ĐẠI HỌC QUỐC GIA HÀ NỘI

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

**BÀI LUẬN NGHIÊN CỨU CỦA NGHIÊN CỨU SINH**

Họ và tên thí sinh: **Quách Công Hoàng**

**Tên bài luận nghiên cứu:**

***NGHIÊN CỨU VÀ PHÁT TRIỂN HỆ THỐNG ĐỊNH VỊ VÀ XÂY DỰNG BẢN ĐỒ BA CHIỀU THỜI GIAN THỰC VỚI NGỮ NGHĨA CHO THIẾT BỊ BAY KHÔNG NGƯỜI LÁI SỬ DỤNG CÔNG NGHỆ HỌC MÁY***

Ngành: **Công nghệ kỹ thuật điện tử truyền thông**

Chuyên ngành: **Kỹ thuật điện tử**

Cán bộ nhận hướng dẫn Nghiên cứu sinh:

1. TS. Lê Vũ Hà
2. TS. Phùng Mạnh Dương
3. PGS.TS Quang Hà

Cán bộ hướng dẫn bài luận:

1. TS. Lê Vũ Hà
2. TS. Phùng Mạnh Dương
3. PGS.TS Quang Hà

**Hà Nội, 2020**

1. **Giới thiệu tổng quát:**

Thiết bị bay không người lái (Unmanned Aerial Vehicles - UAV) đang được kỳ vọng mang đến cuộc cách mạng trong hàng loạt lĩnh vực như vận chuyển và giao thông, giám sát hiện trường, sản xuất công nghiệp, quản lý nông nghiệp, hay thậm chí giải trí thực tại ảo. Các công nghệ cốt lõi để ứng dụng thành công UAV bao gồm thiết kế phần cứng, kỹ thuật điều khiển, giải pháp truyền thông, và công nghệ định vị và dẫn đường. Trong khi các công nghệ phần cứng, điều khiển và truyền thông đã tương đối hoàn thiện, công nghệ định vị và dẫn đường vẫn chủ yếu dựa trên tín hiệu định vị vệ tinh GPS, vốn có sai số lớn, cường độ tín hiệu nhỏ, dễ bị gây nhiễu và không ổn định với đa dạng địa hình và thời tiết. Để giải quyết vấn đề này, khả năng định vị và xây dựng bản đồ thời gian thực (Simultaneous Localization And Mapping - SLAM) với độ chính xác cao đóng vai trò quyết định bởi nó cho phép tổng hợp thông tin từ tất cả các cảm biến của thiết bị bay không người lái để tái tạo lại môi trường ba chiều xung quanh thiết bị bay và từ đó cho phép thiết bị bay không chỉ được định vị mà còn có thể tương tác và làm việc với môi trường hoạt động.

Xuất phát từ tầm quan trọng của bài toán nền tảng này, chúng tôi đề xuất trong nghiên cứu này giải pháp sử dụng kỹ thuật học sâu (Deep Learning) nhằm tổng hợp và phân tích ngữ nghĩa dữ liệu 3D được thu thập từ các cảm biến trên thiết bị bay không người lái để tạo thành một hệ thống SLAM. Tính bền vững của hệ thống SLAM đề xuất dựa trên hai tiêu chí sau: (1) có khả năng phân tích ngữ nghĩa từ dữ liệu cảm biến ảnh để nâng cao tính ổn định trước các biến động của môi trường; và (2) có khả năng biểu diễn dữ liệu bản đồ một cách hiệu quả cùng với mô hình tích hợp phù hợp cho thiết bị bay cỡ vừa và nhỏ. Đây cũng là hai vấn đề khoa học cơ bản trong nghiên cứu bài toán SLAM hiện tại.

1. **Sự cần thiết tiến hành nghiên cứu**
   1. **Ý tưởng khoa học:**

Thiết bị bay không người lái (UAV) đang trở thành một phần tất yếu của các hệ thống IoTs bởi khả năng cơ động và linh hoạt. Chúng có thể đảm nhận vai trò giám sát thu thập dữ liệu môi trường trong các bài toán thực tiễn như đô thị thông minh và nông nghiệp thông minh. Muốn phục vụ một cách hiệu quả và kinh tế các ứng dụng, một chiếc UAV cần phải đảm bảo khả năng di chuyển linh hoạt và an toàn. Do đó nhiệm vụ định vị và xây dựng bản đồ ba chiều cho thiết bị bay cỡ vừa và nhỏ là một trong những ưu tiên hàng đầu của các nhà khoa học.

Khác với các hệ thống xây dựng bản đồ viễn thám hiện nay, sử dụng các thiết bị bay tầm cao trên 500 mét, chúng tôi sử dụng các thiết bị bay không người lái giá rẻ, có trần bay thấp dưới 100 mét, hoạt động len lỏi giữa các tòa nhà. Trong điều kiện này các hệ thống định vị toàn cầu GPS thường hoạt động không ổn định, khả năng định vị phụ thuộc nhiều vào dữ liệu từ cảm biến ảnh và đơn vị đo lường quán tính (IMU). Hệ thống Visual-Inertial SLAM (VIO) bền vững khi hoạt động trong môi trường này cần đảm bảo hai tiêu chí sau: (1) có khả năng phân tích ngữ nghĩa của tín hiệu từ cảm biến ảnh để đảm bảo tính bền vững với biến động của môi trường, (2) có khả năng biểu diễn dữ liệu bản đồ một cách hiệu quả và có mô hình vận hành phù hợp với thiết bị bay cỡ vừa và nhỏ. Hai vấn đề trên là những bài toán cơ bản được các nhà khoa học đề ra trong giai đoạn hiện nay, tuy nhiên số lượng công bố theo hướng này không nhiều và các kết quả chưa triển khai thực tế trên thiết bị bay.

* 1. **Tình hình nghiên cứu trong và ngoài nước:**

1. ***Các vấn đề SLAM trong giai đoạn mới***

Bài toán SLAM đã có một chặng đường hơn 30 năm liên tục phát triển với nhiều dấu mốc đáng nhớ. Quá trình này được chia làm ba thời kỳ phát triển với những bước chuyển lớn trong quan điểm học thuật. Giới khoa học hiện nay cũng xác định hiện nay chúng ta đang ở ngưỡng cửa của giai đoạn thứ ba, hứa hẹn nhiều đột phá để phát triển một hệ thống SLAM có tính bền vững cao.

Giai đoạn 20 năm đầu tiên của SLAM, tính từ 1986 đến 2004, được gọi là giai đoạn cổ điển. Giai đoạn này cho ra đời công thức xác suất của SLAM cùng các lời giải bằng bộ lọc Kalman EKF, bộ lọc hạt, ước lượng khả năng tối đa v.v … Mối liên hệ giữa sự hiệu quả với và sự bền vững của hệ thống SLAM được phác họa trong giai đoạn này.

Giai đoạn thứ hai, tính từ 2005 đến 2015, được gọi là giai đoạn phân tích giải thuật *(algorithmic-analysis age)* [1]. Điểm nhấn của thời kì này là đi vào nghiên cứu những thuộc tính cơ bản của SLAM như tính khả kiến, tính hội tụ và tính nhất quán. Vai trò then chốt của biểu diễn thưa đối với tính hiệu quả của các bộ giải (solver) cho SLAM cũng được hiểu rõ; đồng thời các bộ mã nguồn mở cho SLAM cũng được ra đời trong giai đoạn này [2], [3].

Giai đoạn thứ ba được Cesar Cadena cùng đồng nghiệp định nghĩa là thời kì của nhận thức bền vững *(robust-perception age)* [1] với bốn tiêu chí nền tảng sau:

1. *Vận hành bền vững*: sai số thấp trong điều kiện môi trường thay đổi lớn, đồng thời tích hợp các khả năng sửa lỗi và tinh chỉnh sai số.
2. *Hiểu được ngữ nghĩa bậc cao*: hệ thống giải thuật phải vượt lên trên thông tin hình học thông thường, hiểu được các thông tin bậc cao hơn của môi trường.
3. *Nhận thức được tài nguyên*: hệ thống phải phù hợp với cảm biến và năng lực xử lý hiện tại; đồng thời cung cấp các phương thức để cải thiện hiệu năng tính toán dựa trên những tài nguyên sẵn có đó.
4. *Khả năng nhận thức dựa theo nhiệm vụ*: hệ thống SLAM có thể lựa chọn thông tin cảm nhận và bỏ qua các tín hiệu không liên quan. Khả năng này cho phép hỗ trợ nhiệm vụ của robot và cung cấp các loại bản đồ thích nghi, phù hợp với từng nhiệm vụ đang được thực hiện.

Các tiêu chí nêu trên đều hướng về nhu cầu thiết kế các phương tiện thông minh hoạt động trong môi trường phức tạp, nơi mà công nghệ định vị vệ tinh không còn hiệu quả như mong muốn [4].

1. ***Mô hình 3D SLAM hiện đại và ứng dụng trên thiết bị bay không người lái***

Trong giai đoạn hiện nay (tính từ 2016), các hướng nghiên cứu về SLAM là khá đa dạng. Để giảm thời gian đáp ứng thì việc áp dụng công nghệ học máy và các cảm biến chủ động (lidar, structure light) không phải là giải pháp triệt để do chi phí tính toán lớn và cảm biến tiêu thụ điện năng cao làm hao hụt thời lượng pin. Các thiết bị thông minh cỡ nhỏ như điện thoại di động, kính thực tế tăng cường (AR) hay UAV cỡ nhỏ hiện nay thường sử dụng kết hợp cảm biến đo lường quán tính (IMU) và các cảm biến thụ động để phát triển bài toán SLAM. Tiên phong trong cách tiếp cận này là nhóm nghiên cứu của Daniel Cremers thuộc Đại học Công nghệ Munich. Việc tính toán chuyển vị giữa các khung hình được ước lượng tối ưu bằng bộ lọc Kalman, dựa trên thông tin IMU và vi sai quang được lấy trực tiếp từ cảm biến ảnh mono [5]. Một hướng đi độc đáo và tiềm năng nữa là tích hợp Event-based camera [6] được nhiều nhóm bắt đầu nghiên cứu, đi đầu là nhóm của Davide Scaramuzza thuộc Đại học ETH. Các kết quả nghiên cứu hiện tại cho thấy phương pháp này cho thời gian đáp ứng khoảng 10 mili giây và có tiềm năng để tích hợp trên các thiết bị bay không người lái cỡ nhỏ [7]–[9]. Điều đáng tiếc là hai nhóm này thường đi sâu vào tối ưu mô hình tích hợp cảm biến, không áp dụng các nghiên cứu mới biểu diễn thưa vào hệ thống nên sẽ khó lòng mở rộng sang bài toán SLAM trong môi trường có kích thước lớn.

Thị giác máy và học sâu cho khả năng nhận nhận thức ngữ nghĩa bậc cao trong ảnh màu RGB. Dựa vào thông tin này cho phép phục hồi các thông tin bị thiếu hoặc ẩn từ môi trường như khoảng cách, véc-tơ pháp tuyến v.v… khiến hệ thống kiểu monocular SLAM trở nên bền vững hơn. K. Tateno là người đầu tiên công bố giải pháp này dưới cái tên CNN-SLAM [10] còn nghiên cứu mới nhất hiện nay thuộc về Code-SLAM [11] của nhóm nghiên cứu của A. Davison thuộc Đại học Imperial. Một cách tiếp cận khác là bổ sung khả năng phân đoạn ngữ nghĩa từ dữ liệu front-end với mục đích phân chia đồ thị surfel của SLAM một cách hợp lý hơn [12], [13]. Cách làm này kế thừa và bổ sung các hệ thống RGB-D SLAM vốn đã thành công trong giai đoạn thứ hai [14], [15]. Vấn đề lớn nhất của các nghiên cứu trên là tập trung nhiều vào môi trường trong nhà và phục vụ bài toán thực tế tăng cường, mặt khác mô hình xử lý đòi hỏi lượng tài nguyên tính toán và lưu trữ lớn. Vì vậy các nghiên cứu này cần phải phát triển thêm trước khi có thể áp dụng trên các phương tiện bay cỡ vừa và nhỏ.

Do tính hướng nhiệm vụ là một phần của giai đoạn *robust-perception*, sẽ là thiếu sót khi không đánh giá các nghiên cứu ứng dụng liên quan tới UAV mà 3D SLAM là một phần thiết yếu. Nitin J. Sanket thuộc Đại học Marryland đã sử dụng một số phương pháp 3D SLAM cho nghiên cứu điều khiển drone chui qua khe hợp hẹp của môi trường [16]. Ji Zhang, Đại học CMU, lại nghiên cứu điều khiển tối ưu quỹ đạo của thiết bị bay dựa trên hệ thống 3D SLAM bằng Lidar đặt trên UAV cỡ lớn dùng trong công nghiệp [17], [18]. Nhóm nghiên cứu của Vijay Kumar lại sử dụng thông tin ba chiều từ UAV và một số phương pháp trong thị giác máy tính để làm nhiệm vụ đếm hoa quả [19]. Một ứng dụng nông nghiệp khá thú vị khác là đánh giá sinh trưởng phát triển của cây trồng dựa trên ảnh đa mức phổ và 3D point-cloud thu thập từ thiết bị bay [20]. Sự kết hợp của UAV giá rẻ và 3D SLAM là hướng nghiên cứu mở có nhiều tiềm năng và có tính ứng dụng cao.

Tại Việt Nam, các nhóm nghiên cứu về bài toán SLAM và thiết bị bay UAV nằm rải rác ở một số viện nghiên cứu và trường đại học với kết quả đạt được vẫn còn hạn chế so với tốc độ phát triển nhanh chóng của thế giới. Do nhu cầu phát triển sản phẩm thực tế, các đơn vị như Viettel hay Viện hàn lâm khoa học Việt Nam đã bắt đầu nghiên cứu chế tạo các thiết bị bay UAV để tiếp cận từng phần của công nghệ. Một số mẫu thiết bị bay của Viettel, như VT-Swift và VT-Pigeon, đã được ứng dụng trong các nhiệm vụ cứu hộ cứu nạn và giám sát hiện trường. Tuy nhiên do các đề tài mang nặng tính thiết kế chế tạo phần cứng, lại tối ưu cho trần bay trên 1000 mét và phụ thuộc nhiều vào công nghệ định vị vệ tinh GPS. Vậy nên các sản phẩm này không dễ để có thể đưa sang các ứng dụng dân sự trong thời gian tới. Với sự phát triển của hạ tầng thông minh trong tương lai, chúng ta cần phải nghiên cứu UAV tầm thấp có khả năng 3D SLAM, xem chúng như là một thành phần tất yếu của hệ thống IoT.

1. **Phân tích ngữ nghĩa ảnh RGB-D cho bài toán SLAM trên UAV**

Cảm biến ảnh RGB-D giá rẻ đang trở nên thông dụng và thích hợp với các nghiên cứu thị giác máy và robotics. Một số đã được thiết kế nhỏ gọn, như Intel Realsense, và tích hợp lên UAV để thay thế cho các cảm biến truyền thống như siêu âm và hồng ngoại. Các khả năng như phát hiện và tránh vật cản, bám theo người của UAV là một trong những ứng dụng của phân tích ngữ nghĩa trên ảnh RGB-D.

Đối với hệ thống SLAM đề xuất, việc gán nhãn từng điểm ảnh là vấn đề chúng tôi quan tâm. Khả năng phân đoạn ngữ nghĩa *(semantic segmentation)* trên ảnh RGB-D cho phép phân loại các đối tượng di động trên bản đồ như người, động vật đồng thời xác định các landmark tin cậy cho hệ thống định vị phía sau. Đã có một số nghiên cứu tận dụng khả năng này cho bài toán SLAM [12], [13], tuy nhiên các mô hình mạng học sâu được sử dụng có chi phí tính toán lớn [21], lại không sử dụng trường thông tin khoảng cách nên không phù hợp với các thiết bị có tài nguyên tính toán hạn chế như UAV.

Nghiên cứu phân đoạn ngữ nghĩa trên ảnh hiện nay đều sử dụng các mô hình học sâu với kiến trúc mạng dạng encoder-decoder [22]. Các nghiên cứu có kết quả tốt nhất hiện nay đều sử dụng các phép nhân chập có tính chất thưa như *dilated convolution* [23] hay nhân chập trực tiếp các ten-xơ thưa (*sparse tensor* [24]) để trích xuất được nhiều thông tin hơn. Một số nghiên cứu đã rút gọn mô hình mạng xuống để có thể xử lý thời gian thực trên các thiết bị máy tính nhúng [25]. Tuy nhiên các nghiên cứu xử lý thông tin depth trên ảnh RGB-D là không nhiều, hơn nữa các kết quả hiện tại có thời gian xử lý lớn mà độ chính xác còn thấp [26], [27]. Đây là vấn đề cần phải nghiên cứu và cải thiện trong thời gian tới.

1. **Nhóm nghiên cứu của các nhà khoa học trong và ngoài nước**

* Nhóm nghiên cứu thị giác máy tính cho tự động hóa, Viện hệ thống thông minh Max Planck tại Tubingen, Đức.
* Phòng thí nghiệm robot Dyson, Đại học Imperial, London, Anh.
* Phòng thí nghiệm thị giác máy tính, Đại học Công nghệ Munich, Đức
* Nhóm nghiên cứu về robot có nhận thức (ailabRPG), Đại học ETH, Zurich, Thụy Sĩ
* Nhóm nghiên cứu về thị giác máy (V4RL), Đại học ETH, Zurich, Thụy Sĩ
* Nhóm nghiên cứu về robot có nhận thức (PRG), Đại học Maryland, Mỹ
* Phòng thí nghiệm về robot và tự động hóa GS. Hung Manh La (ARA Labs), Đại học Nevada Reno, Mỹ
* Nhóm nghiên cứu về robot và tự động hóa trong xây dựng của giáo sư GS. Quang Ha, Đại học Công nghệ Sydney, Sydney, Úc
* Nhóm nghiên cứu về robot trồng nho thông minh của TS. Mark Whitty, Đại học New South Wales, Sydney, Úc
* Nhóm nghiên cứu về robot bay, Đại học Hồng Kông, Trung Quốc
* Viện hàng không vũ trụ Viettel, Việt Nam

Bài toán nghiên cứu và ứng dụng SLAM có phạm vi rộng và đòi hỏi nhiều khía cạnh chuyên sâu. Vì vậy các nhóm nghiên cứu quốc tế đều cởi mở chia sẻ các kết quả đã đạt được, từ mã nguồn giải thuật cho đến công cụ mô phỏng và tập dữ liệu để đánh giá.

1. **Những kết quả nghiên cứu của chính chủ nhiệm đề tài đã đạt được theo hướng nghiên cứu của đề tài**

Dưới sự hướng dẫn và giúp đỡ của các thầy cô trong bộ môn Điện tử và kĩ thuật máy tính ĐH Công Nghệ do PGS.TS Trần Quang Vinh chủ nhiệm, tôi đã có một số công bố hội nghị quốc tế trong lĩnh vực robotics và xử lý ảnh trên nền tảng máy tính nhúng. Riêng nội dung liên quan tới thiết bị bay không người lái, tôi đã có một bài hội nghị và một bài tạp chí quốc tế [28]–[31].

Nhóm chúng tôi còn có những hoạt động hợp tác nghiên cứu với các đơn vị trong và ngoài nước. Năm 2014, nhóm chúng tôi dưới sự hỗ trợ của PGS.TS. Chử Đức Trình đã có dự án nghiên cứu phát triển thiết bị bay không người lái phục vụ ứng dụng xây dựng với công ty SIAMA, Úc. Dự án hợp tác trong nước tiêu biểu gần đây là project nghiên cứu cùng với SAMSUNG R&D Việt Nam năm 2015 do TS. Phạm Minh Triển chủ trì, liên quan tới lĩnh vực xử lý ảnh. Năm 2016, tôi có 6 tháng tham gia chương trình trao đổi nghiên cứu tại phòng thí nghiệm “Thị giác máy và hình học rời rạc” của GS. Sugimoto, Viện công nghệ thông tin quốc gia (NII), Tokyo, Nhật Bản. Chuyến đi này đã giúp tôi nắm bắt được các vấn đề cơ bản và quan trọng trong việc phát triển bài toán 3D SLAM trong giai đoạn hiện nay.

Về mặt cá nhân, trong thời gian công tác nghiên cứu tại trường, tôi có hướng dẫn sinh viên của trường tham gia cuộc thi về xe tự hành do FPT tổ chức trong hai mùa và đạt giải nhất năm 2018. Trong thời gian chuẩn bị cho hướng nghiên cứu này, tôi cũng đã tìm hiểu và làm chủ các công cụ nghiên cứu cho thị giác máy tính như PyTorch cùng quy trình rút gọn và tối ưu mạng neuron nhân tạo trên thiết bị nhúng của Nvidia. Bên cạnh đó, các công cụ sinh dữ liệu giả lập hỗ trợ cho nghiên cứu thị giác máy tính trên các phương tiện thông minh [32], [33] cũng được nhóm chúng tôi tiếp cận làm chủ.



Hình 1: Giao diện phần mềm phân tích dữ liệu 3D point-cloud thu thập được từ UAV mà nhóm đã thực hiện trong dự án hợp tác với SIAMA.

Với mong muốn đưa các kết quả nghiên cứu của mình ứng dụng vào thực tế, tôi cùng các bạn sinh viên đã dành nhiều thời gian tham gia tìm hiểu các bộ mã nguồn mở cho robotics là ROS và Dronecode (PX4). Cách thức này giúp chúng tôi nhanh chóng cập nhật thêm kiến thức mới nhất về robot thông minh và thiết bị bay không người lái, nâng cao chất lượng của thực nghiệm, điều này sẽ có lợi trong việc xuất bản các công bố quốc tế trong các hội nghị và tạp chí hàng đầu của ngành robotics.

1. **Những vấn đề chưa được giải quyết mà đề tài lựa chọn để nghiên cứu**

Dựa trên những đánh giá về tình hình trong nước và quốc tế, cùng những kết quả cá nhân đã thực hiện được trong thời gian qua, chúng tôi thấy rằng một hệ thống SLAM bền vững trong tương lai cần sự kết hợp với khả năng phân tích ngữ nghĩa bậc cao cho thông tin 3D point-cloud. Trong phạm vi đề tài nghiên cứu này là thiết bị bay, việc sử dụng tài nguyên tính toán SLAM một cách hiệu quả lại càng phải đề cao. Vậy nên chúng tôi đưa ra hai nhóm vấn đề chính cần được nghiên cứu:

* Vấn đề 1: mô hình tính toán định vị thời gian thực dựa trên việc bổ sung *ngữ nghĩa bậc cao của thị giác máy* để bù đắp khuyết điểm của các phương pháp cũ, vốn gặp tình trạng trôi tự do sóng định vị GPS bị nhiễu hay điều kiện anh sáng kém ổn định.
* Vấn đề 2: phương pháp xây dựng và tối ưu một cách hiệu quả bản đồ 3D point cloud trong không gian lớn.
* Vấn đề 3: xây dựng mô hình biểu diễn và lưu trữ bản đồ 3D để UAV hoạt động hiệu quả, cũng như dễ dàng phát triển các ứng dụng gia tăng như cảnh báo cháy, giám sát cơ sở hạ tầng đô thị.
  1. **Khả năng thành công (giải quyết hoặc cải thiện được vấn đề nghiên cứu).**

Với vấn đề định vị thời gian thực dựa trên thông tin thị giác, các nghiên cứu mới nhất hiện nay đều cho thấy khả năng phân tích đặc trưng hiệu quả của học sâu, đạt tốc độ lên tới 30fps trên máy tính nhúng [25], [34]. Các kết quả này có thể được áp dụng để cải thiện chất lượng của hệ thống định vị state-of-the-art hiện tại như Stereo-Omnidirectional-DSO.

Vấn đề thứ hai và thứ ba liên quan tới phương pháp xây dựng và tối ưu bản đồ 3D một cách hiệu quả. Các phương pháp sử dụng nền tảng toán học thông thường đã được nghiên cứu xong, trong khi đó các mô hình học sâu cho loại dữ liệu 3D point-cloud và ảnh RGB-D vẫn đang được phát triển. Chúng ta có thể tận dụng các nghiên cứu này làm nền tảng cho việc gán nhãn dữ liệu, để phát triển giải pháp cho hai vấn đề trên.

Mặt khác các công cụ mô phỏng dữ liệu phục vụ xử lý ảnh cũng như 3D SLAM đang được cộng đồng nghiên cứu phát triển, cho phép thực nghiệm kiểm chứng các mô hình mới trên bài toán SLAM. Nếu được đầu tư nghiên cứu bài bản và nghiêm túc thì đây lại là một cơ hội để có những công bố trong các tạp chí có uy tín cao.

1. **Nội dung và kế hoạch nghiên cứu**
   1. **Mục tiêu của nghiên cứu**

Trong đề tài này, mục tiêu nghiên cứu chính là xây dựng mô hình giải thuật 3D SLAM bền vững và phù hợp với ứng dụng trên thiết bị bay không người lái, cho phép khả năng vận hành thiết bị bay một cách an toàn trong điều kiện tín hiệu GPS không còn độ tin cậy cao. Đây là nền tảng vững chắc để phát triển các bài toán ứng dụng thiết bị bay UAV có ý nghĩa thực tiễn như cứu hộ cứu nạn, giám sát và đánh giá môi trường.

* 1. **Nội dung nghiên cứu**

Nội dung 1: đề xuất một mô hình ước lượng tối ưu kết hợp giữa phương pháp trực tiếp DSO [5] với thông tin ngữ nghĩa trích xuất từ hình ảnh RGB-D [12] bằng công nghệ học sâu. Cách làm này là mới do tín hiệu RGB-D thường dùng cho bài toán gán nhãn phân loại vật thể. Không có nghiên cứu về các bài toán định vị sử dụng thông tin ngữ nghĩa của ảnh RGB-D.

Nội dung 2: tối ưu bản đồ 3D dựa trên thông tin ngữ nghĩa đã được các tác giả đề cập trong một số hội nghị gần đây [35]. Tuy nhiên các nghiên cứu này mang nặng tính lý thuyết và không có đối tượng ứng dụng cụ thể. Bên cạnh đó, các lý thuyết về biểu diễn thưa và học sâu cũng đang có những chuyển biến cùng sự phát triển của nhiều công cụ tính toán mạnh mẽ [24], [36] hứa hẹn nhiều đột phá mới trong phân tích ngữ nghĩa của 3D point cloud. Đây là cơ sở tốt để cải thiện back-end của SLAM, giúp hệ thống có thể bù lại một cách hiệu quả những những biến dạng tích lũy theo thời gian trên mô hình bản đồ 3D toàn cục.

Các kết quả từ nghiên cứu này sẽ đóng góp trực tiếp các hiểu biết về thị giác máy tính, biểu diễn thưa vào hướng nghiên cứu 3D Visual SLAM. Mặt khác, xây dựng bản đồ ba chiều kết hợp phân tích ngữ nghĩa cũng là bước đệm để phát triển các nghiên cứu về UAV robotics như quy hoạch đường bay (path planning), né tránh vật cản (obstacles avoidance), điều khiển bầy đàn (swarm control).

* 1. **Cách tiếp cận, phương pháp nghiên cứu**

Các nội dung nghiên cứu 3D SLAM được chúng tôi phát triển dựa trên dữ liệu mô phỏng và thực nghiệm. Các tập dữ liệu thực tế cũng như mô phỏng phục vụ bài toán 3D SLAM đã được xây dựng tương đối nhiều trong những năm gần đây [37]–[40]. Các bộ dữ liệu này có kích thước lớn, tuy nhiên độ đa dạng chưa cao do phần lớn trong số đó là dữ liệu indoor, chưa bao quát hết môi trường hoạt động thực tế của UAV.

Để bù đắp cho việc thiếu hụt dữ liệu này, các công cụ mô phỏng dựa trên nền game engine như AirSim hay ESim được phát triển. Đây là các công cụ mạnh mẽ cho phép mô hình hóa hầu hết các loại cảm biến hình ảnh cho thiết bị bay hiện nay như ảnh RGB, Depth, Lidar và Event-Camera. Dữ liệu không chỉ được tự động gán nhãn mà còn được mô phỏng gần với các hiện tượng thường xảy ra trong thực tế như ngược sáng, motion blur v.v…

Việc sử dụng dữ liệu mô phỏng để đánh giá cũng là một phương pháp phổ biến đối với cộng đồng nghiên cứu giải thuật 3D SLAM hiện nay do dữ liệu môi trường rất đa dạng, dữ liệu kiểm chứng (ground-truth) được sinh tự động và không bị phụ thuộc vào các hệ thống định vị đắt tiền như Mocap. Sau khi đánh giá giải pháp trên hệ thống mô phỏng, bộ chính xác của hệ thống 3D SLAM đề xuất trong nghiên cứu có thể được đánh giá thực nghiệm trên thiết bị bay bằng bộ định vị RTK với độ tin cậy khoảng 2cm.

Cách sử dụng dữ liệu để huấn luyện trong hai nội dung đề tài cũng khác nhau:

* Nội dung 1: dữ liệu đầu vào là một khung truyền các dữ liệu cảm biến theo thời gian. Các mô hình học máy (kiểu *point cloud networks* hoặc *kết hợp 2D-3D*) sẽ phân đoạn các dữ liệu dạng 2.5D này trước khi đưa vào mô đun ước lượng trạng thái.
* Nội dung 2: với các point-cloud được xây dựng từ nội dung 1, chúng tôi dự định xây dựng phương án quy hoạch bản đồ 3D thành các đỉnh của đồ thị ngữ nghĩa. Cách làm này một mặt loại bỏ bớt các ràng buộc có độ tin cậy thấp như các ràng buộc được sinh ra từ đặc trưng SIFT, SURF hay ORB; mặt khác xây dựng các ràng buộc mới trong đồ thị pose của camera trước khi đưa vào bước tối ưu hóa.
  1. **Kế hoạch và kiến nghị**

1. **Kế hoạch thực hiện trong thời gian đào tạo**

Hoàn thành các tín chỉ học phần và chuyên đề.

* **Tháng 10 đến tháng 12 năm 2020:**
* Nghiên cứu tổng quan về công nghệ đo lường quán tính & thị giác (visual inertial odometry - VIO) hiện tại.
* Nghiên cứu tìm hiểu các công cụ mô phỏng cho thiết bị bay không người lái
* **Năm 2021:**
* Nghiên cứu xử lý ảnh, học máy và cách tối ưu thực thi trên máy tính nhúng.
* Đề xuất phương pháp trích xuất và phân đoạn ngữ nghĩa trên ảnh RGB-D để cải thiện độ chính xác của hệ đo lường quán tính & thị giác.
* Viết ≥ 01 bài báo hội nghị.
* **Năm 2022:**
* Nghiên cứu về các phương pháp học máy cho biểu diễn thưa.
* Đề xuất phương pháp tối ưu bản đồ 3D khi đóng vòng (loop-closure) dựa trên bản đồ ngữ nghĩa và học máy.
* Viết ≥ 01 bài báo hội nghị.
* **Năm 2023:**
* So sánh các phương pháp đã được đề xuất.
* Viết 01 bài đăng tạp chí.
* Viết luận án.
* Bảo vệ luận án.

1. **Dự kiến kế hoạch hoạt động khoa học sau khi hoàn thành chương trình đào tạo tiến sĩ**

Triển khai các kết quả nghiên cứu trên thiết bị bay không người lái. Phát triển các giải thuật điều khiển và tối ưu lập lịch cho bầy đàn. Tiến hành cho thiết bị bay thu thập dữ liệu trong khu vực nông nghiệp có địa hình đồi núi phức tạp, tạo thành cơ sở dữ liệu lớn cho các vùng canh tác.

1. **Kiến nghị**

Kính mong các thầy cô giáo trong Phòng thí nghiệm tín hiệu và xử lý thông tin, khoa Điện tử - Viễn thông hết sức tạo điều kiện giúp đỡ về mọi mặt và các trang thiết bị cần thiết cũng như hướng dẫn tôi về chuyên môn trong thời gian làm luận án.

Kính mong sự hướng dẫn tận tình của TS. Lê Vũ Hà.

*Hà Nội, ngày 7 tháng 07 năm 2020*

|  |  |
| --- | --- |
|  | **NGHIÊN CỨU SINH**  **Quách Công Hoàng** |

**TÀI LIỆU THAM KHẢO**

[1] C. Cadena *et al.*, “Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age,” *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Jun. 2016.

[2] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, “G2o: A general framework for graph optimization,” in *Proceedings - IEEE International Conference on Robotics and Automation*, 2011.

[3] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, “A Tutorial on Graph-Based SLAM,” *IEEE Intell. Transp. Syst. Mag.*, vol. 2, no. 4, pp. 31–43, 2010.

[4] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, “Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving,” *IEEE Trans. Intell. Veh.*, vol. 2, no. 3, pp. 194–220, 2017.

[5] J. Engel, V. Koltun, and D. Cremers, “Direct Sparse Odometry,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, 2018.

[6] G. Taverni *et al.*, “Dynamic Vision Sensors,” pp. 3–6, 2017.

[7] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, “Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 994–1001, Apr. 2018.

[8] G. Gallego, J. E. A. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza, “Event-Based, 6-DOF Camera Tracking from Photometric Depth Maps,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2402–2412, 2018.

[9] D. Scaramuzza and Z. Zhang, “Visual-Inertial Odometry of Aerial Robots,” in *Encyclopedia of Robotics, Springer, 2019*, 2019, pp. 1–13.

[10] K. Tateno, F. Tombari, I. Laina, and N. Navab, “CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6565–6574, 2017.

[11] M. Bloesch, J. Czarnowski, R. Clark, S. Leutenegger, and A. J. Davison, “CodeSLAM - Learning a Compact, Optimisable Representation for Dense Visual SLAM,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2560–2568.

[12] J. McCormac, A. Handa, A. Davison, and S. Leutenegger, “SemanticFusion: Dense 3D semantic mapping with convolutional neural networks,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 4628–4635.

[13] M. Runz, M. Buffier, and L. Agapito, “MaskFusion: Real-Time Recognition, Tracking and Reconstruction of Multiple Moving Objects,” *Proc. 2018 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2018*, pp. 10–20, 2019.

[14] R. A. Newcombe *et al.*, “KinectFusion: Real-time dense surface mapping and tracking,” in *2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011*, 2011.

[15] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger, “ElasticFusion: Real-time dense SLAM and light source estimation,” *Int. J. Rob. Res.*, vol. 35, no. 14, pp. 1697–1716, Dec. 2016.

[16] N. J. Sanket, C. D. Singh, K. Ganguly, C. Fermüller, and Y. Aloimonos, “GapFlyt: Active Vision Based Minimalist Structure-less Gap Detection For Quadrotor Flight,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2799–2806, Feb. 2018.

[17] J. Zhang and S. Singh, “Laser-visual-inertial odometry and mapping with high robustness and low drift,” *J. F. Robot.*, vol. 35, no. 8, pp. 1242–1264, Dec. 2018.

[18] J. Zhang, R. G. Chadha, V. Velivela, and S. Singh, “P-CAP: Pre-Computed Alternative Paths to Enable Aggressive Aerial Maneuvers in Cluttered Environments,” *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 8456–8463, 2018.

[19] X. Liu *et al.*, “Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1045–1052.

[20] R. Khanna, L. Schmid, A. Walter, J. Nieto, R. Siegwart, and F. Liebisch, “A spatio temporal spectral framework for plant stress phenotyping,” *Plant Methods*, vol. 15, no. 1, p. 13, Dec. 2019.

[21] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2980–2988, 2017.

[22] Y. Zhu *et al.*, “Improving Semantic Segmentation via Video Propagation and Label Relaxation,” Dec. 2018.

[23] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “DeepLabV3+,” *CVPR*, 2018.

[24] C. Choy, J. Gwak, and S. Savarese, “4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks,” Apr. 2019.

[25] S. Mehta, M. Rastegari, L. Shapiro, and H. Hajishirzi, “ESPNetv2: A Light-weight, Power Efficient, and General Purpose Convolutional Neural Network,” Nov. 2018.

[26] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[27] X. Qi, R. Liao, J. Jia, S. Fidler, and R. Urtasun, “3D Graph Neural Networks for RGBD Semantic Segmentation,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, vol. 2017-Octob, pp. 5209–5218.

[28] M. D. Phung, T. Van Thi Nguyen, C. H. Quach, and Q. V. Tran, “Development of a tele-guidance system with fuzzy-based secondary controller,” in *2010 11th International Conference on Control Automation Robotics & Vision*, 2010, pp. 1826–1830.

[29] M. D. Phung, C. H. Quach, D. T. Chu, N. Q. Nguyen, T. H. Dinh, and Q. P. Ha, “Automatic interpretation of unordered point cloud data for UAV navigation in construction,” in *2016 14th International Conference on Control, Automation, Robotics and Vision, ICARCV 2016*, 2017.

[30] M. D. Phung, C. H. Quach, T. H. Dinh, and Q. Ha, “Enhanced discrete particle swarm optimization path planning for UAV vision-based surface inspection,” *Autom. Constr.*, vol. 81, 2017.

[31] C. H. Quach, V. L. Tran, D. H. Nguyen, V. T. Nguyen, M. T. Pham, and M. D. Phung, “Real-time lane marker detection using template matching with RGB-D camera,” in *2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom)*, 2018, vol. 2018-Janua, pp. 152–157.

[32] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles,” pp. 1–14, 2017.

[33] H. Rebecq, D. Gehrig, and D. Scaramuzza, “ESIM: an Open Event Camera Simulator,” in *CoRL*, 2018.

[34] D. Detone, T. Malisiewicz, and A. Rabinovich, “SuperPoint: Self-supervised interest point detection and description,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2018-June, pp. 337–349, 2018.

[35] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, “Probabilistic data association for semantic SLAM,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, vol. 11, no. 8, pp. 1722–1729.

[36] M. Fey and J. E. Lenssen, “Fast Graph Representation Learning with PyTorch Geometric,” no. 1, pp. 1–8, Mar. 2019.

[37] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *Int. J. Rob. Res.*, 2013.

[38] J. McCormac, A. Handa, S. Leutenegger, and A. J. Davison, “SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation?,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[39] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, “ScanNet: Richly-annotated 3D reconstructions of indoor scenes,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.

[40] W. Li *et al.*, “InteriorNet: Mega-scale Multi-sensor Photo-realistic Indoor Scenes Dataset,” *Environ. Sci. Technol.*, vol. 45, no. 24, pp. 10354–10361, Sep. 2018.