

NHẬN DẠNG CÁC TƯ THẾ CON NGƯỜI TỪ DỮ LIỆU CỦA CẢM BIẾN ĐỘ SÂU SỬ DỤNG ĐẶC TRƯNG KHOẢNG CÁCH VÀ MÀU DA

Đinh Đồng Lương^{a*}

^aKhoa Công nghệ Thông tin, Trường Đại học Nha Trang, Khánh Hòa, Việt Nam

Nhận ngày 09 tháng 01 năm 2017 | Chính sửa ngày 11 tháng 04 năm 2017

Chấp nhận đăng ngày 18 tháng 05 năm 2017

Tóm tắt

Bài báo trình bày một cách tiếp cận mới để nhận dạng phục vụ tái tạo tư thế con người từ dữ liệu ảnh của cảm biến độ sâu (Kinect camera), sử dụng đặc trưng khoảng cách trên dữ liệu ảnh độ sâu và xác định vùng da trên ảnh màu tương ứng. Đầu tiên, thông tin từ ảnh độ sâu được xử lý và biểu diễn dạng đồ thị khoảng cách. Những điểm quan tâm trên cơ thể người hay cụ thể là các điểm khớp nối sẽ được xác định dựa vào đặc trưng khoảng và dò tìm các vùng bàn tay, bàn chân và mặt thông qua ảnh màu tương ứng bằng thuật giải dò tìm vùng da (skin detection) trên ảnh màu. Khoảng cách được tính từ các điểm trên đồ thị tới một điểm trọng tâm của cơ thể. Sử dụng các điểm khớp đã được tính để xác định ma trận chuyển đổi trên mỗi thành phần. Các ma trận được xác định sẽ được ánh xạ lên mô hình người 3D. Cuối cùng, mô hình người 3D phản ánh việc tái tạo tư thế của con người. Trong phần thực nghiệm, chúng tôi đã đánh giá kết quả của việc tái tạo lại các tư thế con người bằng phương pháp định tính cho kết quả cải thiện so với việc chỉ dùng đặc trưng khoảng cách và có thể thực hiện thời gian thực.

Từ khóa: 3D human body model; 3D human pose recovery; Depth image; Geodesic distance.

1. GIỚI THIỆU

Ngày nay, việc tái tạo các tư thế của con người dựa trên thông tin ảnh độ sâu (depth image) đang trở thành chủ đề nóng trong lĩnh vực thị giác máy tính, đặc biệt là tái tạo các tư thế phức tạp khi thông tin đầu vào của đối tượng không đầy đủ. Những kết quả của việc phân tích, tái tạo và nhận dạng các tư thế con người có ý nghĩa quan trọng trong các lĩnh vực liên quan tới các dịch vụ giải trí, theo dõi giám sát người nơi công cộng, tự động hóa, nhận dạng các hoạt động của con người, giao tiếp giữa người và máy, trong các lĩnh vực chăm sóc sức khỏe và khoa học thể thao (Moeslund, Hilton, & Kruger, 2006).

* Tác giả liên hệ: Email: luongdd@ntu.edu.vn

Hiện nay có nhiều cách tiếp cận để giải quyết bài toán tái tạo các hoạt động của con người dựa trên ảnh. Tuy nhiên, việc sử dụng ảnh độ sâu phục vụ tái tạo tư thế con người cũng được quan tâm gần đây (Chen, Wei, & Ferryman, 2013). Để tái tạo lại các tư thế, các hoạt động của con người từ các ảnh độ sâu, ba cách tiếp cận chính là tiếp cận dựa vào những cặp ảnh xạ trên hai tập của hai đối tượng (*point set registration*), dựa trên gán nhãn cho các phần trên cơ thể (*body part label*), và dựa trên đồ thị khoảng cách (*geodesic map*).

Trong cách tiếp cận thứ nhất, mục đích chính là việc tìm ra được các cặp điểm tương đương trên đối tượng mẫu và đối tượng cần tái tạo. Nhiều giải thuật cũng đã được đề xuất theo cách tiếp cận này như giải thuật ICP (*Iterative Closest Point*). Đây là giải thuật điển hình trong việc tìm ra các cặp điểm tương đương trên hai tập điểm thuộc hai đối tượng và nó được sử dụng rộng rãi trong lĩnh vực xác định mô hình 3D tương ứng với đối tượng (Tam và ctg., 2013). Trong cách tiếp cận của Kim và Kim (2008) và Mundermann, Corazza, và Andriacchi (2007), các tác giả sử dụng giải thuật ICP để ánh xạ từng phần của mô hình với đối tượng cần tái tạo thông qua các khớp nối. Tuy nhiên hạn chế của giải thuật này là chúng cần khởi tạo bằng việc đưa vào tư thế ban đầu và tư thế này luôn phải đảm bảo có sự tương tự so với mô hình cần tìm. Điều này dẫn đến việc ánh xạ từ các tư thế phức tạp thì điều kiện khởi tạo không được thỏa mãn dẫn đến kết quả không được như mong đợi do sự tối ưu cục bộ được sử dụng trong giải thuật trong việc xác định cặp điểm tương đương trên hai đối tượng dẫn đến kết quả bị sai.

Cách tiếp cận dựa trên gán nhãn cho các thành phần trên cơ thể người như trong nghiên cứu của Shotton và ctg. (2013) và Dinh, Han, Jeon, Lee, và Kim (2013), các tác giả đề xuất một phương pháp hiệu quả trong việc nhận dạng các thành phần cơ thể con người bằng cách nhận dạng từng điểm ảnh (pixel) bằng phương pháp rừng cây ngẫu nhiên RFs (*Random Forests*). Cách tiếp cận này cho phép nhận dạng hiệu quả các thành phần trên cơ thể người dựa vào ảnh độ sâu thu được từ máy ảnh chuyên dụng. Số thành phần được nhận dạng lên tới 31 nhãn. Tuy nhiên phương pháp này gặp phải khó khăn trong việc xây dựng cơ sở dữ liệu cho việc huấn luyện là rất lớn và phải chứa đựng đầy đủ các tư thế cần nhận dạng. Vì lý do này mà độ chính xác trong nhận dạng sẽ thấp đối với các cơ sở dữ liệu không đủ lớn và đầy đủ.

Dựa trên lý thuyết đồ thị để thực hiện các ước lượng vị trí và thông tin cho việc tái tạo mô hình người 3D như đề cập trong các nghiên cứu của Ganapathi, Plagemann, Koller, và Thrun (2010) và Schwarz, Mkhitarian, Mateus, và Navab (2012). Cách tiếp cận này tập trung biểu diễn lại tập điểm của ảnh độ sâu thành dạng đồ thị trọng số và sau đó xây dựng đồ thị khoảng cách được tính từ tất cả các đỉnh tới một điểm trọng tâm của đối tượng thuộc đồ thị. Cách tiếp cận này có chi phí tính toán thấp. Tuy nhiên, việc giải quyết các trường hợp có tư thế phức tạp còn nhiều hạn chế và không ổn định do việc xây dựng đồ thị khoảng cách thiếu thông tin.

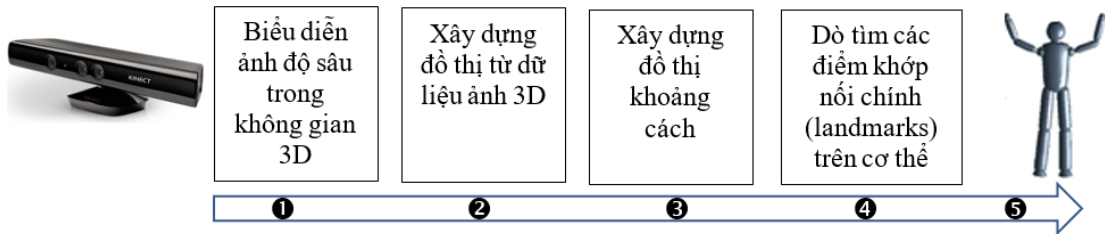
Để loại bỏ những hạn chế như đã trình bày trên, bài báo này trình bày cách tiếp cận để tái tạo tư thế toàn bộ con người từ ảnh độ sâu được thu nhận từ máy ảnh 3D (*depth camera*) và sử dụng đặc trưng khoảng cách kết hợp với việc gán nhãn và dò tìm vùng da của đối tượng trên ảnh màu được thu nhận từ máy ảnh 3D để nâng cao hiệu quả của phương pháp đề xuất. Ý tưởng cơ bản thực hiện gồm: Thông tin 3D của mỗi điểm ảnh được lấy từ ảnh độ sâu sẽ được biểu diễn dưới dạng đồ thị gồm các đỉnh và cạnh. Các điểm được quan tâm trong vùng tương ứng là các điểm khớp (*landmarks*) của mô hình người. Các vị trí này sẽ được tìm dựa trên đồ thị khoảng cách. Vị trí các điểm khớp trên cơ thể được xác định sẽ được ánh xạ sang mô hình người 3D đã được thiết lập sẵn, mô hình cơ thể người gồm khối Ellipsoid, mỗi khối sẽ tương ứng với một bộ phận cơ thể. Kết quả của việc tái tạo tư thế được cập nhật qua mô hình 3D.

Trong bài báo này, Mục 1 là giới thiệu tổng quan về bài toán và kiến thức liên quan đã và đang nghiên cứu; Mục 2 trình bày phương pháp được đề xuất; Mục 3 trình bày một số kết quả đạt được của phương pháp đề xuất. Cuối cùng là kết luận bài báo.

2. PHƯƠNG PHÁP ĐỀ XUẤT

Hình 1 trình bày các bước tái tạo tư thế của con người tự động dựa theo mô hình từ dữ liệu ảnh độ sâu được thu nhận từ cảm biến độ sâu. Mỗi ảnh độ sâu sau khi được loại bỏ thông tin nền (*background*) chỉ còn lại thông tin về toàn bộ cơ thể người (*human depth silhouette*). Tập các điểm ảnh sẽ được biểu diễn thành đồ thị trọng số gồm các đỉnh tương ứng các điểm ảnh. Việc dò tìm vị trí các điểm chính (*primary landmarks*) trên cơ thể người như đầu, hai tay và hai chân được thực hiện trên ảnh độ sâu dựa trên đồ thị khoảng

cách. Những điểm khác (*secondary landmarks*) như khuỷu tay, đầu gối chân được tính từ các điểm chính và các thông tin gắn nhãn trên mỗi điểm ảnh và những ràng buộc trên mô hình cơ thể để đạt được kết quả mong muốn. Trong nghiên cứu này chúng tôi sử dụng đặc trưng khoảng cách Euclidean được tính giữa hai điểm trên cơ thể, biết rằng khoảng cách các điểm trên một đối tượng là không thay đổi khi con người cử động và thay đổi tư thế.



Hình 1. Lưu đồ các bước tái tạo tư thế con người từ ảnh độ sâu dựa vào đặc trưng khoảng cách

2.1. Thu nhận và biểu diễn ảnh độ sâu

Trong nghiên cứu này chúng tôi sử dụng thiết bị Kinect (*Kinect Xbox 360*) của hãng Microsoft để thu nhận ảnh. Để biểu diễn dữ liệu ảnh độ sâu được chụp từ Kinect thành dữ liệu trong không gian 3D. Gọi X, Y, Z là tọa độ trong không gian 3D tương ứng trên các trục x, y , và z . Mối quan hệ giữa các điểm (*pixels*) trong ảnh độ sâu và các điểm trong không gian 3D được tính như sau:

$$X = c \frac{Z}{f}, \quad Y = v \frac{Z}{f}, \quad Z = D \quad (1)$$

Ở đây tham số f là độ dài ống kính (*focal length*), D là giá trị độ sâu của ảnh tương ứng khoảng cách được tính từ máy ảnh tới đối tượng được thu nhận, c và v là chỉ số hàng và chỉ số cột của điểm ảnh. Kết quả chuyển đổi các điểm ảnh sang không gian 3D được minh họa như trong Hình 2.



Hình 2. Biểu diễn ảnh trong không gian 3D
Ghi chú: (a) Ảnh độ sâu (depth image); (b) Ảnh trong không gian 3D

2.2. Xây dựng đồ thị từ dữ liệu ảnh 3D

Sau khi đã biểu diễn các ảnh độ sâu dưới dạng tập điểm trong không gian 3D. Gọi $x_{i,j}$ là một điểm được biểu diễn trong không gian 3D tương ứng với điểm ảnh tại vị trí (i,j) trong ảnh độ sâu. Chúng ta sẽ xây dựng đồ thị $G_t=(V_t, E_t)$, ở đây V_t là tập các đỉnh $x_{i,j}$ và $E_t \subseteq V_t \times V_t$ là các cạnh của đồ thị G_t . Hai đỉnh trong đồ thị được gọi là láng giềng hay không thì dựa vào giá trị khoảng cách giữa hai đỉnh này. Khoảng cách giữa hai đỉnh bất kỳ trong đồ thị được tính như sau:

$$E_t = \{(x_{ij}, x_{kl}) \in V_t \times V_t \mid \|x_{ij}, x_{kl}\| < \delta \cap \|(i, j)^T - (k, l)^T\|_\infty \leq 1\} \quad (2)$$

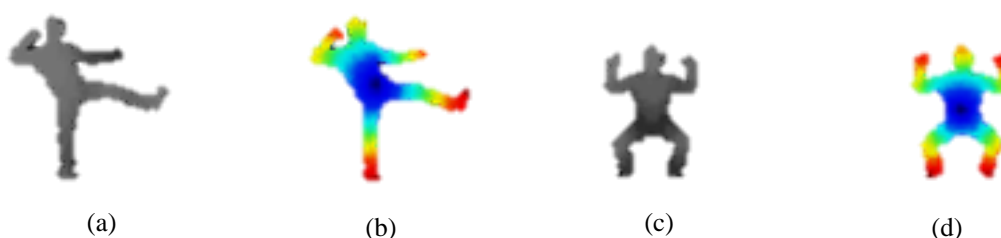
Ở đây $\|x_{ij}, x_{kl}\|_2$ độ dài Euclidean và $\|(i, j)^T - (k, l)^T\|_\infty$ là giá trị lớn nhất được chuẩn hóa và $(i, j)^T, (k, l)^T$ là tọa độ của hai điểm trong ảnh trong không gian 2D. Đối với mỗi cạnh $e=(x,y) \in E_t$, chúng ta lưu trữ một trọng số $w=||x-y||/2$. Vì vậy, việc kết nối các điểm trong không gian 3D với khoảng cách Euclidean của chúng nhỏ hơn δ . Sử dụng G_t để có thể tính khoảng cách (*geodesic distance*) $d_G(x,y)$ giữa hai điểm $x,y \in V_t$ và được tính như trong (3).

$$d_G(x, y) = \sum_{e \in SP(x,y)} w(e) \quad (3)$$

Ở đây $SP(x, y)$ chứa đựng tất cả các cạnh (SP : *Shortest path*) trên đường đi từ x tới y là ngắn nhất. Để thực hiện tìm đường đi ngắn nhất giữa hai đỉnh bất kỳ chúng tôi sử dụng giải thuật Dijkstra.

2.3. Xây dựng đồ thị khoảng cách

Từ đồ thị $G_t=(V_t, E_t)$ đã được tạo, chúng tôi xây dựng đồ thị khoảng cách được tính từ tất cả các điểm của đồ thị G_t tới một điểm trên đồ thị được coi là tâm đồ thị, tương ứng ví trí điểm tâm của cơ thể người (*human centroid point*), chúng ta có thể quan sát các điểm này trên các đối tượng trong Hình 3 là các điểm thường nằm ở vùng “bụng” trên cơ thể người. Khoảng cách các điểm tới điểm trọng tâm này được coi như không thay đổi khi tư thế thay đổi như minh họa trên Hình 3.

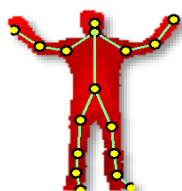


Hình 3. Minh họa các đồ thị khoảng cách được lượng tử hóa thành các mã màu (Blue: Min; Red: Max distance) trên đối tượng có tư thế khác nhau

Ghi chú: (a) Ảnh độ sâu; (b) Đồ thị khoảng cách tương ứng ảnh (a); (c) Ảnh độ sâu; (d) Đồ thị khoảng cách tương ứng ảnh (c)

2.4. Dò tìm các điểm khớp (*anatomical landmarks*)

Sau khi xây dựng đồ thị khoảng cách, chúng tôi xác định vị trí của các điểm khớp trên cơ thể người. Trong bài toán này chúng tôi xác định 17 vị trí khớp trên cơ thể như trình bày trên Hình 4. Trong đó 6 điểm chính (*primary landmarks*) và 11 điểm khác (*secondary landmarks*).



Hình 4. Vị trí của các điểm khớp trên cơ thể người

2.4.1. Định vị các điểm chính (*primary landmarks*)

Các điểm chính trên cơ thể gồm 6 điểm bao gồm điểm vùng đầu, vùng tay trái, tay phải, chân trái, chân phải và điểm trọng tâm. Vị trí của các điểm này được xác định

trong vùng bàn tay trái phải, chân trái phải, đầu và vùng bụng. Đầu tiên là dò tìm vị trí trọng tâm trên cơ thể, đây cũng là điểm trọng tâm của đồ thị khoảng cách. Để dò tìm 5 điểm còn lại được dựa vào đồ thị khoảng cách với giá trị ngưỡng τ đã được lấy bằng thực nghiệm. Từ đồ thị khoảng cách, các vùng có khoảng cách lớn nhất từ điểm trọng tâm sẽ được sử dụng để dò tìm các điểm chính. Để có được các điểm ngưỡng chính xác chúng tôi sử dụng tư thế đầu tiên đứng thẳng hai tay giơ ngang được gọi là *T-pose* cho việc khởi tạo đối mỗi đối tượng là khác nhau.

2.4.2. Định vị các điểm khác (*secondary landmarks*)

Sau khi xác định được các điểm chính, chúng tôi sử dụng các phép đo để nội suy ra 11 điểm khác tương ứng các vị trí khớp nối trên cơ thể gồm cổ, vai trái, phải, đầu gối tay trái, phải, hông, đầu gối, mắt cá chân trái, phải. Để tăng độ chính xác chúng tôi sử dụng thêm thông tin về nhãn đã được gán bằng giải thuật phân lớp Random Forest (Shotton và ctg., 2011; Dinh và ctg., 2013). Năm nhánh của đồ thị khoảng cách được tìm dựa trên 5 điểm chính đã biết trước. Kết hợp đồ thị khoảng cách và đồ thị gán nhãn ta dò tìm các điểm phụ khác (*secondary landmarks*) các bước được chi tiết được trình bày trên Hình 5.



Hình 5. Xác định các điểm phụ (*secondary landmarks*)

2.4.3. Dò tìm các điểm khớp trong tư thế phức tạp

Trong một số tư thế mà ở đó một số thành phần trên đối tượng bị che khuất, dẫn đến thông tin trên ảnh độ sâu không được đầy đủ, điều này dẫn đến việc xây dựng đồ thị khoảng cách bị thiếu thông tin và kết quả dò tìm các khớp có thể bị thiếu hoặc sai lệch. Để giải quyết trường hợp này, chúng tôi sử dụng thêm phương pháp dò tìm vùng da tay bằng phương pháp skin detection và nhận dạng vùng mặt trên ảnh màu để tăng độ chính xác dò tìm các điểm primary landmarks cũng như để nâng cao kết quả dò tìm các secondary landmarks.

2.5. Xây dựng mô hình 3D sử dụng điểm khớp

Mô hình người 3D được xây dựng từ các Ellipsoid và các điểm khớp (Thang và ctg., 2010). Ở đây mỗi phần của cơ thể được biểu diễn tương ứng một Ellipsoid và được điều khiển bởi một ma trận tham số (*kinematic parameter*). Công thức của mỗi Ellipsoid trong không gian 4D được biểu diễn như sau (Thang và ctg., 2010).

$$q(X) = X^T Q_v^T S^T D S Q_v X - 2 = 0 \quad (4)$$

Ở đây $D = \text{diag}[a^2, b^2, c^2]$ là thông tin về kích thước của Ellipsoid, S là vị trí trọng tâm của Ellipsoid trong hệ trục tọa độ, Q_θ là ma trận chuyển đổi và $X = [x, y, z, 1]^T$ là tọa độ của các điểm thuộc Ellipsoid trong 3D. Các khớp nối trong mô hình 3D có thể đạt 3 bậc tự do (*DOF: Degree of freedom*), trong khi đó sự chuyển đổi từ hệ trục toàn cục sang hệ trục cục bộ tại vị trí ‘hông’ trong mô hình đạt 6 bậc tự do (3 phép quay và 3 phép dịch chuyển theo ba trục). Toàn bộ mô hình người 3D được mô tả trong Hình 6. Để giống hơn với con người, chúng tôi sử dụng các khối super-quadrics thay cho Ellipsoid như công thức mô tả ở (5).

$$\left(\frac{x}{a_0}\right)^2 + \left(\frac{y}{b_0}\right)^2 = \left(1 + \frac{sz}{c_0}\right) \left(1 - \left(1 - \frac{2z^d}{c_0}\right)\right), \quad 0 \leq z \leq c_0 \quad (5)$$

Ở đây a_0 , b_0 và c_0 xác định kích thước của khối super-quadric theo các trục x , y và z tương ứng.



Hình 6. Mô hình người 3D với mỗi phần là một Ellipsoid (*super-quadrics*)

Việc ước lượng tổng cộng 17 điểm khớp được sử dụng để cập nhật cho các tham số của mô hình để phản ánh những thay đổi của đối tượng. Cụ thể, chúng tôi ước lượng 8 vector hướng để cập nhật lên 8 khối của mô hình tương ứng: Phần trên, dưới của tay trái, tay phải, phần trên, dưới của chân trái và chân phải. Chi tiết của mô hình và cập nhật các tham số của mô hình chi tiết trong Dinh và ctg. (2013) và Thang và ctg. (2010).

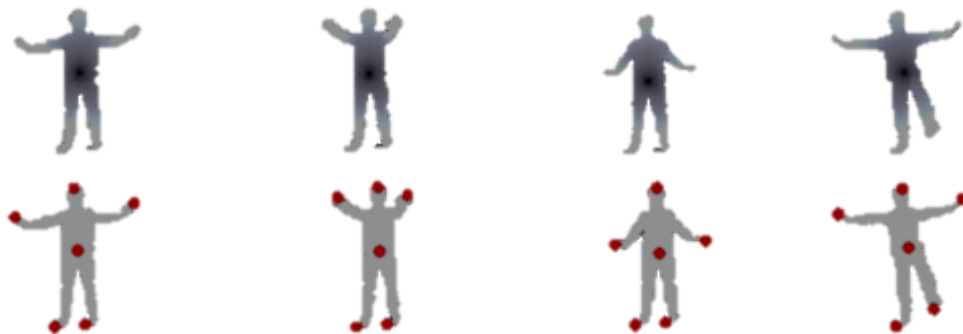
3. THỰC NGHIỆM

3.1. Thiết lập thực nghiệm

Chúng tôi đánh giá việc tái tạo các tư thế con người sử dụng cảm biến ảnh độ sâu (*Kinect camera, PrimeSense*). Kinect cho phép thu nhận được cả ảnh độ sâu và ảnh màu với độ phân giải 640x480 pixels. Mỗi chuỗi dữ liệu được thu để thực hiện trong mỗi thí nghiệm có độ dài 500 khung hình. Đối tượng thực hiện đoạn thí nghiệm được yêu cầu đứng cách cảm biến Kinect từ 2 đến 4m. Đánh giá kết quả đạt được của bài báo được dựa trên ước lượng định tính (*qualitative evaluation*), kết quả tái tạo được thể hiện ngay trên cái bóng của dữ liệu gốc đầu vào của tư thế cần tái tạo.

3.2. Dò tìm các điểm chính

Trong thí nghiệm này, chúng tôi đã xây dựng đồ thị khoảng cách để dò tìm các điểm primary landmarks đối với các tư thế đơn giản. Đầu tiên chúng tôi cố gắng tìm ra các vị trí của các điểm primary landmarks (vùng tay trái, phải, vùng chân trái, phải, vùng đầu và điểm trọng tâm vùng bụng) trên các tư thế đơn giản.

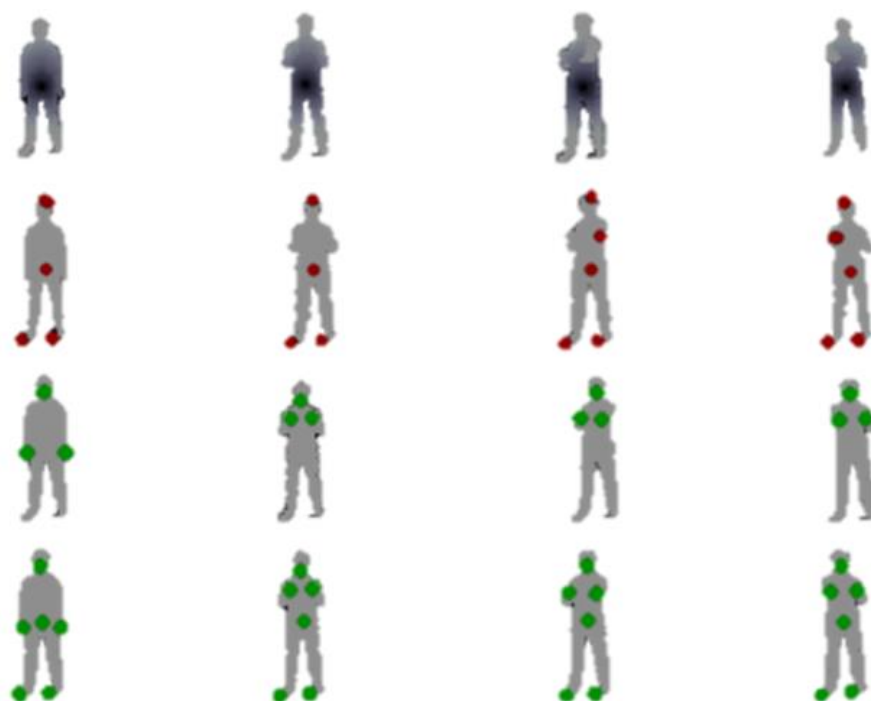


Hình 7. Kết quả của việc dò tìm các điểm chính (primary landmarks)

Ghi chú: Hàng trên là đối tượng biểu diễn bằng đồ thị khoảng cách, hàng dưới là 6 vị trí landmark được xác định trên mỗi tư thế

Việc xác định điểm trọng tâm dựa trên tính trung bình theo ba tọa độ x, y, z của toàn bộ các điểm thuộc cơ thể người. Các điểm trên vị trí đầu, hai tay và hai chân dựa trên đồ thị khoảng cách. Hình 8 trình bày một số tư thế phức tạp, việc xác định các điểm primary landmarks có thể bị thiếu hoặc sai do đồ thị khoảng cách được xây dựng bị thiếu thông tin các đỉnh cũng như khoảng cách. Trong trường hợp này nếu chỉ sử dụng đặc

trung khoảng cách thì kết quả xác định các điểm chính là thấp như trình bày ở hàng 2 của Hình 8. Để nâng cao hiệu quả của hệ thống, chúng tôi có sử dụng thêm kỹ thuật hỗ trợ tìm primary landmarks thông qua dò tìm vùng da bằng thuật giải *skin detection* trên ảnh màu tương ứng để hỗ trợ việc tìm các điểm primary landmarks thường quan sát thấy là vùng da như mặt, hai tay và hai chân.

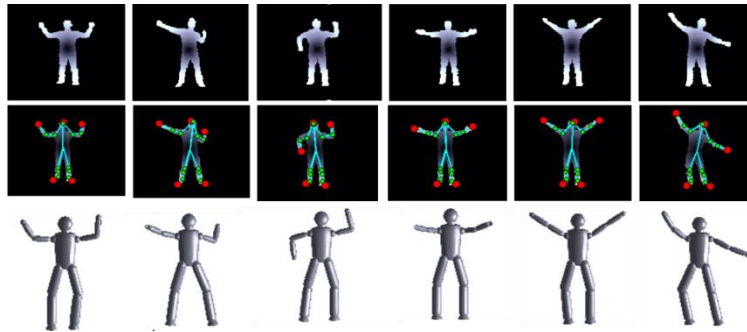


Hình 8. Kết quả dò tìm các primary landmarks cho các tư thế phức tạp

Ghi chú: Hàng thứ 1 là biểu diễn đồ thị khoảng cách được tính từ tất cả các điểm tới điểm trọng tâm đối tượng; Hàng thứ 2 là kết quả dò tìm các primary landmarks sử dụng đặc trưng khoảng cách; Hàng thứ 3 là kết quả dò tìm các primary landmarks, dò tìm vùng da và nhận dạng vùng mặt. Hàng thứ 4 là kết quả dò tìm kết hợp giữa đặc trưng khoảng cách và kỹ thuật dò tìm vùng da và nhận dạng vùng mặt

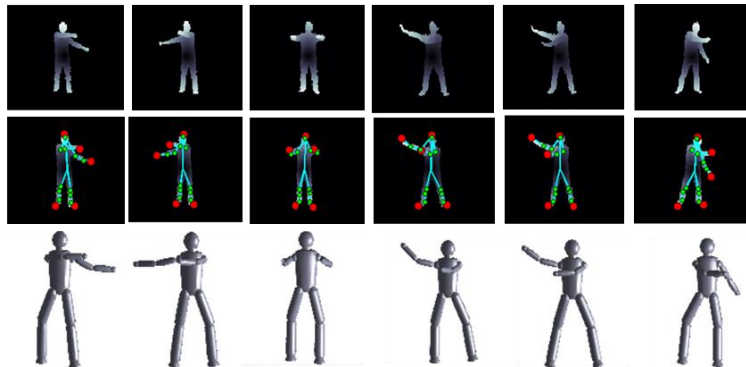
3.3. Tái tạo tư thế trên mô hình 3D

Việc xác định các vị trí các khớp trên ảnh độ sâu được sử dụng để ánh xạ lên mô hình 3D. Thông tin về vị trí của 17 điểm và dựa trên khớp này để tính ra 8 vector hướng tương ứng với 8 khối trên mô hình gồm phần trên, dưới của tay trái, phải, phần trên, dưới của chân trái và phải thông qua các ma trận biến đổi (*transformation matrix*) Q_{θ} cho mỗi phần của trên mô hình 3D. Kết quả được trình bày trên Hình 9.



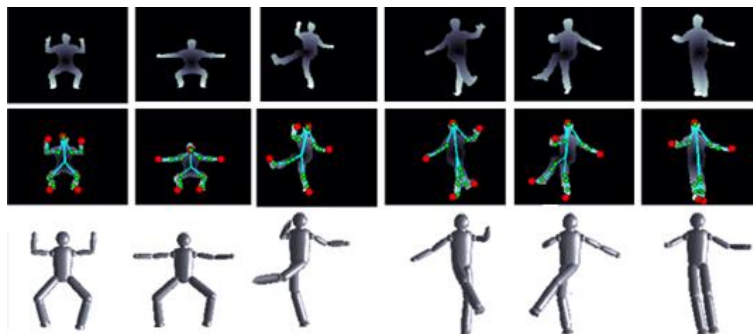
Hình 9. Tái tạo các tư thế đơn giản của con người từ ảnh độ sâu

Ghi chú: Hàng thứ 1 là đồ thị khoảng cách; Hàng thứ 2 là kết quả dò tìm các vị trí khớp trên ảnh;
Hàng thứ 3 là kết quả việc tái tạo các hoạt động con người trên mô hình 3D



Hình 10. Tái tạo các tư thế phức tạp trên hai tay của con người

Ghi chú: Hàng thứ 1 là đồ thị khoảng cách; Hàng thứ 2 là kết quả dò tìm các vị trí khớp trên ảnh;
Hàng thứ 3 là kết quả việc tái tạo các hoạt động con người trên mô hình 3D



Hình 11. Tái tạo các tư thế phức tạp trên hai tay và hai chân của con người

Ghi chú: Hàng thứ 1 là đồ thị khoảng cách; Hàng thứ 2 là kết quả dò tìm các vị trí khớp trên ảnh;
Hàng thứ 3 là kết quả việc tái tạo các hoạt động con người trên mô hình 3D

4. KẾT LUẬN

Chúng tôi đã đề xuất một hệ thống để tái tạo các tư thế con người từ ảnh độ sâu sử dụng đặc trưng khoảng cách kết hợp với việc dò tìm vùng da trên ảnh màu tương ứng để hỗ trợ việc tìm các primary landmarks cho kết quả tốt hơn. Việc xây dựng một đồ thị

khoảng cách từ các điểm trên ảnh độ sâu cho phép dò tìm các điểm khớp trên cơ thể một cách hiệu quả trên các tư thế đơn giản như có thành phần không bị bắt chéo nhau, giao nhau và tiếp xúc nhau. Để nâng cao hiệu quả hệ thống đối với các tư thế phức tạp, chúng tôi đã tích hợp thêm các kỹ thuật gán nhãn cho các vị trí landmarks trong trường hợp phương pháp dùng đặc trưng khoảng cách không xác định được bằng cách dò tìm vùng da trên ảnh màu tương ứng để tăng kết quả tái tạo.

TÀI LIỆU THAM KHẢO

- Chen, L., Wei, H., & Ferryman, J. (2013). A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, 34(15), 1995-2006.
- Dinh, D. L., Han, H. S., Jeon, H. J., Lee, S., & Kim, T. S. (2013). *Principal direction analysis-based real-time 3D human pose reconstruction from a single depth image*. Paper presented at The Fourth Symposium on Information and Communication Technology, Vietnam.
- Ganapathi, V., Plagemann, C., Koller, D., & Thrun, S. (2010). *Real time motion capture using a single time-of-flight camera*. Paper presented at The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA.
- Kim, D., & Kim, D. (2008). *A novel fitting algorithm using the ICP and the particle filters for robust 3D human body motion tracking*. Paper presented at The ACM Workshop on Vision Networks for Behavior Analysis, Canada.
- Moeslund, T. B., Hilton, A., & Kruger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2), 90-126.
- Mundermann, L., Corazza, S., & Andriacchi, T. P. (2007). *Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models*. Paper presented at the 2007 IEEE Conference on Computer Vision and Pattern Recognition, USA.
- Schwarz, L. A., Mkhitarian, A., Mateus, D., & Navab, N. (2012). Human skeleton tracking from depth data using geodesic distance and optical flow. *Image and Vision Computing*, 30(3), 217-226.
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1), 116-124.
- Tam, G. K. L., Cheng, Z. Q., Lai, Y. K., Langbein, F. C., Liu, Y. H., Marshall, D., Martin, R. R., Sun, X. F., & Rosin, P. L. (2013). Registration of 3D point clouds and meshes: A survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics*, 19(7), 1199-1217.

Thang, N. D., Kim, T. S., Lee, Y. K., & Lee, S. (2011). Estimation of 3-D human body posture via co-registration of 3-D human model and sequential stereo information. *Applied Intelligence*, 35(2), 163-177.

HUMAN POSE RECOGNITION USING GEODESIC DISTANCE AND COLOR FEATURES WITH DEPTH CAMERA

Dinh Dong Luong^{a*}

^a*The Faculty of Information Technology, Nhatrang University, Khanhhoa, Vietnam*

^{*}*Corresponding author: Email: luongdd@ntu.edu.vn*

Article history

Received: January 09th, 2017 | Received in revised form: April 11th, 2017

Accepted: May 18th, 2017

Abstract

The paper presents an approach to recover a full-body 3D human pose using geodesic and color features captured by a depth camera. The 3D information obtained from the depth images is employed to represent the points belonging to a human body in the form of a graph. The interest points or landmark locations with definite geodesic distances from the human body centroid are extracted to locate areas of the hand, foot, and face based on color images using a skin detection algorithm. Utilizing the anatomical landmark locations, joint angles of the body parts are computed. The estimated joint angles are then mapped to the body parts of a 3D human body model, which consists of a set of connected parts. Finally, the 3D human model reflects the human pose estimate. In our experiments, we assessed the detection of anatomical landmarks by our algorithm and then the 3D pose recovery for simple hand gestures, complex hand gestures, simple full body movements and complex body movements. The qualitative results of pose approximations depict the fact that the presented methodology is efficient enough to get good estimates of various full body movements.

Keywords: 3D human body model; 3D human pose recovery; Depth image; Geodesic distance.
