Kỷ yếu Hội nghị Quốc gia lần thứ X về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), Đà Nẵng, ngày 17-18/08/2017
DOI: 10.15625/vap.2017.00083

NHẬN DẠNG NGƯỜI TỪ DỮ LIỆU CHUYỂN ĐỘNG SỬ DỤNG CONVOLUTIONAL NEURAL NETWORK

Hoàng Văn Hà [1], Trần Minh Triết[2]

[1] Trường Đại học Công nghệ thông tin, ĐHQG TP. Hồ Chí Minh

[2] Trường Đại học Khoa học Tự nhiên, ĐHQG TP. Hồ Chí Minh

hahv@uit.edu.vn, tmtriet@fit.hcmus.edu.vn

TÓM TẮT: *Dáng chuyển động của mỗi người, thuật ngữ tiếng Anh là gait, được xem là duy nhất và có thể được sử dụng như một đặc điểm sinh trắc học để nhận diện cho mỗi cá nhân. Sự phát triển mạnh mẽ của điện thoại thông minh cũng như các thiết bị đeo thông minh như đồng hồ thông minh, vòng đeo hay theo dõi sức khoẻ giúp cho việc thu thập dữ liệu chuyển động của người dùng một cách dễ dàng thông qua các cảm biến chuyển động được tích hợp sẵn trong các thiết bị này.*

Nhiều phương pháp được đề xuất trong việc nhận dạng người dùng từ dữ liệu chuyển động thông qua việc biến đổi dữ liệu thành các đặc trưng được thiết kế một cách thủ công. Trong nghiên cứu này, chúng tôi đề xuất một phương pháp nhận diện người dùng từ dữ liệu chuyển động sử dụng Convolutional Neural Network - kiến trúc được sử dụng phổ biến và mang lại hiệu quả cao trong trí tuệ nhân tạo và xử lý ảnh số như một bộ trích xuất đặc trưng cấp cao tự động. Kết quả thử nghiệm với dữ liệu của 496 người có được từ bộ dữ liệu dáng chuyển động OU-ISIR của đại học Osaka - bộ dữ liệu được xem là lớn nhất về dữ liệu chuyển động gait - mang lại hiệu quả cao với độ chính xác đạt trên 99%.

Từ khóa: *Nhận diện người dùng từ dáng đi, cảm biến chuyển động, Gait, Convolutional Neural Networks.*

I. GIỚI THIỆU

Sinh trắc học, hay công nghệ sinh trắc học (Biometric) là khoa học, công nghệ sử dụng để đo lường và phân tích những đặc tính sinh học, hành vi của con người như DNA, vân tay, mống mắt. Đặc trưng sinh trắc học thường mang tính duy nhất, có thể dùng để định danh cá nhân ví dụ như các điện thoại iPhone thế hệ mới (từ dòng iPhone 5S) hỗ trợ quét vân tay cho phép mở khóa thiết bị thay vì nhập mật khẩu như thông thường. Tuy nhiên, các giải pháp chứng thực sinh trắc học thường sử dụng các đặc trưng truyền thống như vân tay (Fingerprint), võng mạc (Iris), gương mặt (Face). Việc sử dụng các loại đặc trưng này đòi hỏi thiết bị cần phải tích hợp thêm các thành phần bổ sung như camera, thiết bị chụp ảnh mống mắt hay thiết bị quét vân tay. Điều này không phải lúc nào cũng thực hiện được đối với các thiết bị đi động (mobile device), đặc biệt là các thiết bị đeo (wearable device). Chứng thực dựa trên đặc trưng sinh trắc học dáng chuyển động (gait) là một hướng tiếp cận mới để giải quyết bài toán chứng thực đặt ra trên các loại thiết bị này.

Dáng chuyển động (gait) là một đặc trưng sinh trắc học chứa thông tin về chuyển động của cơ thể người trong một khoảng thời gian. Loại đặc trưng này được đánh giá là khả thi dùng để định danh giữa những người khác nhau và phù hợp hơn các phương thức khác trên điện thoại [1][2]. Cảm biến chuyển động (inertial sensor) - gia tốc kế (accelerometer) hoặc con quay hồi chuyển (gyroscope) - được tích hợp sẵn trong các thiết bị di động cũng như thiết bị đeo thông minh tạo cơ hội cho việc thu thập dữ liệu chuyển động để xây dựng đặc trưng gait phục vụ cho bài toán định danh người dùng. Sử dụng đặc trưng dáng chuyển động có thể tạo ra các hệ thống chứng thực thông minh với mức độ thân thiện cao, ví dụ như người dùng đeo đồng hồ thông minh và vận động thì hệ thống chứng thực sử dụng gait tự động nhận biết người đeo là ai và ghi nhận lại thông tin sức khoẻ của người đó,…

Trong quá trình nhận dạng, dữ liệu chuyển động của một người được phân thành các đoạn (segment) nhỏ hơn. Các đoạn nhỏ này được biến đổi thành các mẫu dáng chuyển động (gait pattern) thông qua việc tính toán và trích rút các đặc trưng thủ công (hand-crafted feature) từ đơn giản [1][2] đến phức tạp [3][4] nhằm phục vụ cho bước định danh sau này. Convolutional Neural Network (CNN) là một thành phần hiệu quả trong việc trích rút các đặc trưng cấp cao từ dữ liệu thô đầu vào một cách tự động, mang lại nhiều thành công trong các bài toán trí tuệ nhân tạo: thị giác máy tính [5], xử lý ngôn ngữ tự nhiên [6]. Ý tưởng áp dụng CNN vào trong bài toán định dạng người dùng từ dữ liệu chuyển động đã được thử nghiệm trong [7] thu được kết quả tích cực tuy nhiên việc đánh giá trong công trình này sử dụng bộ dữ liệu với kích thước nhỏ (chỉ có 24 người). Vì thế, trong bài báo này chúng tôi muốn đề xuất một kiến trúc CNN khác và áp dụng thử nghiệm trên bộ dữ liệu với kích thước lớn hơn.

Tóm lại, những đóng góp chính của nghiên cứu này là nhóm tác giả đã đề xuất một kiến trúc CNN khác, với đặc trưng sử dụng lớp ReLU ngay sau mỗi convolutional layer, đồng thời thử nghiệm và đánh giá kiến trúc đề xuất trên dữ liệu của 496 người trong bộ dữ liệu OU-ISIR của đại học Osaka, Nhật Bản [8]. Bộ dữ liệu này được xem là bộ dữ liệu lớn nhất về dáng chuyển động gait.

Phần còn lại của bài báo được tổ chức với cấu trúc như sau: phần II trình bày các công trình liên quan đến việc nhận dạng người dùng sử dụng dáng chuyển động (gait); phần III trình bày phương pháp chúng tôi đề xuất bao gồm cách thức phân vùng và biến đổi dữ liệu cũng như hướng tiếp cận sử dụng biến thể CNN (sử dụng hàm kích hoạt phi tuyến RELU sau mỗi lớp convolutional) vào việc trích rút đặc trưng cấp cao của mẫu dáng đi (gait patterns) và phương

pháp phân lớp; quá trình thực nghiệm và đánh giá kết quả được trình bày trong phần IV; phần V là phần kết luận rút ra cũng như bàn luận về các hướng nghiên cứu tiếp theo.

II. CÁC CÔNG TRÌNH LIÊN QUAN

Ý tưởng nhận biết dáng chuyển động với thông tin sensor đã được đề xuất hơn 30 năm trước. Việc nghiên cứu sử dụng thông tin từ cảm biến chuyển động để nhận biết người dùng bắt đầu được quan tâm từ đầu thế kỷ 21, và đặc biệt trong những năm gần đây, các nghiên về vấn đề này ngày càng phát triển mạnh.

Về tổng quát, *quy trình nhận biết dáng chuyển động (gait) [9] gồm các giai đoạn: thu thập, tiền xử lý dữ liệu, phân vùng tín hiệu, xây dựng mẫu dáng chuyển động và nhận dạng*. Ở giai đoạn thu thập và tiền xử lý dữ liệu, các cảm biến chuyển động (inertial sensors) được thiết lập để thu thập dữ liệu chuyển động; nội suy tín hiệu để thu được tần số ổn định cũng như áp dụng các bộ lọc để khử nhiễu tín hiệu. Dữ liệu sau đó được phân vùng để thành các chu kỳ chuyển động (gait cycle) [10] hay các đoạn dữ liệu (frame) với kích thước cố định cho trước [11]. Dữ liệu thu được từ bước phân vùng được biến đổi thành mẫu dáng chuyển động (gait pattern) thông qua việc rút trích đặc trưng của chuyển động làm đầu vào cho quá trình nhận dạng. Thủ tục nhận dạng được thực hiện theo 2 cách chính: so khớp (pattern similarity matching) và phân lớp (classification) sử dụng máy học (machine learning). Phương pháp so khớp tính toán độ giống nhau của mẫu thu được so với mẫu dáng đi đã biết. Các độ đo thường được sử dụng để tính toán sự khác nhau giữa các mẫu gồm sự tương quan về histogram [12], khoảng cách Manhattan [8] [10], khoảng cách Euclide [8] [10] [13], hệ số tương quan [8], khoảng cách Tanimoto [8] [14], khoảng cách Hamming [15], độ đo phức tạp DTW hoặc tuỳ biến [1] [16] [17]. Phương pháp phân lớp sử dụng máy học được thực hiện bằng cách mỗi mẫu huấn luyện được gán nhãn (định danh của người có dáng đi đó) tương ứng. Quá trình phân lớp cho phép gán nhãn đã được định nghĩa trong bộ huấn luyện (train) vào các mẫu trong bộ kiểm thử (test). Với hướng tiếp cận này, một số kỹ thuật phân lớp nổi tiếng được sử dụng bao gồm K láng giềng gần nhất (k-NN) [18], SVM [19], cây quyết định [20], mạng neuron [20], …

Trong bước xây dựng mẫu dáng đi (gait pattern), dữ liệu chuyển động thu được sau bước phân vùng sẽ được biến đổi về miền không gian đặc trưng (feature space). Các đặc trưng được tạo ra có thể là các đặc trưng chung (generic feature) hoặc các đặc trưng cấp cao (advanced feature). Đặc trưng chung có thể là các tham số thống kê như độ lớn (magnitute) của vector tín hiệu [11], giá trị trung bình (mean) và sự phân bố histogram [12], tham số trong miền tần số (frequency domain) như hệ số FFT [21]. Một số phương pháp gần đây thực hiện việc trích rút các đặc trưng cấp cao và phức tạp như đặc trưng ảnh động gait (gait dynamic images) [3], đặc trưng dựa trên HOS [4]. Tuy nhiên, các phương pháp trên đều trích rút và tạo nên các đặc trưng thủ công (hand-crafted feature) mà không có phương pháp trích rút một cách tự động từ dữ liệu đầu vào. Bên cạnh đó, việc thiết kế các đặc trưng mang lại hiệu quả cao trong việc nhận dạng tốn nhiều thời gian và công sức.

Kể từ thành công ban đầu của Convolutional Neural Network (CNN) trong kiến trúc LeNet [22] phục vụ việc đọc chữ số, zip code phát triển bởi Yann LeCun trong thập niên 90 của thế kỷ trước, CNN ngày càng trở nên phổ biến và thường được dùng trong các kiến trúc học sâu (deep learning) [5] [6]. CNN cho phép trích rút đặc trưng cấp cao (high-level feature) một cách tự động với hiệu quả cao. Matteo Gadaleta và cộng sự đã tận dụng ưu điểm này để đề xuất một kiến trúc CNN sử dụng trong bài toán xác thực người dùng [7] mang lại kết quả tích cực. Kiến trúc này sử dụng 2 lớp convolutional - lớp convolutional đầu tiên với hàm kích hoạt tuyến tính (linear activation function) và lớp thứ hai sử dụng hàm kích hoạt phi tuyến (non-linear activation function) $tanh$, không sử dụng hàm kích hoạt phi tuyến RELU (Rectified Linear Units) - đi kèm theo sau bởi chỉ một lớp pooling duy nhất. So sánh với các hàm kích hoạt phi tuyến được sử dụng trong CNN, ReLU được sử dụng nhiều hơn do hàm này giúp cho quá trình huấn luyện mạng nhanh hơn với sự thay đổi độ chính xác tổng quát không đáng kể [23]. Bên cạnh đó, việc đánh giá hiệu quả trong công trình [7] được thực hiện trên tập dữ liệu với số lượng người ít (24 người).

Vì thế trong bài báo này, chúng tôi đề xuất một kiến trúc CNN khác: luôn sử dụng hàm kích hoạt phi tuyến ReLU ngay sau mỗi convolutional layer cũng như sử dụng 2 lớp max pooling, đồng thời kiểm tra đánh giá hiệu quả của phương pháp đề xuất trên tập dữ liệu chuyển động với số lượng người lớn (dữ liệu từ tập Gait database OU-ISIR [8]) để minh chứng tiềm năng ứng dụng của CNN trong bài toán nhận dạng người dùng từ dữ liệu chuyển động.

III. PHƯƠNG PHÁP ĐỀ XUẤT

Về cơ bản, việc nhận dạng người dùng dựa trên dữ liệu chuyển động được thực hiện theo quy trình như mô tả ở phần trước, riêng bước biến đổi mẫu dáng đi về miền không gian đặc trưng được áp dụng CNN. Như đã đề cập, chúng tôi sử dụng dữ liệu chuyển động từ bộ dữ liệu OU-ISIR [8] đã được thực hiện bước thu thập và tiền xử lý dữ liệu, chi tiết các bước còn lại trong quy trình được mô tả dưới đây:

A. Phân vùng và biến đổi dữ liệu

1. Phân vùng

Dữ liệu chuyển động thu được của một đối tượng được thể hiện như một ma trận **T** với kích thước ***d x n***; trong đó ***d*** là số dòng, bằng số lượng loại thông số chuyển động (ở đây ***d*** = *3* với 3 loại thông số gia tốc chuyển động Ax, Ay,

Az) và n là số lượng lần lấy mẫu giá trị thông số của đối tượng đó. Phương pháp phân vùng (segmentation) được sử dụng là phương pháp *Fixed Size Overlapping Sliding Window (FOSW)* [24].

Một cửa sổ thời gian **W** với kích thước ***d x k*** được trượt lần lượt từ trái qua phải trên ma trận **T**, trong đó ***k*** là độ rộng của cửa sổ. Mỗi lần trượt thứ ***i*** , với độ dài bước trượt ***u***, phát sinh một phân vùng $\mathbf{S_i}$ với kích thước đúng bằng của sổ thời gian **W**. Tập hợp $\mathbf{P = \{S_i\}}$ thu được chính là các phân vùng dữ liệu được sử dụng trong bước kế tiếp.

2. Biến đổi dữ liệu

Để việc trích rút đặc trưng đạt hiệu quả tốt hơn, với mỗi phân vùng dữ liệu **S** (là một ma trận có kích thước ***d x k***) trong tập **P** thu được trong bước trên, ta thực hiện tính toán một số thông số bổ sung như sau:

Với một giá trị Ax, Ay, Az tại cột thứ j trong **S**, ta tiến hành tính

- Độ lớn của vector tạo bởi 2 trong 3 thông số gia tốc Ax, Ay, Az:
  Ví dụ tính độ lớn của vector tạo bởi Ax, Ay

  $D_{j\ (Ax,\ Ay)} = \sqrt{S_{j\ (Ax)}{}^2 + S_{j\ (Ay)}{}^2}$

  Tương tự cho độ lớn của các vector tạo bởi (Ay, Az) và (Ax, Az).
  Do đó, với mỗi cột j, ta thu được thêm 3 giá trị $D_{j\ (Ax,\ Ay)}$, $D_{j\ (Ay,\ Az)}$, $D_{j\ (Ax,\ Az)}$
- Độ lớn của vector tạo bởi 3 thông số Ax, Ay, Az:

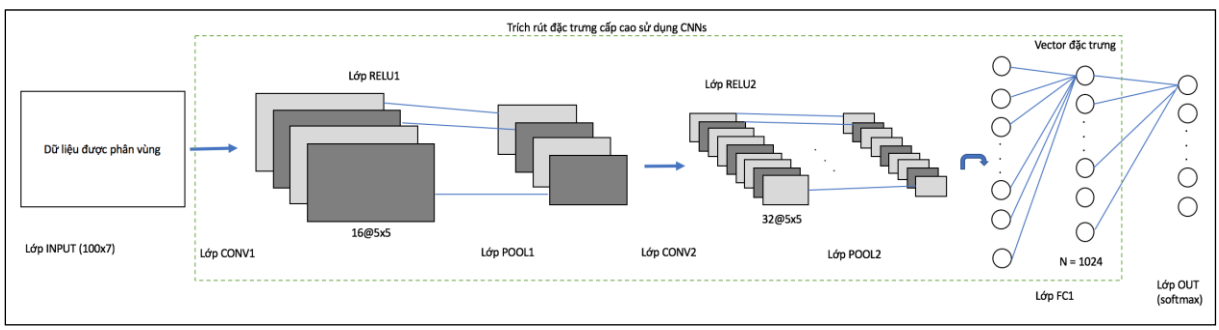
  $D_{j\ (Ax,\ Ay,\ Az)} = \sqrt{S_{j\ (Ax)}{}^2 + S_{j\ (Ay)}{}^2 + S_{j\ (Az)}{}^2}$

Sau bước này, mỗi phân vùng **S** trong tập **P** được biến đổi thành phân vùng **S'** là ma trận với kích thước ***d'x k*** (ở đây ***d'*** = ***d*** + 4, với 4 là giá trị đặc trưng được tính toán theo công thức như trên). Các phân vùng **S'** chính là dữ liệu đầu vào cho bước trích xuất đặc trưng cấp cao (high-level features) sử dụng CNN ở bước kế tiếp.

B. Rút trích đặc trưng cấp cao sử dụng CNN

Ở đây, CNN được sử dụng để trích rút đặc trưng cấp cao một cách tự động từ dữ liệu đã được phân vùng và biến đổi được trình bày trong mục trước. Kiến trúc CNN trong [7] sử dụng 2 lớp convolutional trong đó lớp convolutional đầu tiên áp dụng hàm kích hoạt tuyến tính và lớp thứ hai sử dụng hàm kích hoạt phi tuyến `tanh`, nối tiếp bởi 1 lớp max pooling. *Kiến trúc CNN được đề trong nghiên cứu này (thể hiện trong Hình 1) có điểm khác biệt so với công trình [7]: sau mỗi convolutional layer luôn được áp dụng hàm kích hoạt phí tuyến ReLU (lớp ReLU), nối tiếp bởi một lớp max pooling (nghĩa là có tổng cộng 2 lớp max pooling thay vì 1 lớp như [7]).* Kiến trúc này gồm nhiều lớp (layer) khác nhau theo thứ tự bao gồm: lớp đầu vào INPUT, lớp convolutional CONV1, lớp relu RELU1, lớp max pooling POOL1, lớp convolutional CONV2, lớp relu RELU2, lớp max pooling POOL2 và lớp Full-connected FC1.

Lớp đầu vào INPUT chính là dữ liệu thu được sau khi kết thúc bước A với kích thước 100x7. Lớp CONV1 sử dụng 16 bộ lọc (filter) với kích thước 5x5 trượt trên INPUT cho phép tìm ra mối liên hệ giữa các thành phần dữ liệu theo thời gian. Kết quả thu được là ma trận ba chiều với kích thước 100x7x16 được áp dụng hàm kích hoạt phi tuyến (non-linear activation function) Max trong lớp RELU1, trước khi được giảm kích thước xuống một nửa thông qua lớp max pooling POOL1 (giảm xuống còn 50x4x16). Dữ liệu tiếp tục được xử lý tương tự ở các lớp kế tiếp: CONV2 (với 32 bộ lọc với kích thước 5x5), RELU2 và POOL2. Kết quả thu được sau lớp POOL2 - một ma trận với kích thước 25x2x32 - được biến đổi về dạng vector 1 chiều với 1600 phần tử để làm đầu vào cho lớp full-connected layer FC1 (1024 noron). Kết quả của FC1 được xem là một vector đặc trưng sử dụng cho quá trình phân lớp (định danh người dùng).



Hình 1. Mô tả kiến trúc hệ thống, trong đó thông số sử dụng với mỗi lớp CNN được ký hiệu theo quy ước X@YxZ với X là số bộ lọc (filter), YxZ là kích thước của bộ lọc

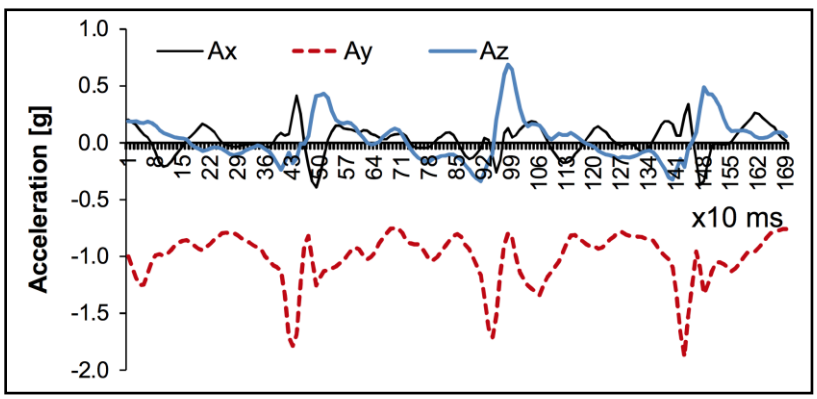
C. Nhận dạng người dùng (bài toán phân lớp - classification)

Trong nghiên cứu này, nhóm tác giả sử dụng Neural Nework thông thường để phân lớp (nhận dạng người dùng). Kết quả thu được của lớp FC1 - vector đặc trưng (feature vector) thể hiện cho một đối tượng - chính là dữ liệu đầu vào cho lớp đầu ra cuối OUT có số nơron bằng số lượng lớp (nhãn) với hàm kích hoạt *Softmax*.

IV. THỬ NGHIỆM VÀ ĐÁNH GIÁ

A. Tập dữ liệu thử nghiệm

Để tiến hành đánh giá phương pháp đề xuất, nhóm tác giả tiến hành lựa chọn bộ dữ liệu dáng chuyển động OU-ISIR của đại học Osaka (Nhật Bản) - bộ dữ liệu được xem là lớn nhất về dữ liệu chuyển động gait [8]. Bộ dữ liệu OU-ISIR gồm 2 tập dữ liệu con: Tập dữ liệu số 1 có dữ liệu của một số lượng lớn đối tượng (744 người - 389 nam và 355 nữ) trong độ tuổi từ 2 đến 78 thu được từ cảm biến chuyển động IMUZ (tích hợp cảm biến gia tốc và con quy hồi chuyển) đặt ở ngay giữa lưng của đối tượng thử nghiệm. Tuy có số lượng đối tượng lớn nhất nhưng tập dữ liệu số 1 lại chỉ tập trung ghi nhận dữ liệu về một loại trạng thái chuyển động duy nhất (di chuyển trên sàn phẳng). Ngược lại, tập dữ liệu số 2 tuy chỉ chứa dữ liệu của 496 đối tượng (ít hơn tập dữ liệu số 1) nhưng lại có dữ liệu đa dạng về các trạng thái chuyển động (di chuyển trên sàn phẳng, di chuyển lên và xuống dốc) được ghi nhận với 2 loại cảm biến chuyển động (3 cảm biến chuyển động IMUZ và 1 điện thoại thông minh Motorola ME860 chỉ có cảm biến gia tốc) đặt ở các vị trí khác nhau trên phần thắt lưng của đối tượng. Trong quá trình thu thập dữ liệu, tần số lấy mẫu f = 100 Hz. Sau khi thu được từ các cảm biến chuyển động, dữ liệu đã được chuẩn hoá bằng cách loại bỏ các giá trị sai lệch do đó chúng ta không cần thiết phải thực hiện thêm thao tác tiền xử lý dữ liệu (nội suy tín hiệu, khử nhiễu) như đã đề cập ở phần trước. Bộ dữ liệu OU-ISIR được mô tả chi tiết và cho phép tải về tại [25]. Do mục tiêu của nhóm nghiên cứu hướng đến hỗ trợ định danh người dùng trong các trạng thái chuyển động khác nhau (đi ngang trên sàn phẳng, đi lên dốc, đi xuống dốc) do đó chúng tôi tập trung thử nghiệm với dữ liệu gia tốc (theo 3 trục - Ax, Ay, Az) trên **bộ dữ liệu số 2** (gồm 496 người và có sự đa dạng về các trạng thái chuyển động).



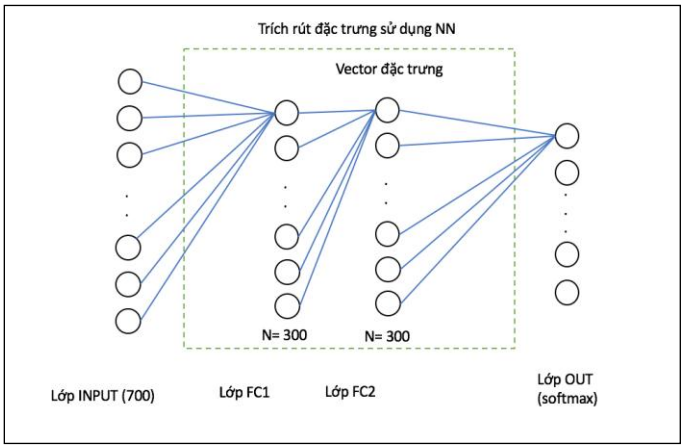
Hình 2. Một ví dụ dữ liệu gia tốc (theo 3 trục x, y, z) của một đối tượng trong bộ dữ liệu OU-ISIR [8]

B. Tiến hành thử nghiệm, kết quả và đánh giá

Với bộ dữ liệu số 2 như đã mô tả, ban đầu chúng tôi tiến hành thực nghiệm trên dữ liệu với trạng thái di chuyển trên sàn phẳng để lựa chọn được kiến trúc mạng CNN đạt hiệu quả cao. Sau đó, chúng tôi dùng chính mô hình mạng đã được xác định này để thử nghiệm với hai loại trạng thái chuyển động còn lại (đi lên dốc và xuống dốc) nhằm kiểm tra tính phù hợp của kiến trúc đề xuất với nhiều loại trạng thái chuyển động khác nhau.

Các phương pháp khác phân vùng dữ liệu bằng cách xác định các chu kỳ chuyển động (gait cycle). Trong nghiên cứu này, chúng tôi hướng tới việc hỗ trợ nhận dạng liên tục (continuous recognition) cho phép tiến hành định danh tại bất kỳ thời điểm nào. Chính vì vậy, chúng tôi cần phải cung cấp đủ các mẫu (sample) cho việc huấn luyện cũng như kiểm thử bằng cách sử dụng phương pháp Fixed Size Overlapping Sliding Window (FOSW) mô tả trong mục A, phần III. Để thử nghiệm, chúng tôi chọn độ rộng $\boldsymbol{k}$ của cửa sổ thời gian **W** với giá trị $\boldsymbol{k}$ = 100 (tương ứng với 1 giây chuyển động do dữ liệu thử nghiệm có tần số lấy mẫu *f = 100 Hz).* Với mỗi đoạn dữ liệu 1 giây, trải qua quá trình khảo sát chúng tôi nhận thấy cần tạo ra 20 biến thể khác nhau, mỗi biến thể lệch nhau một đoạn thời gian là 50 mili giây do đó giá trị độ dài bước trượt $\boldsymbol{u}$ = 5**.** Đầu tiên, chúng tôi tiên hành thực nghiệm với trạng thái di chuyển trên sàn phẳng. Với dữ liệu của một phiên di chuyển trên sàn phẳng của mỗi đối tượng trong tập dữ liệu thử nghiệm, ta tiến hành thực hiện bước phân vùng và biến đổi dữ liệu (mục A, phần III) với thông số $\boldsymbol{k}$, $\boldsymbol{u}$ mô tả ở trên. Kết thúc quá trình này thu được tổng cộng 39.085 mẫu (samples). Số lượng người (số nhãn - label) trong dữ liệu thử nghiệm là 496, lớn hơn nhiều so với số lượng 24 nhãn trong công trình [7].

Để tiến hành kiểm nghiệm mức độ hiệu quả của việc sử dụng CNN so với sử dụng Neural Network (NN) truyền thống, một hệ thống với kiến trúc sử dụng NN được xây dựng. Mô hình hệ thống này được mô tả chi tiết trong Hình 3. Lưu ý ở đây, dữ liệu đầu vào ban đầu là ma trận có kích thước 100x7 được tiến hành biến đổi về dạng vector với số lượng phần tử là 700 để làm dữ liệu đầu vào cho hệ thống.



Hình 3. Mô hình hệ thống sử dụng Neural Network sử dụng 2 hidden layer với 300 neural mỗi layer

Để phân tích hiệu quả của của kiến trúc CNN được đề xuất, ta tiến hành xây dựng một số cấu hình để thử nghiệm như trình bày trong bảng dưới đây:

Bảng 1. Các cấu hình được dùng trong thử nghiệm

Cấu hình	Mô tả cấu hình
A	Sử dụng NN (Hình 3) với số lượng neural trong lớp FC1 và FC2 là 300.
B	Sử dụng CNN. Tuy nhiên, cấu hình sử dụng ở đây tương tự như trong công trình [7]: sau lớp INPUT là một lớp convolutional CONV1 với hàm kích hoạt tuyến tính sử dụng 16 bộ lọc với kích thước 1x10, tiếp đó là một lớp convolutional CONV2 với hàm kích hoạt phi tuyến ***tanh*** sử dụng 32 bộ lọc với kích thước 3x5, nối tiếp bởi một lớp Max Pooling trước khi kết thúc bởi Full-connected layer FC1 (2048 nơron) để tạo ra vector đặc trưng.
C	Sử dụng CNN. Kiến trúc như mô tả trong Hình 1 tuy nhiên kích thước bộ lọc (filter) sử dụng ở 2 lớp CONV1 và CONV2 đổi thành 3x3.
D	Sử dụng CNN. Chính là kiến trúc đề xuất (Hình 1).

Quá trình cài đặt sử dụng thư viện mã nguồn mở Tensorflow [26] với ngôn ngữ lập trình Python. Kỹ thuật K-Fold (cross validation) với K = 10 được sử dụng để đánh giá mức độ hiệu quả của các cấu hình mô tả trong Bảng 1. Kết quả độ chính xác trung bình của các cấu hình sau quá trình thử nghiệm với dữ liệu của trạng thái di chuyển trên sàn phẳng được trình bày trong bảng sau:

Bảng 2. Độ chính xác của việc định danh theo cấu hình thử nghiệm với trạng thái di chuyển trên sàn phẳng

Cấu hình thử nghiệm	Độ chính xác - Accuracy (%)
A	73,25
B	92,68
C	96,30
D	99,19

Kết quả thể hiện trong Bảng 2 cho thấy hiện quả vượt trội của việc sử dụng CNN (cấu hình B, C, D) để trích rút đặc trưng cấp cao so với Nerual Network thông thường (cấu hình A). Cấu hình B (kiến trúc tương tự công trình [7]) mang lại độ chính xác thấp hơn cấu hình C (dựa theo kiến trúc CNN mà chúng tôi đề xuất nhưng sử dụng bộ lọc với kích thước nhỏ 3x3). Việc tăng kích thước của bộ lọc từ 3x3 (cấu hình C) lên 5x5 (cấu hình D) giúp tăng độ chính xác. Kết quả cấu hình D cho thấy kiến trúc CNN đề xuất mang lại hiệu quả cao trên 99% đối với việc nhận diện người dùng từ dữ liệu với trạng thái di chuyển trên sàn phẳng.

Để kiểm chứng xem việc sử dụng lớp ReLU ngay sau mỗi convolutional layer cũng như sử dụng 2 lớp max pooling trong kiến trúc đề xuất có giúp cho tăng tốc quá trình huấn luyện, chúng tôi tính toán thời gian huấn luyện của cấu hình B và cấu hình D trên 100 lần lặp trong điều kiện sử dụng kỹ thuật mini-batch với giá trị ***batch_size =128*** trong cùng điều kiện phần cứng. Kết quả được hiển thị trong bảng dưới đây:

Bảng 3. Kết quả tốc độ huấn luyện trên cấu hình B và D

Cấu hình thử nghiệm	Thời gian chạy 100 lần lặp (giây)
B	53,32
D	22,73

Kết quả Bảng 3 thể hiện rõ hiệu quả của việc ứng dụng lớp ReLU cũng như sử dụng 2 lớp max pooling giúp tăng tốc quá trình huấn luyện hơn 50%.

Ứng với dữ liệu từ các loại trạng thái chuyển động khác (đi lên và đi xuống dốc), chúng tôi muốn đánh giá liệu kiến trúc CNN đề xuất còn đạt hiệu quả. Vì vậy, chúng tôi tiến hành xây dựng các mẫu (sample) và chạy thử nghiệm cấu hình D theo cách giống như khi thử nghiệm với trạng thái di chuyển trên sàn phẳng. Kết quả thực nghiệm được trình bày trong bảng dưới đây:

Bảng 4. Kết quả thử nghiệm cấu hình D với trạng thái di chuyển đi lên và xuống dốc

Trạng thái di chuyển	Độ chính xác - Accuracy (%)
Lên dốc (Slope up)	99,01
Xuống dốc (Slope down)	98,50

Qua thực nghiệm, chúng tôi nhận thấy mặc dù độ chính xác của kiến trúc CNN đề xuất (cấu hình D) ứng với trạng thái di chuyển đi lên dốc và đi xuống dốc có kém hơn so với kết quả ứng với trạng thái di chuyển trên sàn phẳng nhưng mức độ hiệu quả vẫn cao (lớn hơn 98%) cho thấy mô hình đề xuất có tính tương thích cao với nhiều trạng thái di chuyển khác nhau.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong nghiên cứu này, chúng tôi đề xuất phương pháp nhận diện người dùng từ dữ liệu chuyển động với một kiến trúc CNN đề xuất sử dụng kết hợp các Convolutional layer đi kèm với lớp RELU và lớp MAX POOLING. Kết quả thử nghiệm trên nhiều trạng thái chuyển động khác nhau của bộ dữ liệu có số lượng đối tượng lớn (496 người) mang lại hiệu quả với độ chính xác cao. Điều này khẳng định CNN không chỉ hoạt động hiệu quả đối với các lĩnh vực truyền thống như thị giác máy tính, xử lý ngôn ngữ tự nhiên mà còn có thể được ứng dụng hiệu quả trong các bài toán nhận dạng liên quan đến dữ liệu chuyển động.

Tuy nhiên, việc đánh giá và xây dựng mô hình bước đầu chỉ diễn ra trên tập dữ liệu số 2 với số lượng người không phải tối đa (496 người) của bộ OU-ISIR. Việc nghiên cứu, đánh giá dự kiến được mở rộng trên dữ liệu với số lượng người tối đa (tập dữ liệu số 1 gồm 744 người), từ đó bổ sung hoàn thiện phương pháp và tiến hành so sánh với các hướng tiếp cận khác đã có thử nghiệm trên tập dữ liệu số 1. Ngoài ra, các kiến trúc mạng khác như ResNet [5], Inception [27], Recurrent Neural Network,… có thể được thử nghiệm vào việc trích rút đặc trưng cấp cao thay thế cho CNN truyền thống.

LỜI CÁM ƠN

Công trình nghiên cứu này được tài trợ bởi Đại học Quốc gia Thành phố Hồ Chí Minh (ĐHQG-HCM) trong đề tài nghiên cứu khoa học mã số B2015-18-01.

TÀI LIỆU THAM KHẢO

[1] M. O. Derawi, C. Nickely, P. Bour, C. Busc, “Unobtrusive User- Authentication on Mobile Phone susing Biometric Gait”, Proceedings of 6th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 306-311, 2010.

[2] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore, “Cell Phone- Based Biometric Identification”, Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS 2010), pp. 1-7, 2010.

[3] Zhong, Y.; Deng, Y, “Sensor orientation invariant mobile gait biometrics”. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Clearwater, FL, USA, pp. 1-8, 2014.

[4] Sprager, Sebastijan, and Matjaz B. Juric, “An efficient HOS-based gait authentication of accelerometer data”. IEEE Transactions on Information Forensics and Security 10.7: pp. 1486-1498, 2015.

[5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[6] Conneau, Alexis, Holger Schwenk, L. Barrault, and Y. Lecun. "Very deep convolutional networks for text classification." arXiv preprint arXiv:1606.01781, 2016.

[7] Gadaleta, Matteo, Luca Merelli, and Michele Rossi. “Human authentication from ankle motion data using convolutional neural networks”. Statistical Signal Processing Workshop (SSP), 2016 IEEE. IEEE, 2016.

[8] T. T Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, “The largest inertial sensor-based gait database and performance evaluation of gait- based personal authentication”, Pattern Recognition Volume 47, Issue 1, pp. 228-237, 2014.

[9] Sprager, Sebastijan, and Matjaz B. Juric. "Inertial sensor-based gait recognition: a review." Sensors 15.9 : 22089-22127, 2015.

[10] M. Derawi, P. Bours, “Gait and activity recognition using commercial phones”, Computers & Security, Volume 39, Part B, pp.137-144, 2013.

[11] Raziff, Abdul Rafiez Abdul, et al. "Gait identification using One-vs-one classifier model." Open Systems (ICOS), 2016 IEEE Conference on. IEEE, 2016.

[12] D. Gafurov, K. Helkala, T. Søndrol, "Biometric gait authentication using accelerometer sensor". Journal of computers, vol. 1, no. 7, pp. 51-59, 2006.

[13] C. Nickel, M. O. Derawi, P. Bours, C. Busch, "Scenario test of accelerometer-based biometric gait recognition", In Proceedings of the Third International Workshop on Security and Communication Networks (IWSCN), Gjovik, Norway; pp. 15-21, 2011.

[14] R. Subramanian, S. Sarkar; M. Labrador, K. Contino; C. Eggert, O. Javed, J. Zhu, H. Cheng (2015), "Orientation invariant gait matching algorithm based on the Kabsch alignment", In Proceedings of the IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), Hong Kong, China, pp. 1-8, March 2015.

[15] T. Hoang, D. Choi, T. Nguyen, "Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme", International Journal of Information Security, Volume 14, Issue 6, pp 549-560, 2015.

[16] M. Muaaz, C. Nickel, "Influence of different walking speeds and surfaces on accelerometer-based biometric gait recognition", In Proceedings of the 35th International Conference on Telecommunications and Signal Processing (TSP), Prague, Czech Republic; pp. 508-512, 2012.

[17] N. T. Trung, Y. Makihara, H. Nagahara, R. Sagawa, Y. Mukaigawa, Y. Yagi, "Phase registration in a gallery improving gait authentication", In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Washington, DC, USA, pp. 1-7, 2011.

[18] J. Frank, S. Mannor, J. Pineau, D. Precup, "Time Series Analysis Using Geometric Template Matching", IEEE Trans. Pattern Anal. Mach. Intell, 35, pp.740-754, 2013.

[19] J. Frank, S. Mannor, D. Precup, "Activity and Gait Recognition with Time-Delay Embeddings", In Proceedings of the AAAI, Atlanta, GA, USA, 11-15 July 2010.

[20] Y. Watanabe, "Influence of Holding Smart Phone for Acceleration-Based Gait Authentication", In Proceedings of the 2014 Fifth International Conference on Emerging Security Technologies (EST), Alcala de Henares, Spain, pp. 30-33, September 2014.

[21] Hoang, Thang, et al. "Adaptive Cross-Device Gait Recognition Using a Mobile Accelerometer." JIPS 9.2: 333, 2013.

[22] LeCun, Yann, et al. "Gradient-based learning applied to document recognition". Proceedings of the IEEE 86.11: pp. 2278-2324, 1998.

[23] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems, 2012.

[24] S. D. Bersch, D. Azzi, R. Khusainov, I. E. Achumba, and J. Ries, "Sensor data acquisition and processing parameters for human activity classification" Sensors, vol. 14, no. 3, pp. 4239-4270, 2014.

[25] The OU-ISIR Gait Database, Inertial Sensor Dataset [Online]. Available: http://www.am.sanken.osaka-u.ac.jp/BiometricDB/InertialGait.html [Accessed: 10- April- 2017].

[26] TensorFlow, An open-source software library for Machine Intelligence [Online]. Available: https://www.tensorflow.org/ [Accessed: 10- April- 2017]

[27] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

HUMAN IDENTIFICATION FROM INERTIAL DATA USING CONVOLUTIONAL NEURAL NETWORK

Hoang Van Ha, Tran Minh Triet

ABSTRACT: *Gait of each person is considered to be unique and can be used as a biometric characteristic to identify each individual. The strong growth of smartphones as well as wearable devices such as smartwatches, fitness trackers makes it easy to collect motion data through motion sensors which are built-in components in these devices.*

Many methods are proposed for identifying users from inertial data through the transformation of data into hand-crafted features. In this study, we propose a method for identifying users from motion data using the Convolutional Neural Network - a widely used and highly effective architecture in artificial intelligence and digital image processing - as an automatic high-level feature extractor. Results from experiments with the data of 496 persons from Osaka University's OU-ISIR Gait Database, which is considered as the largest inertial sensor-based gait database, show the high efficiency with an accuracy of over 99%.

Keywords*: Gait, Human Identification, Convolutional Neural Networks, Inertial Sensors, Accelerometer, Gyroscope.*