

Mô hình điều khiển drone tự phát hiện và bám theo hướng di chuyển của con người

■ **ThS. PHẠM TRUNG MINH; ThS. NGUYỄN HOÀNG THÙY TRANG**
Trường Đại học Hàng hải Việt Nam

TÓM TẮT: Trong lĩnh vực điều khiển drone tự động thực hiện hành động bay bám theo sự di chuyển của đối tượng được giám sát, sự thay đổi đột ngột hướng di chuyển của đối tượng có thể dễ dàng gây mất dấu, mất phương hướng cho drone trong việc xác định vị trí mới của đối tượng giám sát. Phương pháp sử dụng chuỗi khung hình biểu diễn các điểm đặc trưng kết hợp mạng nơ-ron hồi quy bộ nhớ ngắn dài có thể giải quyết vấn đề trên. Nội dung bài báo đề xuất mô hình tự động phát hiện hướng xoay của của con người, từ đó điều khiển hướng bay của drone một cách phù hợp nhằm nâng cao khả năng bám mục tiêu của drone.

TỪ KHÓA: Drone, chuỗi khung hình, mạng nơ-ron hồi quy bộ nhớ ngắn dài.

ABSTRACT: The sudden change in the direction of the subject's movement causes the surveillance drone to lose track of and orient the subject's position. This problem can be solved by combining frame sequences with Long Short Term memory neural network.

KEYWORDS: Drone, frame sequences, long short term memory.

1. ĐẶT VẤN ĐỀ

Trong các ứng dụng điều khiển drone tự động bám theo con người, thông qua hình ảnh thu được từ camera của drone, vị trí cơ thể hoặc khuôn mặt của con người được xác định, quá trình tính toán sự sai khác giữa vị trí của con người so với điểm trung tâm của camera. Giá trị sai khác vị trí được sử dụng để đưa ra các quyết định hành động bay phù hợp nhằm giúp drone thay đổi vị trí trong không gian, cố gắng luôn đưa vị trí cơ thể hoặc khuôn mặt của người vào vị trí trung tâm của khung hình camera.

Đã có nhiều công trình được công bố trong lĩnh vực này. Trong [1], nhóm tác giả sử dụng phương pháp phân lớp đặc trưng Haar Cascade để phát hiện vị trí tâm khuôn mặt con người, sau đó tính toán sự khác biệt trên các trục tọa độ x,y và z để xác định phương hướng bay của drone. Phương pháp này có độ nhạy và chính xác cao, tuy nhiên nếu khuôn mặt con người không ở hướng trực diện với camera của drone thì tính hiệu quả suy giảm rõ rệt. Thuật

toán Kanade-Lucas-Tomasi được sử dụng trong [2] đã giúp cải thiện vấn đề góc khuôn mặt con người không nhất thiết phải ở hướng trực diện với camera, nhưng thuật toán này bị giới hạn tốc độ di chuyển con người và góc nghiêng tối đa của khuôn mặt con người. Nguyên nhân của các điểm hạn chế này là do cơ chế xử lý từng khung hình. Tại một thời điểm, chỉ nhận dạng khuôn mặt người trong 1 khung hình, thiếu sự liên kết kết quả nhận dạng giữa các khung hình với nhau.

Từ vấn đề nhược điểm trên, nhóm tác giả đề xuất một mô hình nhận dạng các điểm đặc trưng khung xương của con người, đồng thời sử dụng mạng nơ-ron hồi quy bộ nhớ dài-ngắn LSTM (Long Short Term Memory networks) [3] để liên kết các kết quả xử lý một chuỗi khung hình liên tiếp. Từ đó, mô hình có thể xác được định hướng di chuyển của con người, cho phép tự động điều khiển drone thực hiện hành động bay một vòng cung giúp lấy lại vị trí của con người ở giữa khung camera của drone.

2. MÔ HÌNH KIẾN TRÚC HỆ THỐNG ĐIỀU KHIỂN TELLO DRONE

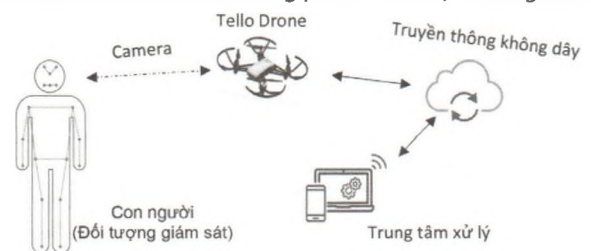
2.1. Mô hình kiến trúc hệ thống

Mô hình kiến trúc tổng quan của hệ thống điều khiển Tello Drone tự động bay theo hướng xoay của con người được chỉ ra trong Hình 2.1, bao các khối chính:

- *Con người:* Đây là đối tượng cần giám sát của Tello Drone. Trong khuôn khổ nghiên cứu này, đối tượng giám sát đứng thẳng, khi hoạt động sẽ quay sang trái hoặc sang phải với góc tối đa là 90°.

- *Khối Tello Drone:* Đây là thiết bị bay cần thực hiện các hành động bay sao cho hướng camera của Tello Drone luôn hướng vuông góc về phía trước mặt của con người.

- *Khối trung tâm xử lý:* Thu nhận hình ảnh từ camera của Tello Drone, đồng thời xử lý và dự đoán con người đã thực hiện việc quay sang trái hay bên phải, từ đó ra lệnh cho Tello Drone thực hiện hành động bay tương ứng với mục tiêu đảm bảo Tello Drone luôn ở hướng phía trước mặt con người.



Hình 2.1: Mô hình kiến trúc hệ thống

Với mô hình kiến trúc như trên, mục đích hoạt động hệ thống thực thi được công việc: thu nhận các hình ảnh từ camera của Tello Drone, tiến hành xác định được hành động của con người (bao gồm: quay sang trái, quay sang phải...). Với các kết quả dự đoán, lệnh điều khiển bay được truyền đi nhằm điều khiển Tello Drone thực hiện hành động bay một vòng cung sang bên trái hoặc bên phải nhằm bám sát khuôn mặt của con người. Quá trình trao đổi dữ liệu và tín hiệu điều khiển giữa Tello Drone và trung tâm xử lý thực hiện bằng phương pháp sử dụng giao thức UDP/IP, môi trường truyền thông không dây sử dụng sóng wifi.

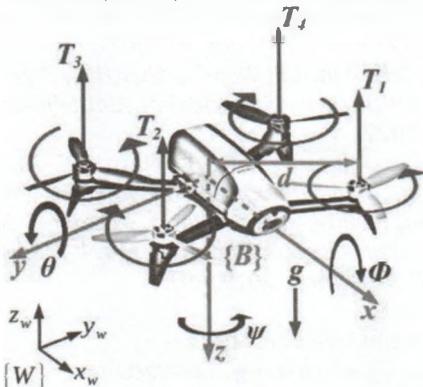
2.2. Thiết bị bay không người lái Tello Drone

Trong khuôn khổ đề tài nghiên cứu này, nhóm tác giả sử dụng Tello Drone do hãng Ryze chế tạo [4]. Đây là loại drone có cấu hình thấp, nhỏ gọn, kích thước (98*92,5*41)m, trọng lượng 87 g. Tello Drone được trang bị camera chuẩn HD720P30, có chỉ số phạm vi quan sát (Field of view- FOV) 82.6°. Nguồn năng lượng cung cấp cho Tello Drone là một pin Lionthium 1,1 Ah/3,8 V, cho phép hoạt động bay tối đa 13 phút với tốc độ lớn nhất là 8 m/s. Với các thông số kỹ thuật đã nêu, nhóm tác giả xác định Tello Drone phù hợp để nghiên cứu và thực nghiệm các thuật toán khoa học mà nhóm tác giả đề xuất.

Tello Drone có 4 động cơ cánh quạt rotor được lắp thẳng đứng ở 4 góc tạo thành dạng chữ thập. Các động cơ này hoạt động theo nguyên tắc: trong khi cặp 2 động cơ chéo nhau quay cùng chiều kim đồng hồ thì cặp 2 động cơ còn lại quay ngược chiều kim đồng hồ. Việc tăng giảm tốc độ quay của các cặp động cơ theo các trường hợp khác nhau một cách đồng thời sẽ cho phép thân Tello Drone xoay theo các trục X, Y và Z trong không gian, tương ứng lần lượt hành động: *Pitch* (gật), *Roll* (ngiên) và *Yaw* (hướng). Các phép xoay này sẽ tạo các hoạt động bay khác nhau cho Tello Drone, bao gồm: *hover* (bay lơ lửng), *forward* (bay tiến), *backward* (bay lùi), *left* (dịch trái), *right* (dịch phải), *rotate left* (xoay trái), *rotate right* (xoay phải), *up* (tăng độ cao), *down* (giảm độ cao), *take off* (cất cánh) và *landing* (hạ cánh).

3. HỆ THỐNG ĐIỀU KHIỂN TELLO DRONE TỰ ĐỘNG BAY BẮM THEO HƯỚNG HÀNH ĐỘNG QUAY CỦA CON NGƯỜI

3.1. Mô hình chuyển động bay của Tello Drone



Hình 3.1: Sự tương quan giữa hệ quy chiếu quán tính trái đất và hệ quy chiếu khung thân Tello Drone

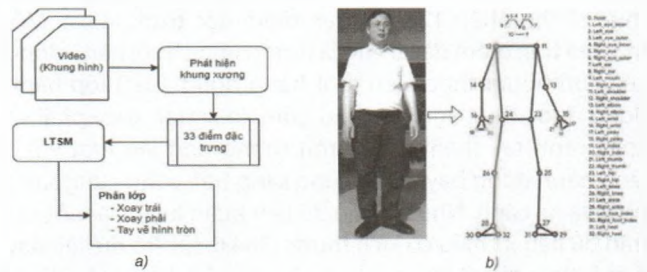
Hình 3.1 mô tả mô hình chuyển động bay của Tello Drone được xét trong hai hệ quy chiếu, bao gồm hệ quy chiếu quán tính của trái đất và hệ quy chiếu của khung thân Tello Drone $\{R_b\}$. $\{R_w\}$ định nghĩa bởi các trục tọa độ x_w, y_w và z_w trong đó z_w là trục hướng lên trên. Đối với $\{R_b\}$, gốc hệ tọa độ là trọng tâm của khung thân Tello Drone, trong đó trục x thể hiện hướng tiến về phía trước, trục z luôn vuông góc tương mặt phẳng chứa 4 động cơ rotor trong trường hợp Tello Drone ở trạng thái bay treo lơ lửng cân bằng trong không gian. Cặp động cơ rotor số 1 và số 3 nằm lần lượt ở phần âm và phần dương của trục x. Tương tự, cặp rotor số 2 và số 4 ở vị trí dương và âm trên trục y. Hệ phương trình góc Euler Z- X- Y được sử dụng nhằm mô hình hóa góc quay của khung thân Tello Drone, khi đó cần phải tham chiếu từ $\{R_b\}$ sang $\{R_w\}$. Trước hết, cần quay trục z theo một góc hướng Ψ (yaw), tiếp theo quay trục x theo một góc nghiêng Φ (roll), cuối cùng quay trục y theo một góc gạt θ (pitch). Hệ phương trình thể hiện sự định hướng của Tello Drone (biểu thị bằng ba góc Euler) được biểu diễn qua của ma trận xoay R như sau:

$$R = \begin{bmatrix} c\psi c\theta - s\psi s\theta & -c\psi s\theta & c\psi s\theta + c\theta s\psi \\ c\theta s\psi + c\psi s\theta & c\psi c\theta & s\psi s\theta - c\psi c\theta \\ -c\psi s\theta & s\theta & c\psi c\theta \end{bmatrix} \quad (1)$$

Trong đó: c và s - Lần lượt các phép lượng giác $\cos()$ và $\sin()$.

3.2. Mô hình phát hiện hướng quay của con người

Nhóm tác giả xây dựng kịch bản, triển khai hệ thống điều khiển Tello Drone luôn giữ vị trí bay ở phía thẳng trước mặt con người - đối tượng giám sát, khoảng cách là 2 m sao cho khuôn mặt con người nằm ở vị trí trung tâm camera của Tello Drone. Trong trường hợp con người di chuyển với tốc độ bình thường theo phương ngang hoặc tiến/lùi thì Tello Drone được hệ thống điều khiển hoạt động bay theo các phương tương ứng một cách phù hợp nhằm đảm bảo vị trí của Tello Drone luôn đảm bảo đúng quy định. Tuy nhiên, trong trường hợp con người xoay người sang bên trái hoặc sang bên phải một góc 90° sẽ xảy ra trường hợp Tello Drone cần phải thực hiện hành động bay một vòng cung $\pm 45^\circ$ tương ứng nhằm đảm bảo vị trí của Tello Drone theo yêu cầu.



Hình 3.2: Nguyên lý phát hiện hành động xoay của con người: a) - Tiến trình xử lý; b) - Các vị trí điểm đặc trưng của khung xương

Quá trình hành động xoay người sang bên trái hoặc bên phải được camera của Tello Drone thu nhận, vấn đề đặt ra là hệ thống cần tự động nhận dạng được hành động xoay đó của con người, cũng như trả lời được câu hỏi: đối

tượng được giám sát đã xoay sang bên trái hay bên phải? Hình 3.2a thể hiện tiến trình xử lý đã được nhóm tác giả đề xuất để giải quyết vấn đề này.

Tại trung tâm xử lý, khi nhận được luồng video truyền từ camera của Tello Drone, từng khung hình được khối phát hiện khung xương xử lý. Nếu trong khung hình có hình ảnh của con người, khối này sẽ xác định 33 vị trí tương ứng với 33 điểm đặc trưng [5] được biểu diễn trong Hình 3.2b. Đối với mỗi điểm sẽ bao gồm 3 tọa độ x , y và z , như vậy đối với mỗi một khung hình sẽ thu được được một ma trận như sau:

$$\begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ \dots & \dots & \dots \\ x_{32} & y_{32} & z_{32} \end{bmatrix} \quad (2)$$



Hình 3.3: Chuỗi khung hình các tư thế thể hiện hành động bước đi của con người

Hành động xoay người dưới góc độ hình ảnh video là một chuỗi các khung hình liên tiếp, trong mỗi khung hình là ma trận các điểm đặc trưng của khung xương tương ứng với tư thế của con người tại khung hình đó. Hình 3.3 là ví dụ thể hiện khi con người đang thực hiện việc bước đi thông thường, các khung hình thu nhận được thể hiện chuỗi tư thế khung xương tương ứng với từng khung hình. Nhóm tác giả đã lựa chọn đối với mỗi hành động xoay người sang bên trái hoặc bên phải sẽ tương ứng với 6 khung hình (tương ứng 0,25 giây). Như vậy, chuỗi hành động các tư thế tương ứng với hành động xoay người được thể hiện thành ma trận điểm đặc trưng 33x6:

$$\begin{bmatrix} [x_0, y_0, z_0] & [x_1, y_1, z_1] & \dots & [x_{32}, y_{32}, z_{32}] \\ [x_0, y_0, z_0] & [x_1, y_1, z_1] & \dots & [x_{32}, y_{32}, z_{32}] \\ \dots & \dots & \dots & \dots \\ [x_0, y_0, z_0] & [x_1, y_1, z_1] & \dots & [x_{32}, y_{32}, z_{32}] \end{bmatrix} \quad (3)$$

Ma trận trên được đưa tiếp vào mạng nơ-ron hồi quy bộ nhớ dài-ngắn LSTM(Long Short Term Memory networks) nhằm xác định sự phân lớp loại hành động tương ứng của con người đã thực hiện. Việc tạo bộ dữ liệu huấn luyện được thực hiện bằng cách thực nghiệm trên thực tế thu nhận 128 ma trận điểm đặc trưng 33x6. Mỗi mỗi ma trận điểm đặc trưng là tương ứng chuỗi hành động của con người thực hiện một hành động. Có 3 lớp hành động được thực nghiệm, bao gồm xoay trái, xoay phải và xoay cánh tay thành hình tròn tương ứng lần lượt với 3 lệnh hành động bay: vòng cung sang trái, vòng cung sang phải và hạ cánh. Như vậy, bộ dữ liệu huấn luyện là một ma trận dữ liệu 3 chiều có kích thước [384,6,33]. Bộ dữ liệu này được phân tập thành bộ huấn luyện và bộ kiểm tra theo tỉ lệ 0,2. Mô hình huấn luyện sử dụng mạng LTSM với các thông số $epochs=2000$, $batch_size=16$.

Trong quá trình Tello Drone hoạt động, cơ chế điều khiển hành động bay của drone được thực thi theo thuật toán dưới đây:

While Not Landing:

`Frame=Drone.ReadFrameFromCamera;`

`KeypointExtract=DetecSkeleton(Frame)`

`MatrixKeypoints.append(KeypointExtract)`

`MatrixKeypoints=MatrixKeypoint[-6]`

If len(MatrixKeypoints)==6:

`action_class=model.predic(MatrixKeypoints)`

Switch action_class.Label if action_class.accuracy>=99:

"Xoay trái": `Drone.Arc_Left`

"Xoay phải": `Drone.Arc_Right`

"Hình tròn": `Drone.Landing`

4. KẾT LUẬN

Nhóm tác giả đã đề xuất mô hình điều khiển drone dựa trên việc xử lý ma trận các điểm đặc trưng khung xương của con người. Ma trận này được xây dựng thông qua việc xử lý chuỗi 6 khung hình liên tiếp thu nhận từ camera của Tello Drone. Trong từng khung hình thực hiện việc phân tích xác định 33 điểm đặc trưng khung xương. Mạng nơ-ron hồi quy LTSM được áp dụng nhằm xác định lớp hành động quay trái, quay phải và xoay tay hình tròn của con người. Dựa trên kết quả lớp hành động đã thu nhận được các lệnh điều khiển hành động bay vòng cung sang trái, sang phải và hạ cánh được truyền tới Tello Drone. Mô hình này đã tăng cường khả năng bám đối tượng cần giám sát của drone, giảm thiểu các rủi ro mất dấu đối tượng trong quá trình theo dõi và bám theo của drone.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong Đề tài mã số DT21-22.57.

Tài liệu tham khảo

- [1]. A. S. Priambodo, F. Arifin, A. Nasuha and A. Winursito (2021), *Face Tracking for Flying Robot Quadcopter based on Haar Cascade Classifier and PID Controller*, J. Phys. Conf. Ser. vol.2111, no.1, doi: 10.1088/1742-6596/2111/1/012046.
- [2]. M. Karahan, H. Kurt, and C. Kasnakoglu (2020) *Autonomous Face Detection and Tracking Using Quadrotor UAV*, no.1, pp.7-10.
- [3]. F. Carrara, P. Elias, J. Sedmidubsky and P. Zezulka (2019), *LSTM-based real-time action detection and prediction in human motion streams*, *Multimed. Tools Appl.*, vol.78 no.19, pp.27309-27331, doi: 10.1007/s11042-019-07827-3.
- [4]. *Tello Official Website-Shenzhen Ryze Technology Co.,Ltd.* (<https://www.ryzerobotics.com/tello-edu>) accessed Mar. 28, 2022.
- [5]. C. L. Gugaresi, J. Tang, H. Nash... C. M.-T. W. on and U. (2019), *MediaPipe: A Framework for Perceiving and Processing Reality*, xr.cornell.edu, Accessed: Jan. 03, 2022 (https://xr.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf).

Ngày nhận bài: 28/6/2022

Ngày chấp nhận đăng: 17/7/2022

Người phản biện: TS. Phạm Việt Hưng

TS. Nguyễn Trọng Đức