

# KHAI THÁC DỮ LIỆU ẢNH ĐỘ SÂU TỪ CẢM BIẾN KINECT: SỬA LỖI VÀ ỨNG DỤNG

Đoàn Thị Hương Giang<sup>1,2</sup>, Vũ Hải<sup>1</sup>, Trần Thị Thanh Hải<sup>1</sup>

<sup>1</sup>Viện Nghiên cứu Quốc Tế MICA, HUST CNRS/UMI 2954 - Grenoble INP, Đại học Bách Khoa Hà Nội

<sup>2</sup>Trường Cao đẳng nghề công nghiệp Hà Nội

huong-giang.doan@mica.edu.vn, hai.vu@mica.edu.vn, thanh-hai.tran@mica.edu.vn

**TÓM TẮT**— Thiết bị Microsoft (MS) Kinect ngày càng được sử dụng rộng rãi trong các ứng dụng thị giác máy, tương tác người máy, robot dẫn đường... Khác với các thiết bị Time-of-flight camera đặt tiền trước đây, Kinect có ưu điểm là giá rẻ nhưng vẫn cung cấp dữ liệu về độ sâu đồng thời với ảnh RGB ở độ phân giải chấp nhận được. Tuy nhiên Kinect được thiết kế đóng kín cả về phần cứng và phần mềm, do đó phát triển ứng dụng trên Kinect thường gặp một số vấn đề khó khăn như: không tương ứng giữa tọa độ hình ảnh và độ sâu, nhiễu và lỗi đo trong ảnh độ sâu... Bài báo này sẽ trình bày một số hướng giải quyết những vấn đề căn bản này, tạo nền tảng phát triển các ứng dụng trên thiết bị Kinect. Đầu tiên, chúng tôi giới thiệu về phần cứng Kinect và các bộ thư viện mở hỗ trợ Kinect, đặc biệt là bộ thư viện mở OpenKinect. Tiếp theo, chúng tôi trình bày phương pháp căn chỉnh, sửa méo, làm khớp (Kinect Calibration) dữ liệu hình ảnh và độ sâu. Khai thác thông tin về độ sâu là một trong những điểm mạnh của Kinect, chúng tôi sẽ trình bày phương pháp sử dụng ảnh độ sâu để tách nền, ứng dụng trong bài toán phát hiện cử chỉ tay nhằm hỗ trợ trong việc nhận dạng trong các bài toán tiếp theo. Kết quả tách được so sánh đánh giá với phương pháp sử dụng dữ liệu ảnh RGB truyền thống, cho thấy ưu việt vượt trội của thiết bị Kinect trong giải quyết vấn đề tách đối tượng ra khỏi nền trong các ứng dụng thị giác máy tính.

**Từ khóa**— Kinect, Computer Vision, Detect hand, Gesture Recognition, ROC curve, histogram

## I. GIỚI THIỆU CHUNG

Cảm biến Kinect ra đời đã mở ra một lĩnh vực mới trong ngành khoa học thị giác máy tính, lĩnh vực mà các nội dung nghiên cứu từ trước đến nay chỉ tập trung ở ảnh đơn sắc (Gray scale) hoặc ảnh màu (RGB). Microsoft Kinect (MS Kinect<sup>1</sup>) ngoài khả năng cung cấp ảnh màu thì nó còn cung cấp cả ảnh độ sâu. Ứng với mỗi pixel trên ảnh màu RGB, ảnh độ sâu cho khoảng cách từ đối tượng đến thiết bị. Các ứng dụng khoa học thị giác máy tính gần đây dựa trên phân tích dữ liệu ảnh độ sâu đã là cảm hứng để phát triển rất nhiều lĩnh vực như: nhận dạng hoạt động cử chỉ người [11], phát hiện sự kiện bất thường [13], hỗ trợ robot di chuyển [14], nhận dạng đối tượng trong không gian 3-D .... Hiện nay, ở Việt nam đã có một số nhóm nghiên cứu sử dụng Kinect như: phát hiện sự kiện ngã [12], nhận dạng hoạt động của người [15]... Trong nghiên cứu này chúng tôi tập trung phân tích một số vấn đề thường gặp khi khai thác dữ liệu ảnh độ sâu như làm khớp (joint) dữ liệu độ sâu và ảnh màu; sửa nhiễu trên ảnh độ sâu. Chúng tôi đồng thời trình bày một phương pháp hiệu quả trong bài toán trừ nền chỉ sử dụng ảnh độ sâu, ứng dụng trong bài toán nhận dạng phát hiện cử chỉ tay.

### A. Các thành phần phần cứng của Microsoft Kinect

Hình 1 giới thiệu về các thành phần của một cảm biến Kinect và một ví dụ về dữ liệu hình ảnh, độ sâu thu từ cảm biến này.



Hình 1. Camera Kinect

Các thành phần bao gồm:

Đèn báo trạng thái: là đèn LED nhỏ, khi thiết bị hoạt động có màu xanh.

Bộ phận xoay: giúp cho cảm biến quay ngang  $75^\circ$ , dọc  $43^\circ$  và nghiêng  $\pm 27^\circ$ .

Máy ảnh màu (RGB): cho ảnh màu 8 bit và có hai chế độ thu nhận ảnh có độ phân giải tương ứng với tốc độ thu nhận như sau:

<sup>1</sup> Microsoft Kinect – MS Kinect : Trong bài báo này chúng tôi gọi tắt là cảm biến Kinect. Mặc dù thực tế không chỉ Microsoft phát triển Kinect; Một số sản phẩm Kinect khác như Asus, Primsense, SoftKinect.

- Tốc độ thu nhận 30 ảnh/giây cho ảnh có độ phân giải 640 x 480 điểm ảnh.
- Tốc độ thu nhận 10 ảnh/giây cho ảnh có độ phân giải 1280 x 1024 điểm ảnh.

Cảm biến độ sâu 3D: gồm máy chiếu hồng ngoại và máy ảnh hồng ngoại kết hợp cho ra ảnh độ sâu 11bit trong đó 3bit đầu định địa chỉ và chỉ có 8 bit sau là dữ liệu ảnh. Khoảng cách cho phép 0.8-3.5m. Cũng như máy ảnh màu, thiết bị kinect thu ảnh độ sâu ở hai chế độ có độ phân giải tương ứng với tốc độ thu nhận như sau:

- Tốc độ thu nhận 30 ảnh/giây cho ảnh có độ phân giải 640 x 480 điểm ảnh.
- Tốc độ thu nhận 10 ảnh/giây cho ảnh có độ phân giải 1280 x 1024 điểm ảnh.

Dây microphone: gồm các microphone bố trí dọc thiết bị MS kinect. Mỗi thiết bị có 4 bộ thu âm và xử lý âm thanh 16bit cho mỗi kênh, với tốc độ lấy mẫu 16kHz. Dây microphone này được sử dụng cho các ứng dụng âm thanh như giọng nói.

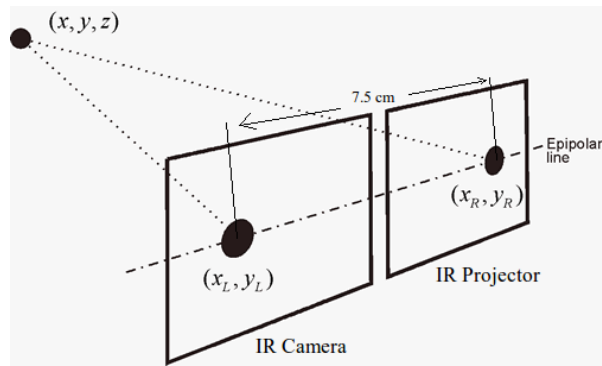
### **B. Các bộ thư viện làm việc với Kinect**

Thiết bị Kinect được thiết kế là một sản phẩm trò chơi của Microsoft, nó được đóng gói kín cả về sản phẩm phần cứng và phần mềm. Tuy nhiên các tính năng ưu việt của nó đã thu hút được sự quan tâm của đông đảo các nhà khoa học cũng như các nhà phát triển phần mềm trong lĩnh vực xử lý ảnh. Cho đến nay đã có một số thư viện nổi bật hỗ trợ phát triển ứng dụng Kinect:

- Microsoft Kinect SDK: Đây là bộ công cụ phát triển phần mềm Kinect cung cấp bởi hãng Microsoft, được cộng đồng sử dụng thiết bị kinect sử dụng hiệu nhất, phát hành phiên bản đầu vào 16/6/2011 chỉ sử dụng được trên windows 7 và chỉ cung cấp cho các mục đích phi thương mại, với các mục đích thương mại thì phải sử dụng phiên bản Kinect for Windows (K4W). Hơn nữa, đây cũng không phải là mã nguồn mở và nó chỉ hỗ trợ việc theo dõi toàn bộ đối tượng mà lại thiếu nhiều tính năng cơ bản hỗ trợ trong xử lý ảnh khi theo dõi từng bộ phận như nhận dạng cử chỉ, phát hiện tay, .... Bộ công cụ này cho phép truy xuất nhiều Kinect cùng một lúc. Địa chỉ web truy cập để tham khảo thông tin về bộ thư viện này tại <http://www.microsoft.com/en-us/kinectforwindows/>. Phiên bản của bộ công cụ này cho đến 2/2014 là v1.8 với mỗi bộ công cụ gồm KinectSDK-v1.8-Setup và KinectDeveloperToolkit-v1.8.0-Setup. Đến 10/2014 sẽ có phiên bản 2.x
- OpenNI (Open Natural Interaction): OpenNI là thư viện hỗ trợ đa ngôn ngữ trên nhiều platform khác nhau, do đó nó cho phép chạy trên nhiều hệ điều hành, giúp cho các việc viết các ứng dụng trên Kinect thuận tiện hơn. Cho phép truy xuất nhiều kinect cùng một lúc. Đây có thể được coi như là một bộ công cụ trung gian để giao tiếp với Kinect cung cấp bởi một tổ chức làm việc với mục đích phi lợi nhuận, cho phép xây dựng các ứng dụng thương mại hóa. Bộ công cụ này hỗ trợ theo dõi người cũng như việc theo dõi cử chỉ bàn tay và nhận dạng cử chỉ bàn tay, hỗ trợ bám cơ thể người. Tuy nhiên, đây không phải là mã nguồn mở. Địa chỉ web truy cập để tham khảo thông tin về bộ thư viện này tại <http://www.openni.org/>. Phiên bản cho bộ thư viện này đến tháng 2/2014 là 2.2.0.11. Đến tháng 4/2014 thuộc sở hữu của hãng Apple.
- LibFreeNect: là thư viện được tạo ra và duy trì bởi một cộng đồng mở người dùng thiết bị Kinect (OpenKinect) để sử dụng cho các ứng dụng sử dụng thiết bị Kinect. Cộng đồng OpenKinect làm việc hoàn toàn tự nguyện và không vì mục đích lợi nhuận, họ phát triển Libfreenect thành một thư viện mã nguồn mở cho các hệ điều hành khác nhau sử dụng cho thiết bị Kinect trong các ứng dụng xử lý ảnh như: Windows, Linux và OS X. Hiện tại, Libfreenect được đóng gói cho việc sử dụng trên Python, C, C++, C#, Java JNI, Java JNA, Javascript. Hơn nữa lập trình viên có thể sử dụng thư viện mã nguồn mở này cho bất cứ dự án nào, nó cũng hỗ trợ nhiều tính năng trong xử lý ảnh, cho phép truy xuất nhiều Kinect cùng một lúc. Địa chỉ web truy cập để tham khảo thông tin về bộ thư viện này tại <https://github.com/OpenKinect/libfreenect/> hoặc <http://developkinect.com/tags/libfreenect>.

### **C. Cơ chế hình thành ảnh độ sâu và các vấn đề với ảnh độ sâu**

Cơ chế hình thành ảnh độ sâu của Kinect là nó sử dụng cặp gồm camera hồng ngoại (IR camera) bộ phát ánh sáng hồng ngoại (IR Projector) phát ra ánh sáng có cấu trúc (structured light) để tạo ra giá trị độ sâu bằng công nghệ Light coding của PrimeSense. Kỹ thuật light coding dùng nguồn sáng hồng ngoại của bộ phát ánh sáng hồng ngoại chiếu liên tục vào môi trường xung quanh kết hợp với việc sử dụng máy ảnh hồng ngoại chụp lại, sau đó tính toán để thu được ảnh độ sâu. Bằng việc so sánh giữa mẫu quan sát được và mẫu tham khảo biết trước, Kinect dự đoán về khoảng cách từ Kinect đến đối tượng. Kinect có thể cung cấp ảnh độ sâu trong điều kiện ánh sáng rất tối và ít chịu ảnh hưởng của chất lượng bề mặt vật thể. Hơn nữa, thiết bị Kinect có giá thành rẻ, nhỏ gọn và rất dễ sử dụng, độ phân giải và tốc độ thu nhận ảnh chấp nhận được (1280 x 1024 điểm ảnh hoặc 640 x 480 điểm ảnh). Tuy nhiên, trong thực tế, phương pháp đo độ sâu của Kinect thường gặp một số lỗi, hoặc sai số đo phụ thuộc khoảng cách [9]. Ví dụ, khi Kinect đo ảnh độ sâu với đối tượng là một bề mặt phẳng, một số lượng nhiễu đáng kể trên ảnh độ sâu có thể quan sát được. Một vấn đề khác là phối ghép các kết quả ảnh độ sâu thu được từ các Kinect khác nhau. Các vấn đề về sửa lỗi và Calib Kinect đối với ảnh độ sâu được trình bày trong phần dưới đây.

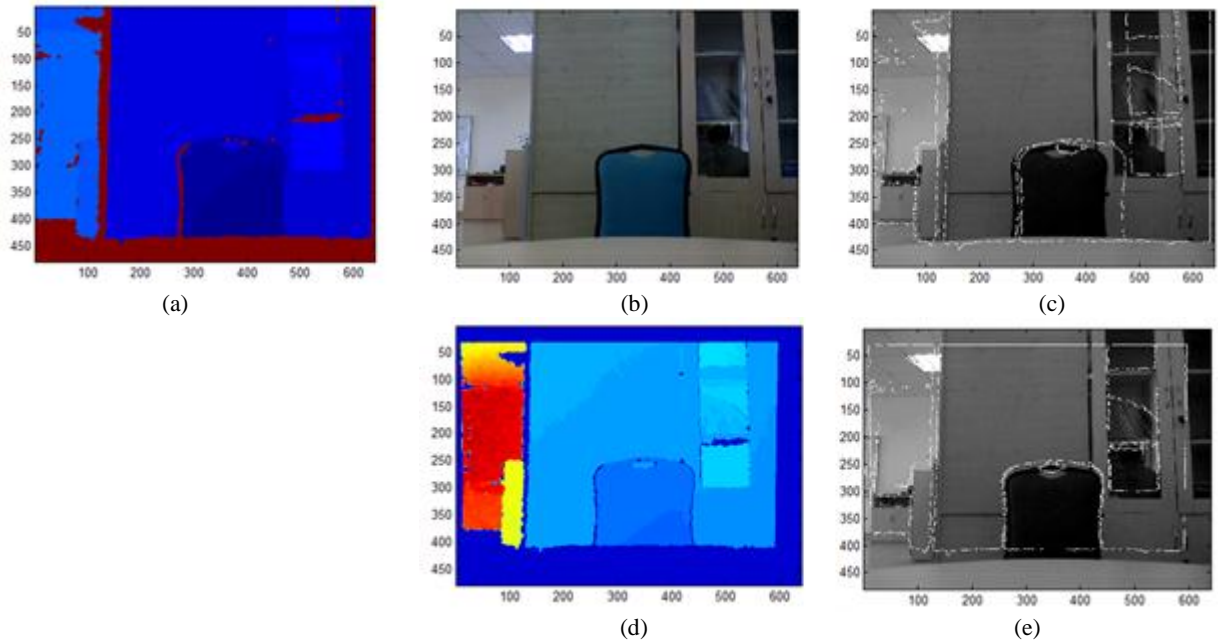


Hình 2. Phương pháp đo giá trị độ sâu của Kinect

## II. CĂN CHỈNH DỮ LIỆU ĐỘ SÂU (DEPTH) VÀ ẢNH (RGB)

### A. Kinect Calibration

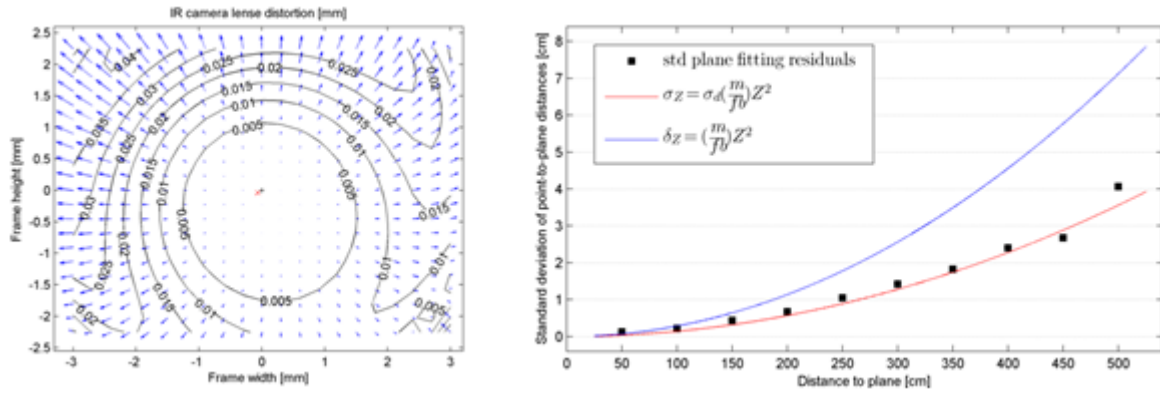
Do ảnh thu thập từ 02 cảm biến đặt ở các góc nhìn khác nhau làm cho dữ liệu thu thập có sự lệch tọa độ giữa ảnh màu và ảnh độ sâu. Calibration là phương pháp tìm một biến đổi gồm: dịch chuyển – translation và xoay (rotation) giữa các ảnh màu và độ sâu về cùng một hệ tọa độ. Một số phương pháp nghiên cứu về Calibration để joint dữ liệu độ sâu và ảnh RGB đã được nghiên cứu bởi Herrera et al [7], Yamanazoe [6]. Herrera et al. [7] sử dụng hình ảnh độ phân giải cao thu được từ một camera gắn ngoài. Herrera et al. sau đó sử dụng một mô hình không tham số. Trong phương pháp này, một số lượng lớn tham số cần phải dự đoán trước. Do đó, [7] sử dụng phương pháp tối ưu hóa để tìm ra bộ tham số này. Hiệu quả của phương pháp calib [7] thể hiện trong Hình 3. Nếu không có quá trình calib, tọa độ trên ảnh độ sâu và ảnh RGB là không tương ứng. Điều này được minh họa trên Hình 3(c), nơi các đường biên của ảnh độ sâu không vẽ đúng trên ảnh RGB. Sau khi sửa dữ liệu depth bằng phương pháp Calib [7] (Hình 3(d)), đường biên của dữ liệu depth đã Calib được vẽ trùng khớp trên ảnh RGB (Hình 3(e)). Các phương pháp calib đồng thời cũng được sử dụng khi phối ghép các dữ liệu thu từ nhiều cảm biến Kinect. Điều này đặc biệt có ý nghĩa khi dùng Kinect để tạo tập các điểm 3-D point cloud data; thường sử dụng trong các bài toán phát hiện và nhận dạng đối tượng 3-D



Hình 3. Ảnh độ sâu và ảnh màu sau khi hiệu chỉnh. (a) Ảnh độ sâu, (b) Ảnh màu tương ứng, (c) Đường biên của ảnh độ sâu vẽ trên ảnh màu, (d) Ảnh độ sâu đã được calib, (e) Đường biên của ảnh độ sâu đã calib vẽ trên ảnh màu.

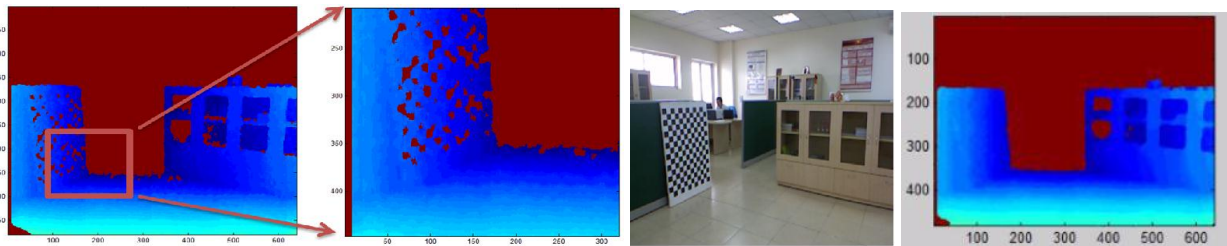
### B. Khử nhiễu với KINECT

Các cảm biến đo độ sâu (Time-of-flight) camera và Kinect đều có những vấn đề về nhiễu trên dữ liệu đo. Có một số nguyên nhân chính gây ra nhiễu trên ảnh độ sâu như: chất lượng cảm biến, phương pháp setup cách đo, thuộc tính trên bề mặt của đối tượng đo. Ngoài ra độ chính xác của phương pháp đo độ sâu cũng phụ thuộc vào khoảng cách [9]. Hình 4 (a) thể hiện lỗi trên cảm biến IR, vùng đối tượng nằm trong vùng diện tích ở tâm của khung ảnh trong bán kính xung quanh 0,005mm thì không bị méo, càng ra ngoài biên ảnh thì độ méo của hình ảnh độ sâu càng tăng; và Hình 4(b) thể hiện lỗi đo phụ thuộc vào khoảng cách, càng ở khoảng cách xa thì lỗi đo càng tăng lên.



Hình 4. Một số lỗi đối với ảnh độ sâu. (a) Lỗi méo trên ảnh độ sâu, (b) Lỗi do theo khoảng cách

Ngoài lỗi có tính hệ thống trên ảnh độ sâu, dữ liệu độ sâu đồng thời còn gặp một số lượng lớn lỗi có tính ngẫu nhiên. Đặc biệt ở vùng biên của đối tượng. Ví dụ: hình 5 thể hiện nhiều lớn khi đo trên đối tượng là một mặt phẳng. Hình 5(b) thể hiện nhiều với cường độ rất mạnh xung quanh ở các đường biên của check-board. Một số phương pháp sửa nhiễu đã được đề xuất (ví dụ [12]). Kết quả sửa nhiễu theo phương pháp [12] được thể hiện ở Hình 5(d).



Hình 5. Cơ chế giải nhiễu ảnh độ sâu. (a) Ảnh độ sâu với nhiễu. (b). Phóng to vùng nhiễu gần sát biên của đối tượng. (c). Ảnh màu RGB tương ứng. (d). Ảnh độ sâu đã được khử nhiễu

### III. SỬ DỤNG ẢNH ĐỘ SÂU TRONG BÀI TOÁN TRỪ NỀN

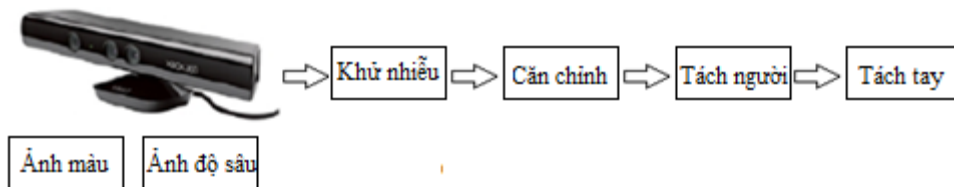
#### A. Giới thiệu bài toán trừ nền

Bài toán trừ nền nhằm mục đích tách đối tượng quan tâm ra khỏi ngữ cảnh là một trong những bài toán rất quan trọng trong các ứng dụng xử lý ảnh như: phát hiện, theo bám người,... Để giải quyết những bài toán này, những nhà khoa học có thể sử dụng đối tượng là ảnh màu, ảnh độ sâu, ảnh 3D, ... Với mỗi loại ảnh đã có nhiều kỹ thuật khác nhau như: dựa vào điểm ảnh tương ứng trên hai ảnh, dựa vào việc chia ảnh thành các khối sau đó so sánh các khối tương ứng, sử dụng đặc trưng phân bố các thuộc tính của đối tượng trong ảnh, hoặc kết hợp các thuộc tính này cũng đem đến nhiều các kỹ thuật khác nhau để trừ nền. Các bài toán trừ nền sử dụng ảnh màu đang gặp một số vấn đề hạn chế đó là các đặc trưng nêu trên không bất biến với sự thay đổi của điều kiện chiếu sáng, hay sự giống nhau của đối tượng và nền,... Gần đây, dữ liệu ảnh độ sâu của camera TOF hay Kinect đã đem đến nhiều thuận lợi cho bài toán trừ nền khác nhau và đặc biệt là bài toán trừ nền với ngữ cảnh trong nhà.

Ở đây, chúng tôi thấy rằng, ảnh độ sâu thu được từ Kinect đem đến thông tin rất tốt về khoảng cách từ các đối tượng đến camera. Hơn nữa, với hình ảnh mà chúng tôi nghiên cứu, thông thường thì đối tượng người thường gần camera hơn và tay lại gần camera hơn người (do tay ở phía trước người). Do đó, lược đồ của đối tượng người cũng như đối tượng tay sẽ có đặc trưng riêng trong ảnh. Vì vậy, chúng tôi đề xuất phương pháp sử dụng để tách nền là sử dụng đặc trưng phân bố các thuộc tính của các đối tượng trong ảnh.

#### B. Phương pháp thực hiện

Chúng tôi đã nghiên cứu nhiều phương pháp khác nhau khi sử dụng ảnh độ sâu cho bài toán trừ nền. Ở đây chúng tôi đề xuất phương pháp với các bước như sơ đồ khối trong hình 6 dưới đây:



Hình 6. Quá trình tách hình ảnh bàn tay từ ảnh thu thập từ cảm biến Kinect

Các bước hiệu chỉnh ảnh độ sâu như khử nhiễu và căn chỉnh ảnh (calibration) đã được mô tả ở phần trên.



## 1. Tách ảnh người ra khỏi nền

Tách người ra khỏi nền là một bài toán rất phổ biến trong các ứng dụng như theo bám người trong hệ thống giám sát, nhận dạng hoạt động của người (human activities). Mặc dù đã có nhiều hướng tiếp cận sử dụng các phương pháp tách nền truyền thống như so sánh khác biệt giữa frame và mô hình nền, mô hình nền sử dụng kỹ thuật GMM (Gaussian Mixture Model). Các phương pháp này thường gặp phải vấn đề khó khăn vì phần cơ thể người không phải là dạng bất biến (non-rigid object). Trong nghiên cứu này, chúng tôi khai thác khoảng dữ liệu hiệu quả của ảnh độ sâu vì thông thường đối tượng người thường gần camera hơn và tay lại gần camera hơn người (do tay ở phía trước người). Đưa vào một ảnh độ sâu  $D(x,y)$ , lược đồ (histogram) ảnh độ sâu được xây dựng như sau:

$$P_D(i) = \frac{\sum_{x,y \in (M,N)} \delta\{D(x,y)\}}{n} \quad (1)$$

Trong đó:

- M,N: là kích thước của ảnh độ sâu
- L: là giá trị độ sâu lớn nhất của ảnh
- n: là tổng số các điểm ảnh của ảnh độ sâu
- $D(x,y)$ : là ảnh độ sâu

Hình 7 minh họa quá trình xây dựng lược đồ từ ảnh độ sâu cho trước.



**Hình 7.** (a) - Ảnh độ sâu thu thập từ camera kinect. (b) - Lược đồ dữ liệu ảnh độ sâu;

Để xác định khoảng dữ liệu ảnh độ sâu hiệu quả cho việc tách người ra khỏi nền, chúng tôi tìm kiếm trên lược đồ mức xám (hình 7(b)) một ngưỡng với tiêu chuẩn lựa chọn như sau:

$$S = \sum_{i=k}^{k+m} P(i) < area \quad (2)$$

$$\text{Với } \begin{cases} P(k) > 0 \\ P(k+m) = 0 \end{cases}$$

k là giá trị bin bắt đầu của vùng diện tích

m là khoảng cách từ k đến giá trị cuối của vùng diện tích

Việc khó nhất là phải lựa chọn được ngưỡng area để có thể thu được hình ảnh người tách khỏi nền chính xác nhất. Và để chọn tham số đó, chúng tôi tiến hành so sánh ảnh thực tế phần cơ thể người mà chúng tôi đã tách nền với ảnh Ground True cơ thể người dựa vào nguyên tắc của đường cong ROC (receiver operating characteristic).

## 2. Tách tay ra khỏi người

Trừ nền bằng cách sử dụng đặc trưng phân bố các thuộc tính của các đối tượng trong ảnh dựa vào lược đồ ảnh độ sâu để thu được hình ảnh chỉ có người. Sau đó tạo lược đồ của ảnh cơ thể người đã trừ nền ở trên. Hình 8 minh họa quá trình xây dựng lược đồ từ ảnh đã tách người khỏi nền ở phần trước:



**Hình 8.** (a) - Ảnh tách người khỏi nền; (b) - Lược đồ dữ liệu ảnh cơ thể người đã tách nền

Tách tay bằng cách sử dụng đặc trưng phân bố các thuộc tính của các đối tượng trong ảnh dựa vào lược đồ và tìm kiếm giá trị vùng diện tích hiệu quả để có thể tách được hình ảnh bàn tay ra khỏi người cũng tương tự như phương pháp tách người khỏi nền. Hình vẽ sau đây minh họa hình ảnh tách tay khỏi người:



Hình 9. Ảnh bàn tay tách từ ảnh cơ thể người

**C. Đánh giá kết quả**

Chúng tôi tiến hành thử nghiệm trên nhiều video khác nhau và đánh giá kết quả trên các video số 5, 27, 56 và 58 của bộ cơ sở dữ liệu của Chalearn2013. Trong mỗi video có khoảng 20 khung hình, chúng tôi lấy ngẫu nhiên 20% số khung hình của mỗi video. Đồng thời sử dụng một tập các giá trị ngưỡng trong công thức [2]: [0,004; 0,005; 0,015; 0,03; 0,035; 0,05; 0,1; 0,15] và so sánh với ảnh Ground Truth bằng cách thực hiện việc tính các giá trị như bảng 1 dưới đây:

**Bảng 1.** Bảng các tham số TP (True Possitive), TN (True Negative), FP (False Possitive), FN (False Negative)

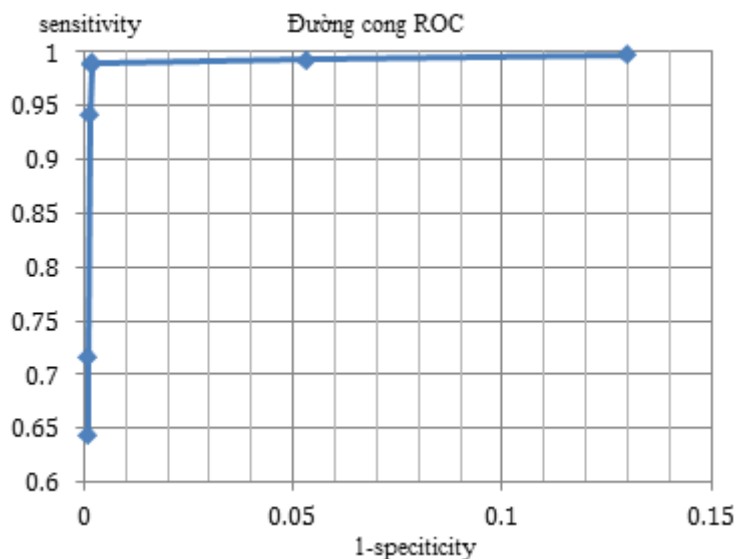
		Ảnh bàn tay tách được thực tế	
		Vùng bàn tay	Vùng ngoài bàn tay
Ảnh Ground True	Vùng bàn tay	TP	FP
	Vùng ngoài bàn tay	FN	TN

Sau khi có được các tham số TP, TN, FP, FN trên cho mỗi khung hình, chúng tôi tiến hành tính toán các tham số của sensitivity và speciticity công thức sau:

$$\text{sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{speciticity} = \frac{TN}{FP+TN} \quad (4)$$

Sau đó, tính trung bình các kết quả cặp tham số trên và vẽ đường cong ROC trong đó giá trị trục hoành là sensitivity và trục tung là 1-speciticity. Một phương pháp trừu nền tốt khi cả giá trị Sensitivity và Specivity gần sát với 1. Kết quả tính trung bình các giá trị và vẽ đường cong ROC thu được như hình 9 như sau:



Hình 10. Đồ thị đường cong ROC

Với bộ tham số thực hiện thì giá trị đạt được kết quả được xem là tốt nhất đó là giá trị area = 0,035. Và với bộ tham số này, kết quả tách tay thu được ổn định hơn, đạt được kết quả tách tay tốt hơn so với phương pháp chỉ sử dụng ảnh RGB truyền thống. Với phương pháp tách tay chỉ sử dụng ảnh RGB như trước đây gặp rất nhiều khó khăn với bài toán điều kiện chiếu sáng thay đổi, do bóng của người hoặc điều kiện nền phức tạp và nền thay đổi.

#### IV. THẢO LUẬN VÀ KẾT LUẬN

Bài báo này trình bày về việc khai thác dữ liệu ảnh độ sâu từ Kinect: chúng tôi phân tích lỗi và một số phương pháp sửa lỗi ảnh độ sâu của Kinect. Chúng tôi đề xuất phương pháp tách bàn tay khỏi nền sử dụng dữ liệu ảnh độ sâu đã được hiệu chỉnh trước đó. Phương pháp tách mà chúng tôi đề xuất tương đối đơn giản, nhanh chóng và rất hiệu quả khi chúng tôi tiến hành thử nghiệm trên một số bộ cơ sở dữ liệu có sẵn. Trong thời gian tới, chúng tôi sẽ tiếp tục nghiên cứu và cải tiến, cũng như kết hợp với ảnh màu để có thể triển khai bài toán nhận dạng cử chỉ động của tay sử dụng Kinect.

#### V. TÀI LIỆU THAM KHẢO

- [1] Kinect: <http://www.microsoft.com/en-us/kinectforwindows/>
- [2] Kinect calibration. <http://nicolas.burrus.name/index.php/Research/KinectCalibration>.
- [3] Libfreenect. [http://openkinect.org/wiki/Main\\_Page](http://openkinect.org/wiki/Main_Page).
- [4] Microsoft Kinect for Windows. <http://www.microsoft.com/en-us/kinectforwindows>.
- [5] Fabrizio Pedersoli, Nicola Adami, Sergio Benini, Riccardo Leonardi, “XKin - eXtensible Hand Pose and Gesture Recognition Library for Kinect”, Department of Information Engineering University of Brescia Brescia, Italy, 2013.
- [6] Hirotake Yamazoe, Hitoshi Habe, Ikuhisa Mitsugami, and Yasushi Yagi, “Easy Depth Sensor Calibration”, In the Proceeding of the 21st International Conference on Pattern Recognition (ICPR 2012)
- [7] D. Herrera C., J. Kannala, and J. Heikkila. Joint depth and color camera calibration with distortion correction. IEEE Trans. Pattern Analysis and Machine Intelligence,
- [8] Ahmed Elgammal, “Figure-ground segmentation - pixel-based”, Visual Analysis of Humans, Springer, 2011, pp. 31-51
- [9] K. Koshelham, S.O. Elberink, “Accuracy and resolution of Kinect depth data for indoor mapping applications”, Sensors 2012
- [10] J. Kramer et al. “Hacking the Kinect”, Apress 2012
- [11] Mao Ye et al., “A Survey on Human Motion Analysis from Depth Data”, Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, LNCS 8200, pp. 149-187, 2013
- [12] NGUYEN Van Toi et al. “Background Subtraction with KINECT data: An Efficient Combination RGB and Depth”, The first NAFOSTED Conference on Information and Computer Science (NICS 2014), pp 160–169, Hanoi, March 2014,
- [13] Caroline Rougier et al., “Fall Detection from Depth Map Video Sequences”, ICOST 2011, LNCS 6719, pp. 121–128, 2011
- [14] R. Mojtahedzaded, “Robot Obstacle Avoidance using the Kinect”, Master of Science Thesis, ETH, Sweden, 2011
- [15] TRAN Thi Thanh Hai, LE Thi Lan, MOREL Jérémy , "An analysis on human fall detection using skeleton from Microsoft Kinect", ICCE 2014

## UTILIZING DEPTH IMAGE FROM KINECT SENSOR: ERROR ANALYSIS AND ITS APPLICATIONS

Doan Thi Hương Giang<sup>1,2</sup>, Vũ Hải<sup>1</sup>, Trần Thị Thanh Hải<sup>1</sup>

<sup>1</sup> International Research Institute MICA, HUST

CNRS/UMI 2954 - Grenoble INP, Hanoi University of Science and Technology

<sup>2</sup> Industrial Vocational College Hanoi

*huong-giang.doan@mica.edu.vn, hai.vu@mica.edu.vn, thanh-hai.tran@mica.edu.vn*

**ABSTRACT**— Microsoft (MS) Kinect widely used in machine vision applications: interactive robotics, navigation robot... Kinect device, which is different from the expensive time-of-flight camera, is low cost and the ability of both depth and RGB data acquisition. However, Kinect is designed closely in both hardware and software, thereby developing Kinect applications still has some troubles such as: no correspondence between depth and RGB image coordinates, noise and measurement errors in depth image. In this paper, we propose some solutions for these and create a platform to develop applications on Kinect. First, the Kinect hardware and the Kinect support libraries, especially OpenKinect are introduced. Next, we present the calibration, distortion correction, fitting methods for color and depth image. Exploiting depth information is one of the strength of Kinect sensors, we will present a background subtraction method for depth image, applied to detect hand gestures and support for later identification problem. The experimental results shows that it overcomes to other traditional methods in background subtraction.