) on the feature space.

2. Face-Track Extraction

- We used Kanade–Lucas–Tomasi (KLT) method to create and track key/interest points between frames.
- The number of key points that pass through pairs of faces in consecutive frames was computed to make decision on grouping faces into face-tracks.
- Several treatments are proposed to handle tracking traps in news video.
 - Flash-frame detector.
 - Adaptively generating key points.

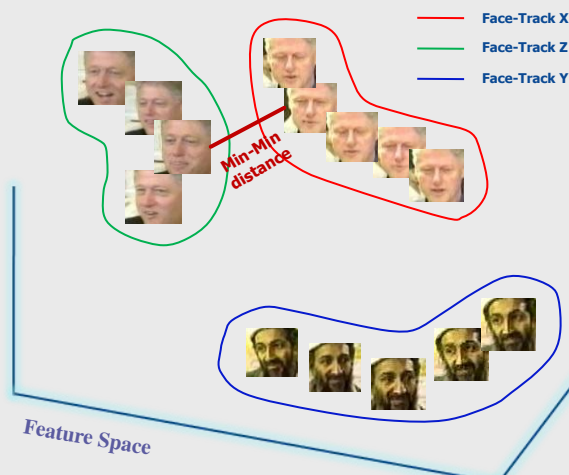


Figure 4. Apply Min-Min method for face-track matching.

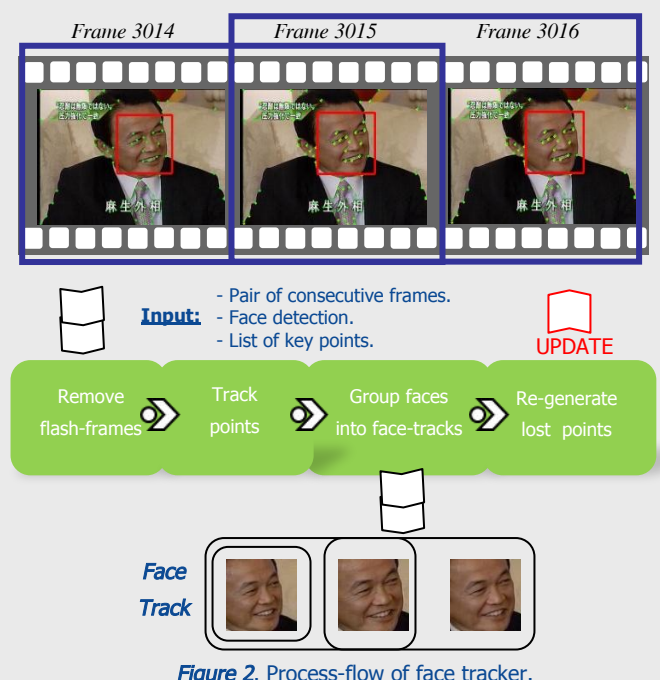


Figure 2. Process-flow of face tracker.

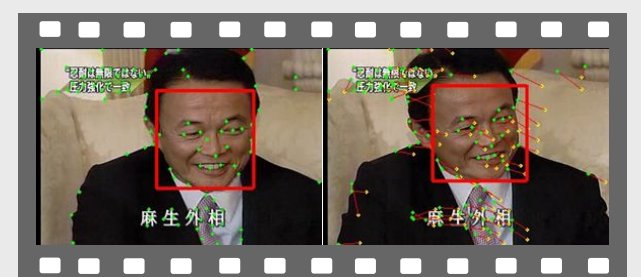


Figure 3. Key/Interest points (plotted as green dots) in the left frame are tracked in the right frame. Small lines from green dots are motion of these points. Two faces in these frames share 22/23 points.

3D-MOMENTS FROM NEAR-DUPLICATE PHOTOS

Trần Thành ^{1,2}

Nguyễn Duy Phúc ^{1,2}

Hoàng Quang Vũ ^{1,2}

¹ Trường Đại học Công nghệ Thông tin
TPHCM

² Đại học Quốc gia
Thành phố Hồ Chí Minh

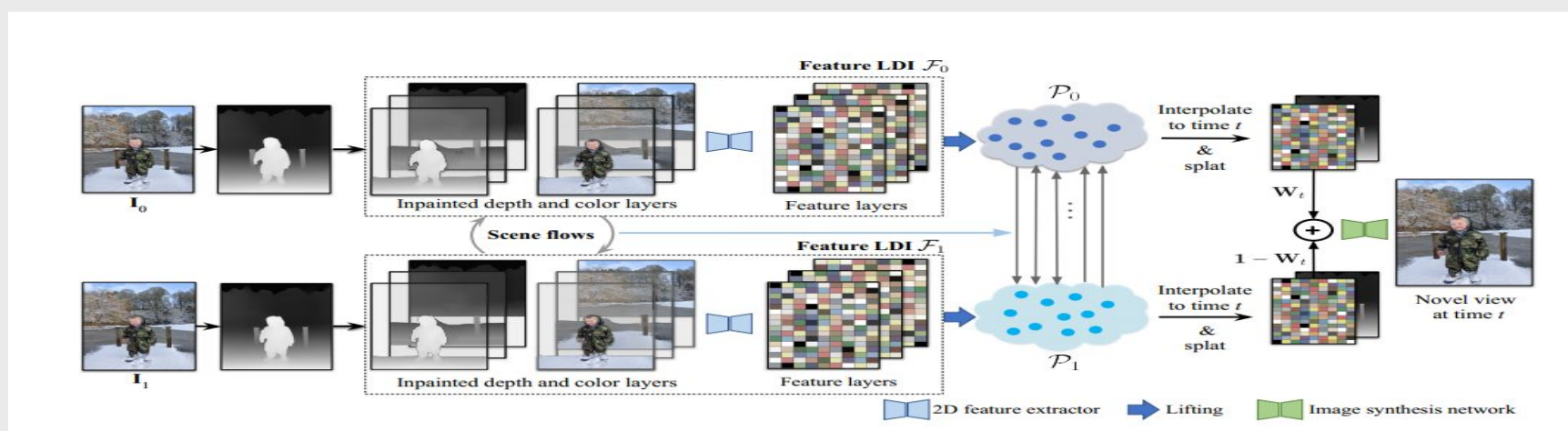
What ?

- Tạo ra 3D Moments từ các bức ảnh gần giống nhau của các cảnh động.
- Tạo ra một biểu diễn mới dựa trên các LDI đặc trưng được tăng cường với các scene flow.
- Một mô hình có thể được huấn luyện để tạo ra 3D Moments

Why ?

- Thay vì chọn ra tấm ảnh đẹp nhất trong loạt ảnh chụp liên tục thì có được một video 3D nhằm lưu giữ khoảnh khắc.

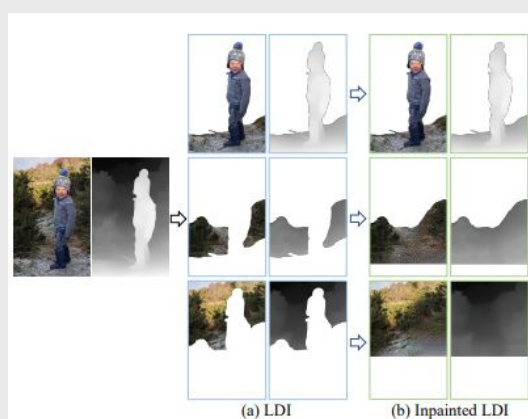
Overview



Description

1. Nội dung

- Cho cặp hình ảnh gần trùng nhau I_0, I_1 , chúng tôi căn chỉnh chúng bằng một homography và dự đoán dense depth map của mỗi tấm hình. Đổi từ hệ màu RGBD sang hệ LDI, với những vùng bị khuyết được lấp đầy bởi depth-aware inpainting. Áp dụng bộ trích xuất đặc trưng 2D cho mỗi lớp màu của ảnh LDI inpainted để thu được các lớp đặc trưng, kết quả thu được các đặc trưng LDI (F_0, F_1), nơi mà các màu trong vùng inpainted LDI đã được thay thế bởi các đặc trưng.
- Để mô hình hóa được cảnh chuyển động, chúng tôi tính toán scene flow của mỗi điểm ảnh trong ảnh LDI sử dụng độ sâu đã dự đoán trước đó và dòng quang học giữa 2 tấm ảnh input.
- Để render ra một novel view ở thời điểm t , chúng tôi đẩy các đặc trưng LDI lên thành 1 cặp 3D point cloud (P_0, P_1) và di chuyển 2 chiều các điểm giữa các scene flow đến thời điểm t . Sau đó chúng tôi chiếu và làm phẳng những điểm 3D đặc trưng này lên form bản đồ 2D đặc trưng trước và sau (từ P_0 và P_1 , theo thứ tự) và depth maps tương ứng. Chúng tôi kết hợp tuyến tính các map này với trọng số map là W_t tính được từ các tín hiệu không thời gian, và chuyển kết quả đến mạng tổng hợp hình ảnh để tạo ra hình ảnh cuối cùng.



2. Phương pháp

- Tạo ra dạng biểu diễn LDI từ cặp hình ảnh gần trùng nhau: sử dụng công cụ ước tính độ sâu bằng một mắt DPT để thu được dạng hình học cho mỗi hình ảnh.
- Biểu diễn cảnh không thời gian: Tính toán luồng quang học 2D giữa các hình ảnh, xác định sự tương quan lẫn nhau giữa các điểm ảnh bằng cách kiểm tra tính nhất quán trước sau rồi sử dụng giá trị độ sâu để tính tọa độ 3D và vector scene flow cho những điểm tương quan nhau này.
- Làm phẳng 2 chiều và render.
- Huấn luyện: Hệ thống được huấn luyện bằng cách sử dụng hai bộ dữ liệu, một bộ chứa các video clip có chuyển động camera nhỏ và bộ kia chứa các video clip về cảnh tĩnh với chuyển động camera đã được biết trước.

3. Kết quả dự kiến

- Mô hình hóa quang cảnh dưới dạng một cặp LDI đặc trưng được tăng cường bằng các scene flow.
- Tạo ra các video space-time có ảnh chân thực từ các cặp gần như trùng lặp mà không có các hiện vật trực quan đáng kể hoặc sự không nhất quán về thời gian.
- Phương pháp này vượt trội so với các phương pháp baseline cả về mặt định lượng và chất lượng trong các nhiệm vụ tổng hợp chế độ xem space-time.