

NHẬN DẠNG TƯ THỂ NGƯỜI VỚI CAMERA KINECT VÀ MÁY HỌC VÉC-TƠ HỖ TRỢ

Phạm Nguyễn Khang¹ và Huỳnh Nhật Minh¹

¹ Khoa Công nghệ Thông tin & Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 19/09/2015

Ngày chấp nhận: 10/10/2015

Title:

Posture recognition using camera Kinect and Support Vector Machine

Từ khóa:

Máy học vector hỗ trợ, nhận dạng tư thế, Kinect

Keywords:

Machine Learning, SVM, Support Vector Machine, camera Kinect, posture recognition

ABSTRACT

Human posture recognition classifies a posture captured by a camera into pre-defined postures such as stand, sit, lay. One person will do a posture in front of the camera and the system will recognize what the posture is. This paper presents the ability to recognize 20 human postures with data provided by Kinect. The advantage of using skeleton data provided by Kinect is that the result of posture recognition is invariant to the change of light condition or noise of the picture. This paper also proposes 4 feature extraction methods from the data. After that, this data will be trained by support vector machine (SVM) model. The experiments showed that the accuracy of human posture recognition is above 98%.

TÓM TẮT

Nhận dạng tư thế người là phân lớp một tư thế thu được từ camera vào một trong các tư thế đã được định nghĩa trước đó ví dụ như: đứng, ngồi, nằm. Người mô tả tư thế sẽ đứng trước camera và hệ thống sẽ nhận dạng tư thế đó là tư thế gì. Trong bài viết này, chúng tôi trình bày về khả năng nhận dạng 20 tư thế người với dữ liệu thu được từ camera Kinect, dữ liệu thu được từ nhiều người với chiều cao khác nhau và góc thu dữ liệu khác nhau. Lợi thế của việc sử dụng dữ liệu khung xương thu từ camera Kinect là không bị ảnh hưởng bởi sự thay đổi của ánh sáng hay độ nhiễu của hình ảnh. Nghiên cứu cũng sẽ đưa ra 4 phương pháp trích đặc trưng từ dữ liệu khung xương thu thập được từ camera Kinect. Sau đó, bộ dữ liệu sẽ được đem đi huấn luyện bằng mô hình máy học véc-tơ hỗ trợ (SVM). Qua thực nghiệm cho thấy độ chính xác khi nhận dạng tư thế người đạt hơn 98%.

1 GIỚI THIỆU

Nhận dạng tư thế người là một đề tài được nhiều người quan tâm và nghiên cứu do có thể ứng dụng vào nhiều lĩnh vực như:

- Y tế: có thể giúp cho bệnh nhân tập vật lý trị liệu [1, 2], theo dõi bệnh nhân từ xa báo cho bác sĩ hay y tá khi bệnh nhân khi bị ngã, cần giúp đỡ hay thay đổi tư thế nằm nếu bị sai.
- Giải trí: các trò chơi mang tính tương tác.

- Hướng dẫn tập thể dục hay tập võ: giúp cho người tập có thể biết được khi nào mình tập sai tư thế, chấm điểm cho mỗi tư thế hay biết được số lượng carlo tiêu hao khi thực hiện đúng một tư thế.

Đã có rất nhiều nghiên cứu về đề tài nhận dạng tư thế người tuy nhiên hầu hết sử dụng các thông tin có được từ ảnh màu được chụp bởi camera thường [3, 4, 5]. Trở ngại chính của các phương pháp giải quyết truyền thống là việc trích xuất đặc trưng từ hình ảnh thu được bởi camera thông thường còn nhiều khó khăn do nhiễu, góc chụp,

ánh sáng, ảnh hưởng của môi trường. Trong khi đó, Microsoft đã phát triển thiết bị Kinect, thiết bị này ngoài khả năng thu được ảnh màu còn có thể cung cấp dữ liệu về độ sâu và theo dõi khung xương của người đứng trước camera.

Hiện nay, có một số đề tài nhận dạng tư thế người dựa trên dữ liệu cung cấp từ camera Kinect như: TS. Lê Thị Lan thực hiện 7 thực nghiệm với 4 cách trích xuất dữ liệu từ khung xương được cung cấp bởi thiết bị Kinect [6], kết quả của đề tài cho thấy độ chính xác cao khi nhận dạng 4 tư thế đứng, ngồi, nằm và cúi người. Đề tài “Human gesture recognition using Kinect camera” [7] của Orasa Patsadu, Chakarida Nukoolkit và Bunthit Watanapa, đề tài này đưa ra sự so sánh giữa 4 phương pháp phân loại là mạng nơron lan truyền ngược, SVM, cây quyết định và Bayes thơ ngây hay “Gesture recognition from Indian classical dance using Kinect” [8] của Sripara Saha, Shreya Ghosh, Amit Konar, Atulya K. Nagar sử dụng tọa độ của 11 khớp xương ở phần thân trên để nhận dạng 5 cử chỉ khác nhau.

Những đề tài trên đều đạt được độ chính xác cao khi sử dụng dữ liệu khung xương từ camera Kinect, tuy nhiên số lượng tư thế của các đề tài này

khá ít (3-5 tư thế). Trong bài viết này, chúng tôi sẽ trình bày khả năng nhận dạng 20 tư thế khác nhau với dữ liệu thu được từ camera Kinect.

Phần còn lại của bài viết sẽ được trình bày như sau: phần hai trình bày cụ thể về cách thu thập dữ liệu, 4 phương pháp trích xuất đặc trưng và mô hình máy học véc-tơ hỗ trợ. Kết quả và thảo luận sẽ được trình bày ở phần 3 và tiếp theo là kết luận và hướng phát triển.

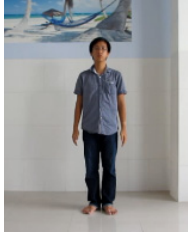
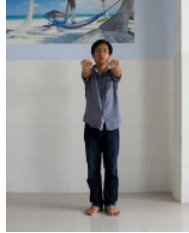


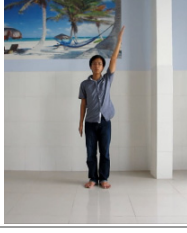

2 PHƯƠNG PHÁP NGHIÊN CỨU


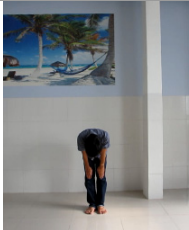
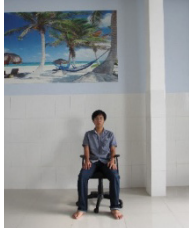





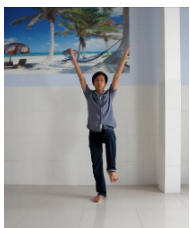





2.1 Thu thập dữ liệu

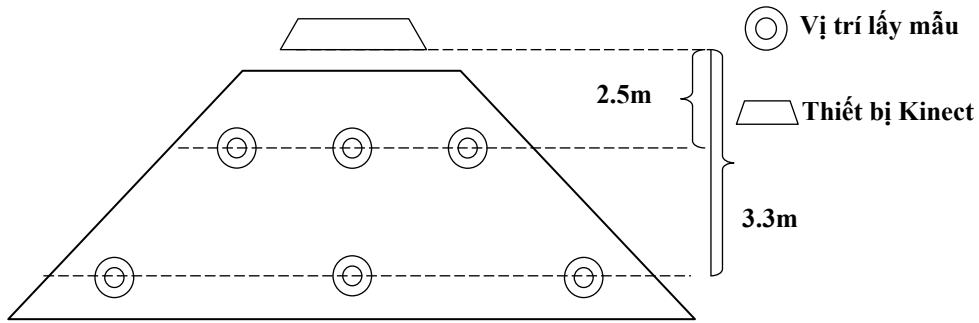
Bộ dữ liệu được thu thập từ 5 người với chiều cao từ 1.5 m-1.8 m theo quy trình sau:

- Mỗi người thu 20 tư thế khác nhau (tham khảo Bảng 1).
- Mỗi tư thế được thu ở khoảng cách 2.5 m ± 0.2 m và 3.3 m ± 0.2 m.
- Mỗi khoảng cách sẽ có 3 vị trí đứng: bên trái, ở giữa và bên phải.
- Mỗi tư thế lấy 3 góc độ: đối diện kinect, xoay trái 30° và xoay phải 30°.
- Mỗi góc độ sẽ được thu dữ liệu 10 khung xương.

Bảng 1: Hình ảnh tư thế (Tác giả đặt tên lại cho phù hợp)

TT	Hình ảnh tư thế	Mô tả	Hình ảnh tư thế	Mô tả	
1		Tư thế đứng: ứng viên đứng thẳng trên hai chân, hai tay thả lỏng, mắt nhìn thẳng về phía trước.	2		Tư thế giơ tay trước mặt: ứng viên đứng thẳng trên hai chân, hai tay giơ về phía trước vuông góc với thân người, mắt hướng về phía trước.
3		Tư thế giơ hai tay sang ngang: ứng viên đứng thẳng trên hai chân, hai tay giơ sang ngang tạo thành hình chữ T, mắt hướng về phía trước.	4		Tư thế giơ hai tay lên trời: ứng viên đứng thẳng trên hai chân, hai tay giơ lên trời, mắt hướng về phía trước.
5		Tư thế giơ tay trái lên trời: ứng viên đứng thẳng trên hai chân, tay trái giơ lên trời, tay phải thả lỏng, mắt hướng về phía trước.	6		Tư thế giơ tay phải lên trời: ứng viên đứng thẳng trên hai chân, tay phải giơ lên trời, tay trái thả lỏng, mắt hướng về phía trước.

7		<p>Tư thế khoanh tay: ứng viên đứng thẳng trên hai chân, hai tay khoanh trước ngực, mắt hướng về phía trước.</p>	8		<p>Tư thế cúi người: ứng viên cúi người, hai tay chạm gối, mắt nhìn xuống đất.</p>
9		<p>Tư thế ngồi: ứng viên ngồi trên ghế, hai tay đặt lên đùi, mắt hướng về phía trước.</p>	10		<p>Tư thế co chân trái: ứng viên đứng trên một chân, chân trái co lên căng chân vuông góc với đùi, chân phải thẳng, hai tay thả lỏng, mắt hướng về phía trước.</p>
11		<p>Tư thế co chân phải: ứng viên đứng trên một chân, chân phải co lên căng chân vuông góc với đùi, chân trái thẳng, hai tay thả lỏng, mắt hướng về phía trước.</p>	12		<p>Tư thế giơ tay chữ U: ứng viên đứng thẳng trên 2 chân, hai tay giơ lên trời khuỷu tay vuông góc, mắt hướng về phía trước</p>
13		<p>Tư thế giơ tay trái sang ngang: ứng viên đứng thẳng trên hai chân, tay trái giơ sang ngang, mắt hướng về phía trước.</p>	14		<p>Tư thế giơ tay phải sang ngang: ứng viên đứng thẳng trên hai chân, tay phải giơ sang ngang, mắt hướng về phía trước.</p>
15		<p>Tư thế tay lên trời và giơ chân trái: ứng viên đứng trên một chân, chân phải đứng thẳng, chân trái co lên căng chân vuông góc với đùi, hai tay giơ lên trời hình chữ V, mắt nhìn thẳng.</p>	16		<p>Tư thế tay lên trời và giơ chân phải: ứng viên đứng trên một chân, chân trái đứng thẳng, chân phải co lên căng chân vuông góc với đùi, hai tay giơ lên trời hình chữ V, mắt nhìn thẳng.</p>
17		<p>Tư thế tay bắt chéo: ứng viên đứng thẳng trên hai chân, hai tay bắt chéo tạo thành hình chữ X trước bụng, mắt hướng về phía trước</p>	18		<p>Tư thế Jack Feet: ứng viên đứng 2 chân dang rộng, 2 tay chắp trước ngực, mắt hướng về phía trước</p>
19		<p>Tư thế Jack Feet có sử dụng tay: ứng viên đứng 2 chân dang rộng, 2 tay chắp vào nhau và giơ lên trời, mắt hướng về phía trước</p>	20		<p>Tư thế khụy gối và tay giơ trước mặt: ứng viên đứng khụy gối, hai tay giơ trước mặt vuông góc với thân người, mắt hướng về phía trước.</p>



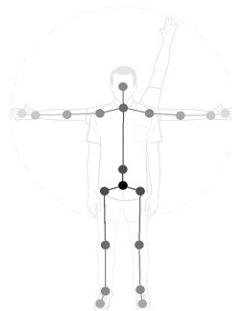
Hình 1: 6 vị trí thu dữ liệu

2.2 Trích xuất đặc trưng cơ thể người với camera Kinect

Như đã giới thiệu, camera Kinect cho phép theo dõi khung xương của người đứng trước camera, cụ thể là với mỗi khung hình camera Kinect thu được 20 khớp xương tương ứng với 20 tọa độ (x, y, z). Mỗi khớp có mỗi ID khác nhau và có gốc với ID là Hip Center.

Trong bài viết này, chúng tôi đề xuất 4 phương pháp dựa trên tọa độ của 20 khớp xương để trích đặc trưng tư thế người từ dữ liệu thu được từ camera Kinect.

- Phương pháp 1: Phương pháp này sẽ sử dụng tọa độ tuyệt đối (Absolute Position) với 3 giá trị (x, y, z) của 20 khớp xương.
- Phương pháp 2: sử dụng vị trí tương đối (Relative Position). Phương pháp 2 sử dụng tọa độ tương tự như Phương pháp 1 tuy nhiên tọa độ của 20 khớp xương được dời lại với góc tọa độ là phần đầu.
- Phương pháp 3: sử dụng 20 véc-tơ thể hiện cho góc xoay tuyệt đối (Absolute Rotation) mỗi góc bao gồm 4 giá trị (x, y, z, w).
- Phương pháp 4: trong phương pháp này 20 véc-tơ thể hiện cho góc xoay tương đối bao gồm 4 giá trị (x, y, z, w).



Hip Center				
Spine			Hip Left	Hip Right
Shoulder Center			Knee Left	Knee Right
Shoulder Left	Head	Shoulder Right	Ankle Left	Ankle Right
Elbow Left		Elbow Right	Foot Left	Foot Right
Wrist Left		Wrist Right		
Hand Left		Hand Right		

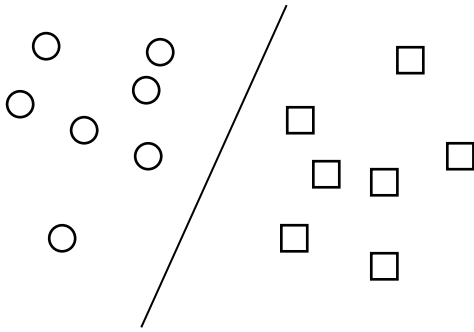
Hình 2: 20 khớp xương thu được từ camera Kinect

2.3 Nhận dạng tư thế

Sau khi đã trích đặc trưng, giai đoạn kế tiếp là huấn luyện một mô hình phân lớp có khả năng nhận dạng đúng tư thế dựa trên các đặc trưng. Phần này sẽ trình bày mô hình máy học SVM và ứng

dụng nó vào giai đoạn nhận dạng.

Để đơn giản ta sẽ xét bài toán phân lớp nhị phân sử dụng mô hình tuyến tính. Sau đó sẽ mở rộng lên bài toán phân loại nhiều lớp. [9]



Hình 3: Ví dụ về phân lớp nhị phân

Xét bài toán như sau, cho bộ dữ liệu huấn luyện \mathcal{D}

$$\mathcal{D} = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \{-1, +1\}\}_{i=1}^n$$

Với mỗi véc-tơ đầu vào $x_i \in \mathbb{R}^p$ ta có một nhãn lớp $y_i \in \{-1, +1\}$ tương ứng, dựa vào bộ dữ liệu như trên mục tiêu là tìm một siêu phẳng nhằm phân loại được các mẫu có nhãn $y = 1$ và các mẫu có nhãn $y = -1$ sao cho lề (margin) từ siêu phẳng tới các mẫu dương và mẫu âm gần siêu phẳng là cực đại.

Hàm của mặt siêu phẳng có dạng:

$$w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b = 0$$

$$\Leftrightarrow w \cdot x + b = 0 \quad (H)$$

với w là véc-tơ trọng số
 b là độ dời

* biểu thị tích vô hướng

Nếu bộ dữ liệu khả tách tuyến tính, ta có thể chọn 2 siêu phẳng để phân tách 2 lớp sao cho không có điểm nào nào giữa 2 siêu phẳng này.

Ta có 2 phương trình siêu phẳng như sau:

$$w \cdot x + b = 1 \text{ với } y = +1 \quad (H_1)$$

$$w \cdot x + b = -1 \text{ với } y = -1 \quad (H_2)$$

Vậy các điểm thuộc lớp $y = 1$ và $y = -1$ có điều kiện tương ứng là

$$w \cdot x_i + b \geq 1 \text{ với } x_i \text{ thuộc lớp } y = +1$$

$$w \cdot x_i + b \leq -1 \text{ với } x_i \text{ thuộc lớp } y = -1$$

Kết hợp 2 điều kiện ta có $y(w \cdot x + b) \geq 1$.

Khoảng cách giữa siêu phẳng (H_1) và (H_2) tới (H) là

$$d_1 = d_2 = \frac{|w \cdot x_k + b - 1|}{\|w\|} = \frac{1}{\|w\|}$$

với x_k là một điểm thuộc (H_1)

d_1 là khoảng cách từ (H_1) đến (H)

d_2 là khoảng cách từ (H_2) đến (H)

$\|w\|$ là độ dài của véc-tơ w

Vậy ta có khoảng cách giữa (H_1) và (H_2) (lề) là:

$$d = d_1 + d_2 = \frac{2}{\|w\|}$$

Vậy bài toán ban đầu trở thành bài toán tìm

$\min_{w,b} \|w\|$ với điều kiện $y(w \cdot x + b) \geq 1$ hay có thể chuyển sang một bài toán đơn giản hơn là tìm

$\min_{w,b} \frac{1}{2} \|w\|^2$ với điều kiện $y(w \cdot x + b) \geq 1$. Lời giải của bài toán này là cực tiểu hóa hàm Lagrange:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i (w \cdot x_i + b) - 1]$$

với α là các hệ số Lagrange, $\alpha \geq 0$.

Sau đó, ta chuyển thành bài toán đối ngẫu là cực đại hóa hàm $W(\alpha)$:

$$\max_{\alpha} W(\alpha) = \max_{\alpha} (\min_{w,b} L(w, b, \alpha))$$

Từ đó giải để tìm được các giá trị tối ưu cho w , b và α . Về sau, việc phân loại một mẫu mới chỉ là việc kiểm tra dấu của hàm $w \cdot x + b$.

Trong trường hợp không khả tách tuyến tính ta có thể sử dụng các hàm nhân (Kernel) để chuyển từ không gian véc-tơ ít chiều sang không gian nhiều chiều.

Một số hàm nhân cơ bản:

- Linear kernel: $K(x, y) = x \cdot y$

- Polynomial kernel: $K(x, y) = (x \cdot y + 1)^d$

- Gaussian Radial basis function kernel:

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

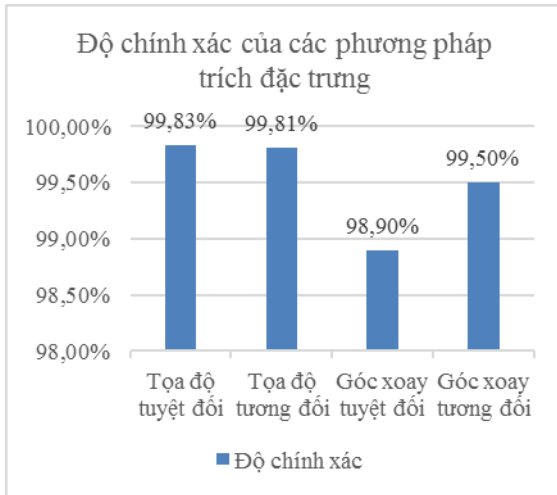
- Hyperbolic tangent kernel:

$$K(x, y) = \tanh(\alpha \cdot x \cdot y - b)$$

3 KẾT QUẢ VÀ THẢO LUẬN

Áp dụng mô hình máy học véc-tơ hỗ trợ, dữ liệu thu được từ 5 người có chiều cao khác nhau bao gồm 20 tư thế, mỗi tư thế thu được 500 mẫu. Kết quả phân lớp bằng SVM với các phương pháp

trích đặc trưng khác nhau được trình bày trong Hình 3.2. Các tham số của SVM hàm nhân tuyến tính, hằng số $c = 1000$.



Hình 4: Đồ thị biểu diễn độ chính xác của các phương pháp trích xuất đặc trưng

Dựa vào kết quả trên cho thấy phương pháp 1

đạt độ chính xác cao nhất 99.83%, kế tiếp là phương pháp 2 với 99.81%. Trong khi đó, phương pháp 3 đạt độ chính xác thấp hơn nhưng vẫn ở mức chấp nhận được là 99.50%.

Tuy nhiên, về mặt trực quan cho thấy độ chính xác của phương pháp 1 là không cao nhưng lại đạt kết quả khá tốt, về mặt dữ liệu phương pháp 1 sử dụng tọa độ các khớp làm đặc trưng, đặc trưng này có giá trị y là không đổi và giá trị z chỉ có thể xấp xỉ 2.5 hoặc 3.3. Chính vì vậy, độ chính xác của phương pháp 1 có thể không đúng, để kiểm chứng thêm độ chính xác, chúng tôi sẽ thực hiện thêm 3 thực nghiệm như sau:

- Thực nghiệm 1: Phân loại dữ liệu thu tại vị trí *bên trái* bằng cách huấn luyện bộ dữ liệu chỉ có vị trí ở *giữa* và *bên phải*.
- Thực nghiệm 2: Phân loại dữ liệu thu tại vị trí ở *giữa* bằng bộ dữ liệu thu ở vị trí *bên trái* và *bên phải*.
- Thực nghiệm 3: Phân loại dữ liệu thu tại vị trí *bên phải* bằng bộ dữ liệu thu ở vị trí *bên trái* và ở *giữa*.

Bảng 1: Độ chính xác các phương pháp trích xuất đặc trưng theo từng thực nghiệm

	Tọa độ tuyệt đối	Tọa độ tương đối	Góc xoay tuyệt đối	Góc xoay tương đối
TN 1	96.03%	97.88%	87.85%	97.21%
TN 2	98.40%	99.58%	92.35%	98.21%
TN 3	95.70%	98.29%	87.26%	97.13%
Trung bình	96.71%	98.58%	89.16%	97.51%

Dựa vào kết quả trên cho thấy độ chính xác của phương pháp 1 và 3 không cao khi loại bỏ các vị trí thu mẫu. Trong khi đó phương pháp 2 đạt độ chính xác trung bình cao nhất 98.58%.

Do đó, phương pháp sử dụng tọa độ tương đối và góc xoay tương đối nên được sử dụng để giải quyết vấn đề của bài toán bởi sự ổn định cũng như độ chính xác cao của phương pháp.

4 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Dựa vào kết quả đạt được, một mô hình nhận dạng tư thế người có rất nhiều ứng dụng như:

- Hướng dẫn tập thể dục, yoga,... Các ứng dụng dạng này có thể chấm điểm, kiểm tra các động tác có tập đúng hay không.
- Nhận dạng tư thế nằm của bệnh nhân.

Tuy nhiên, trong quá trình thu mẫu dữ liệu các bộ phận trên cơ thể bị che khuất dẫn đến khung xương không được chính xác (các điểm trên khung xương bị suy biến). Để giải quyết vấn đề này có thể sử dụng nhiều thiết bị Kinect đặt ở các vị trí khác

nhau để thu mẫu được chính xác hơn từ đó giúp cho việc nhận dạng đạt được độ chính xác cao hơn. Đây là một phương án khả thi do thiết bị Kinect có chi phí thấp.

Đề tài cũng tạo tiền đề để phát triển từ nhận dạng tư thế sang nhận dạng cử chỉ. Trong tương lai, các nghiên cứu trên có thể sử dụng thiết bị Kinect v2 và có khả năng sẽ đạt được độ chính xác cao hơn so với nghiên cứu này.

TÀI LIỆU THAM KHẢO

1. M. Peck, "Defense News," 17 December 2012. [Online]. Available: <http://www.defensenews.com/article/20121217/TSJ01/312170003/Microsoft-Wants-Kinect-Pentagon>. [Accessed 2015].
2. "InfoStrat," Information Strategies, Inc, 2015. [Online]. Available: <http://www.infostrat.com/solutions/remotion360>. [Accessed 2015].
3. I. Cohen and H. Li, "Inference of Human Postures by Classification of 3D Human

- Body Shape," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures, IOCV 2003*, 2003.
4. H.-C. Mo, J.-J. Leou and C.-S. Lin, "Human Behavior Analysis Using Multiple 2D Features and Multicategory Support Vector Machine," *MVA 2009 APR Conference on Machine Vision Applications*, Yokohama, 2009.
 5. Z. L. Haiyong Zhao, "Human Action Recognition Based on Non-linear SVM Decision Tree," pp. 7: 7 2461-2468, 2011.
 6. T.-L. Le, M.-Q. Nguyen and T.-T.-M. Nguyen, "Human posture recognition using human skeleton," *IEEE*, pp. 340-345, 2013.
 7. O. Patsadu, C. Nukoolkit and B. Watanapa, "Human gesture recognition using Kinect camera," in *2012 Ninth International Joint Conference on Computer Science and Software Engineering*, Bangkok, 2012.
 8. S. Saha, S. Ghosh, A. Konar and A. K. Nagar, "Gesture recognition from Indian classical dance using Kinect," in *CICSYN '13 Proceedings of the 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks*, Kolkata, 2013.
 9. C. M. Bishop, in *Pattern Recognition and Machine Learning (Information Science and Statistics)*, New Jersey, Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006, p. 341.