

ỨNG DỤNG CẢM BIẾN CAMERA KINECT TRONG NHẬN DẠNG NGÔN NGỮ KÍ HIỆU HỖ TRỢ VIỆC GIAO TIẾP VỚI NGƯỜI ĐIẾC

NGUYỄN THỊ PHƯƠNG*

Ngày nhận bài: 30/10/2017; ngày sửa chữa: 02/11/2017; ngày duyệt đăng: 15/11/2017.

Abstract: Automatic sign language recognition involves the use of a computer to automatically identify speech-language statements through information channels derived from the sensor to aid communication between deaf people with others or with computers. Kinect is a motion detector peripheral, integrated depth sensor, developed by Microsoft for Xbox 360 and Windows. In this paper, we propose a method for identifying sign language by capturing data from a Kinect camera.

Keywords: Sign language, Kinect camera sensor application, deaf.

1. Đặt vấn đề

Ngôn ngữ kí hiệu (NNKH) được định nghĩa như là một loại hình ngôn ngữ tự nhiên, được xây dựng phục vụ cho việc giao tiếp trong cộng đồng người Điếc, có nhiều biến thể tùy theo từng quốc gia, vùng miền. NNKH cho tiếng Việt được coi là một ngôn ngữ riêng biệt, có nhiều đặc điểm riêng nên đòi hỏi những nghiên cứu riêng. Hơn nữa, nhu cầu hỗ trợ cho việc giao tiếp của người điếc ở Việt Nam là không nhỏ. Theo kết quả từ báo cáo tổng điều tra dân số và nhà ở Việt Nam năm 2009, nước ta có 6,7 triệu người khuyết tật, chiếm khoảng 7,8% dân số, trong đó có 2,5 triệu người điếc.

Nhận dạng tự động NNKH liên quan đến việc sử dụng máy tính để nhận dạng tự động các câu nói thuộc NNKH, thông qua các kênh thông tin có được từ cảm biến (thường là video màu) nhằm hỗ trợ việc giao tiếp với người điếc.

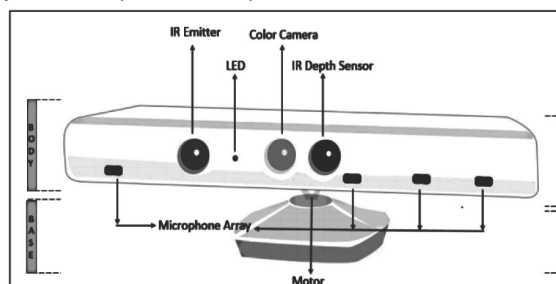
Gần đây, với sự ra đời của Camera Kinect, việc trích vị trí các khớp xương trên cơ thể người có thể được thực hiện dễ dàng. Nhiều công trình liên quan đến việc ứng dụng dữ liệu thu được từ Kinect đã được công bố như: Nhận dạng tư thế người, Nhận dạng cử chỉ và nhận dạng NNKH. Trong Agarwal và Thakur, các tác giả trình bày một phương pháp để nhận dạng các số (từ 0 đến 9) trong NNKH. Bài viết giới thiệu về Camera Kinect và cơ chế hoạt động của Camera Kinect trong việc thu nhận NNKH hỗ trợ giao tiếp với người điếc.

2. Nội dung nghiên cứu

2.1. Thu nhận dữ liệu với Camera Kinect. Kinect là thiết bị cảm biến ngoại vi thu chuyển động, tích hợp cảm biến chiều sâu, camera màu và một nhóm microphones cùng nhiều vật khác đặt bên trong một hộp nhỏ và phẳng. Hộp phẳng này được gắn kèm với một động cơ nhỏ để có thể cho phép thiết bị thay đổi góc quay theo phương ngang. Được phát triển bởi

Microsoft dành cho máy console Xbox 360 và Windows. Hiện tại, Microsoft đưa ra 2 phiên bản Kinect: một phiên bản cho Xbox 360 và 1 phiên bản cho PC (Kinect for PC). Microsoft ra mắt bộ công cụ phát triển phần mềm Kinect cho Windows 7 (SDK - Software Developer Kit), ngày 16/06/2011. Bộ công cụ phát triển phần mềm này cho phép các nhà phát triển viết ứng dụng cho Kinect ở ngôn ngữ C#, C++/CLI, Visual Basic 2005. Hiện tại, phiên bản mới nhất của bộ phần mềm này là SDK 1.8.

Thiết bị cảm biến Kinect bao gồm những thành phần sau (xem hình 1):



Hình 1. Hình ảnh về Camera Kinect

- Color camera.
- Infrared (IR) emitter.
- IR depth sensor.
- Tilt motor.
- Microphone array (gồm 4 microphone giống nhau hoạt động ở 16bit với tốc độ lấy mẫu là 16KHz).
- LED.
- Cảm biến gia tốc.

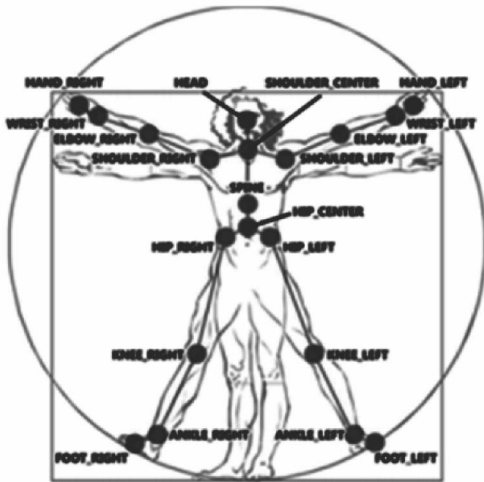
Dây microphone: Thiết bị Kinect trình diễn sự hỗ trợ tốt cho âm thanh, với sự giúp đỡ của dây microphone. Dây microphone gồm 4 microphone khác nhau, đặt theo thứ tự (3 microphone được đặt ở

* Trường Cao đẳng Sư phạm Trung ương

bên phải và 1 chiếc còn lại được đặt ở bên trái, phía cuối của thiết bị).

Mục đích của dãy microphone không chỉ là thiết bị lấy âm thanh mà còn xác định hướng của âm thanh tới. Ưu điểm chính của việc có một dãy microphone hơn việc chỉ có 1 microphone là quá trình bắt và nhận giọng nói được hỗ trợ tốt hơn, với cơ chế giảm ồn, cắt tiếng vang và công nghệ hình thành chùm âm thanh. Nó cho phép thiết bị Kinect có thể là microphone 2 chiều, xác định nguồn của âm và nhận diện ra nhiều và tiếng vang của môi trường.

Thiết bị Kinect cho phép chụp ảnh màu và ảnh độ sâu cùng một lúc. Ngoài ra, với phiên bản hiện tại, Kinect còn cho phép thu được vị trí của 20 khớp xương trên cơ thể các khớp xương được thể hiện (xem hình 2).



Hình 2. Mô hình 20 khớp xương camera Kinect có thể thu nhận

2.2. Thu nhận NNNK với Camera Kinect. Khi người ra dấu (signer) đứng đối diện với camera, dữ liệu thu thập được chính xác hơn mặc dù Kinect cho phép người ra dấu quay một góc 30° so với chính diện. Dữ liệu thu thập được từ Kinect là một chuỗi các khung, mỗi khung bao gồm 3 kênh: i) dữ liệu khung xương, mỗi khung xương gồm tọa độ của 20 khớp xương, ii) Ảnh màu (tương đương với hình ảnh thu được, với máy ảnh thông thường) và ảnh độ sâu tính từ camera. Từ dữ liệu khung xương, ta có thể trích xuất dễ dàng góc quay giữa các khớp xương để phục vụ các tác vụ khác như: nhận dạng tư thế.

Đối với NNNK, dữ liệu thu nhận được là một đoạn video ngắn, khoảng 30-35 khung (frame), cần trích thông tin quan trọng từ dữ liệu này để hệ thống phân tích và phản hồi lại trên màn hình máy tính.

Cách thức hoạt động của một hệ thống đơn giản, hỗ trợ cho việc giao tiếp với người điếc được

hình dung như sau: Giả sử có một người điếc và một người không biết NNNK. Giả sử người điếc muốn hỏi một câu với nội dung “Anh bạn là người nước nào?”, người đó sẽ thể hiện một chuỗi các hành động trước camera để thể hiện các từ “anh bạn”, “nước nào” (ở đây, việc lựa chọn này chỉ mang tính minh họa, phụ thuộc vào đặc điểm NNNK, là một nội dung cần nghiên cứu). Hệ thống máy tính thực hiện nhận dạng: xử lý thông tin thu nhận được sẽ hiển thị lên màn hình các cụm từ gốc “anh bạn”, “nước nào”. Người bình thường đọc các từ đó sẽ hiểu được ý của người khiếm thính muốn nói, trả lời bằng cách gõ nội dung vào máy tính “Tôi là người Việt Nam”. Máy tính dựa vào câu nhập sẽ tổng hợp ra một đoạn video có người mô phỏng sử dụng NNNK thể hiện các cụm từ “tôi”, “người Việt Nam” chiếu lên màn hình cho người điếc có thể xem và hiểu. Việc giao tiếp được thực hiện lặp đi lặp lại theo cùng một cách thức như vậy. NNNK và ngôn ngữ nói thông thường là khác nhau, nên một câu trong ngôn ngữ thông thường sẽ tương đương với một chuỗi các hành động trong NNNK. Mỗi hành động thể hiện một từ (cụm từ). Chúng được tạo ra bằng cách trích các từ khóa trong câu. Việc ánh xạ câu trong ngôn ngữ thông thường sang các cụm từ thể hiện các chuỗi của NNNK và ngược lại, có thể coi là bài toán dịch tự động, cần được quan tâm nghiên cứu.

Từ đó cho thấy, có hai thành phần quan trọng trong hệ thống là nhận dạng NNNK và tổng hợp NNNK. Trong đó, tập trung vào thành phần nhận dạng tự động NNNK, cụ thể là khi người đứng trước thiết bị nói (ra dấu hiệu), chương trình tính toán từ đầu vào là các tín hiệu thu được từ các kênh cảm biến để chuyển tự động chúng thành các chuỗi từ, người không biết về NNNK có thể hiểu được.

Bài toán nhận dạng NNNK sử dụng nhiều kỹ thuật của hai ngành Thị giác máy tính và Nhận dạng tiếng nói tự động. Mỗi câu nói trong NNNK tương ứng với một chuỗi các hành động (có thể liên quan tới tay hoặc không). Các hành động này có thể diễn ra song song trong một giai đoạn nào đó, nhưng không bắt buộc phải đồng bộ tuyệt đối với nhau. Các hành động tay bao gồm: dạng bàn tay, vị trí cấu âm, cách thức, hướng di chuyển,... Các hành động khác bao gồm: điều bộ cơ thể và biểu cảm của khuôn mặt.

Tuy nhiên, bên cạnh đó cũng có những trở ngại, đó là sự biến đổi lớn của các cử chỉ của riêng một người hay nhiều người khác nhau, hiệu ứng đồng

cấu âm, phân lớp cử chỉ phải phụ thuộc vào ngữ cảnh; tính không thống nhất về cách thức kí âm cho từng cử chỉ hoặc một phần của cử chỉ, tốc độ ra dấu hiệu quá cao làm mờ chuyển động, bỏ sót đặc trưng và sự cần thiết phải có hệ thống theo dõi khuôn mặt và bàn tay tự động với độ chính xác cao là những trở ngại để thấy cho việc nhận dạng tự động NCKH dựa trên hình ảnh video.

2.3. Giải thuật nhận dạng NCKH. Để hệ thống học được NCKH và có thể xuất ra dạng văn bản hoặc dạng lời nói, cần tiến hành lập trình cho Kinect, có nhiều giải thuật được đề cập tới bài toán nhận dạng NCKH như: giải thuật tối ưu bầy đàn PSO, giải thuật PCA và SVM,... tùy vào góc độ tiếp cận phân tích dữ liệu thu được trên cử chỉ tay là dạng vectơ, điểm hay dạng đồ họa mà chúng ta sẽ lựa chọn một thuật toán biểu diễn phù hợp.

3. Kết luận

Trên đây, chúng tôi đã trình bày quy trình một phương pháp mới trong việc nhận dạng NCKH với dữ liệu thu được từ camera Kinect. Chương trình sau khi hoàn thiện sẽ là một hệ thống giao tiếp giữa người điếc và người nghe một cách thuận lợi và gần gũi. Đây là một hướng đi mới để các nhà nghiên cứu tiếp tục đưa ra các giải pháp công nghệ,

giúp đỡ những người điếc trong việc tái hòa nhập cộng đồng. □

Tài liệu tham khảo

- [1] Huỳnh Hữu Hưng - Nguyễn Trọng Nguyên - Võ Đức Hoàng - Hồ Viết Hà (2012). *Nhận dạng ngôn ngữ kí hiệu tiếng Việt sử dụng mạng Neuron nhân tạo*. Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng, số: 12 (61); pp: 75-80, 2012.
- [2] Duong Van Hieu - Supot Nitsuwat (2009). *Sign Language recognition for hearing-impaired people using trajectory feature based on the Fuzzy Hidden Markov Models*. Hội thảo quốc gia lần thứ 12, một số vấn đề chọn lọc của công nghệ thông tin và truyền thông: chủ đề phát hiện tri thức từ dữ liệu, Biên Hòa, 2009.
- [3] Le Thi Lan - Minh Quoc Nguyen - Thi Thanh - Mai Nguyen (2013). *Human posture recognition using human skeleton provided by Kinect*. in Proceedings of the International Conference on Computing, Management and Telecommunications, 340-345. Agarwal, A. and M.K., Thakur. In proceedings of the 6th International Conference on Contemporary Computing (IC3), 181-185.
- [4] Tổ chức Lao động Quốc tế (2010). *Báo cáo khảo sát về đào tạo nghề và việc làm cho người khuyết tật tại Việt Nam*.
- [5] Bộ Kế hoạch và Đầu tư - Tổng cục Thống kê (2016). *Báo cáo điều tra lao động việc làm quý IV năm 2016*.

THỂ LỆ ĐĂNG BÀI BÁO KHOA HỌC

1. Bài gửi đăng trên Tạp chí Giáo dục chưa và không gửi đăng trên các sách, báo, tạp chí khác. Tạp chí không nhận đăng các bài đã đăng trên những ấn phẩm khác và không trả lại các bài không được đăng.

2. Bài viết được trình bày theo trình tự như sau: *tóm tắt* (ý tưởng và nội dung bài báo, tóm tắt bằng tiếng Việt và tiếng Anh, bao gồm cả tên bài báo, không quá 200 từ); *từ khóa* (bằng tiếng Việt và tiếng Anh); *mở đầu* (tóm tắt tình trạng nghiên cứu trong nước và thế giới, tính thời sự của vấn đề nghiên cứu,...); *nội dung bài báo* (trình bày về phương pháp nghiên cứu, phương tiện nghiên cứu, đối tượng nghiên cứu, kết quả nghiên cứu và thảo luận,...); *kết luận và thảo luận, tài liệu tham khảo*.

3. Bài gửi đăng không quá 5.000 từ, dùng font chữ Times New Roman, cỡ chữ 14, khổ A4 và không quá 10 trang (có thể gửi bản mềm dạng file word và kèm bản in); công thức toán có thể dùng phần mềm Mathtype, công thức hóa học có thể dùng phần mềm ACD/Chem Sketch hoặc Science Helper for Word; hình vẽ rõ ràng, đánh số thứ tự và tên hình vẽ phía dưới hình vẽ; bảng, biểu rõ ràng, đánh số thứ tự và tên bảng, biểu phía trên; tên riêng người, địa phương, thuật ngữ tiếng nước ngoài để nguyên vẹn, không phiên âm sang tiếng Việt; các đoạn trích dẫn trong bài để trong ngoặc kép, in nghiêng.

4. Tài liệu tham khảo để ở cuối bài có đánh số theo thứ tự trích

dẫn; có trình tự như sau: Tên tác giả (tên các tác giả cách nhau bởi dấu gạch ngang) (năm xuất bản). *Tên tài liệu tham khảo* (in nghiêng). Tên nhà xuất bản.

Chẳng hạn như sau:

[1] Nguyễn Xuân Bình (2011). *Vấn đề tự học của sinh viên năm thứ nhất Trường Cao đẳng Y tế Hà Nội hiện nay*. Tạp chí Giáo dục, số 270, tr 57-59.

[2] Đỗ Hữu Châu (1985). *Từ vựng - ngữ nghĩa tiếng Việt*. NXB Giáo dục.

[3] Trần Thị Quốc Minh (1996). *Phân tích tâm lí tình huống có vấn đề trong mối quan hệ giữa giáo viên và trẻ mẫu giáo*. Luận án tiến sĩ. Trường Đại học Sư phạm Hà Nội.

[4] Lesh, R - Caylor, B (2007). *Modeling as application versus modeling as a way to create mathematics*. International Journal of Computers for Mathematical Learning, 12, pp. 173-194.

[5] Van de Walle, J. A (2004). *Elementary and middle school mathematics: Teaching developmentally*. Pearson Education Publisher.

5. Họ và tên, học vị, chức danh khoa học, nơi công tác, email và số điện thoại của tác giả cần được ghi ở đầu bài viết, sau tên bài báo.

6. Bài viết xin gửi về Tòa soạn theo địa chỉ: Số 04, Trịnh Hoài Đức, Đống Đa, Hà Nội hoặc qua thư điện tử: tapchigiaoduc@moet.edu.vn.