Kỷ yếu Hội nghị Khoa học Quốc gia lần thứ IX "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR'9)"; Cần Thơ, ngày 4-5/8/2016
DOI: 10.15625/vap.2016.00012

ĐÁNH GIÁ HIỆU NĂNG GIẢI THUẬT FAB-MAP* ĐỊNH VỊ ROBOT TRONG NHÀ SỬ DỤNG THÔNG TIN HÌNH ẢNH

Nguyễn Quốc Hùng[12], Vũ Hải[1], Trần Thị Thanh Hải[1], Nguyễn Quang Hoan[3]

[1] Viện Nghiên cứu quốc tế MICA, Trường ĐHBK HN - CNRS/UMI - 2954 - INP Grenoble

[2] Trường Cao đẳng Y tế Thái Nguyên

[3] Trường Đại học Sư phạm Kỹ thuật Hưng Yên

{Quoc-Hung.NGUYEN, Thanh-Hai.TRAN, Hai.VU}@mica.edu.vn, quanghoanptit@yahoo.com.vn

TÓM TẮT— *Bài báo này trình bày tóm lược giải thuật FAB-MAP* định vị robot sử dụng thông tin hình ảnh trong môi trường trong nhà với ý tưởng chính là việc xác định vị trí robot bởi việc phép toán xác xuất có điều kiện giữa quan sát hiện tại với tập các quan sát mà robot di chuyển qua, các quan sát này được huấn luyện từ trước dựa vào đặc trưng phân loại cảnh và cây khung nhỏ nhất liên kết của các từ điển môi trường đồng xuất hiện. Từ đó quyết định vị trí chính xác trên bản đồ môi trường đã xây dựng từ trước. Trong bài báo này chúng tôi tập chung thực hiện đánh giá hiệu năng giải thuật FAB-MAP* trên CSDL được thu thập tại thư viện Tạ Quang Bửu (Việt Nam) và Milano-Bicocca (Italy). Kết quả cho thấy giải thuật định vị FAB-MAP* có tính khả thi trong bài toán định vị đối tượng trong nhà, làm cơ sở xây dựng các ứng dụng các bài toán SLAM cho robot trong tương lai.*

Từ khóa— *Giải thuật FAB-MAP*, Định vị hình ảnh, Robot.*

I. GIỚI THIỆU CHUNG

Dẫn đường robot được mô tả như một quá trình xác định đường đi hợp lý và an toàn từ điểm khởi đầu đến một điểm đích để robot có thể di chuyển giữa chúng. Rất nhiều cảm biến khác nhau như GPS, Lidar, Wifi đã được sử dụng để giải quyết bài toán này. Tuy nhiên những dữ liệu đó không luôn có sẵn hoặc thuận tiện cho việc thu thập, đặc biệt trong môi trường nhỏ hoặc vừa. Ví dụ, hệ thống định vị GPS chỉ cung cấp các dịch vụ bản đồ trong điều kiện khắt khe như thời tiết tốt, môi trường lớn, ngoài trời và không hỗ trợ trong môi trường nhỏ như trong các tòa nhà [1]. Các hệ thống sử dụng Lidar đòi hỏi phải đầu tư chi phí [2]. Hệ thống định vị Wifi cũng không dễ dàng triển khai, cài đặt, ngay cả đối với các môi trường diện hẹp [3].

Trong 30 năm trở lại đây, hướng tiếp cận dẫn hướng dựa trên hình ảnh đã thu hút sự quan tâm đặc biệt của các nhà nghiên cứu và phát triển. Các hướng tiếp cận dẫn hướng nói chung và sử dụng hình ảnh nói riêng có thể phân thành hai loại: loại thứ nhất dựa trên bản đồ môi trường đã được xây dựng từ trước bởi chính robot hoặc bởi một công cụ khác; loại thứ hai vừa định vị vừa tự xây dựng bản đồ (hay còn gọi chung là SLAM). Các phương pháp thuộc hướng tiếp cận thứ hai chỉ cho phép dẫn hướng cho robot nhưng không cho phép tìm đường đi giữa hai vị trí trong môi trường. Trong khi ngữ cảnh bài toán đặt ra là robot dẫn đường từ hai vị trí biết trước vì vậy phải có một bản đồ môi trường được xây dựng ở pha ngoại tuyến và được sử dụng ở pha trực tuyến để định vị và tìm đường.

Trong khuôn khổ bài báo này, chúng tôi đi theo hướng tiếp cận xây dựng bản đồ môi trường ở pha ngoại tuyến và định vị sử dụng nguồn thông tin hình ảnh thu thập được từ camera. Ưu điểm chính của việc sử dụng camera là giá thành rẻ hơn rất nhiều so với các cảm biến khác trong khi cung cấp nguồn thông tin hình ảnh có giá trị phục vụ cho nhiều bài toán khác nhau như xây dựng bản đồ, định vị và phát hiện vật cản bằng các thiết bị thông thường như camera cầm tay. Đặc biệt là các kết quả đánh giá giải thuật FAB-MAP* với các CSDL thu thập tại thư viện Tạ Quang Bửu (Việt Nam) và Milano-Bicocca (Italy) từ đó cho thấy được điểm mạnh của giải thuật được đề xuất.

Bài báo này được bố cục như sau: Phần I giới thiệu bài toán định vị cho robot. Phần II trình bày các nghiên cứu liên quan. Phần III tóm tắt hệ thống đề xuất, trong đó trình bày các bài toán liên quan đến robot dẫn đường. Phần IV đánh giá thử nghiệm trên CSDL thu nhận tại thư viện Tạ Quang Bửu (Việt Nam) và Milano-Bicocca (Italy). Phần V là kết luận và hướng phát triển trong tương lai.

II. MỘT SỐ NGHIÊN CỨU LIÊN QUAN

Như đã giới thiệu ở trên, bài toán định vị và xây dựng bản đồ môi trường cho robot có thể sử dụng rất nhiều loại cảm biến khác nhau. Tuy nhiên bài báo này đi theo hướng tiếp cận sử dụng camera. Vì vậy, chúng tôi chỉ tập trung trình bày những nghiên cứu liên quan về định vị và xây dựng bản đồ môi trường sử dụng thông tin hình ảnh.

A. Hướng tiếp cận chỉ sử dụng 01 camera

Broida 1990 [4], Broida và Chellappa 1991 [5] đề xuất thuật toán đệ quy tính toán sử dụng 01 camera thu thập một chuỗi hình ảnh của một đối tượng di chuyển để ước tính cấu trúc và động học của đối tượng sử dụng 01camera, đây có thể coi là nghiên cứu đầu tiên theo hướng tiếp cận Monocular SLAM. Việc thực hiện dự đoán vị trí được thực hiện kết hợp với bộ lọc lặp Kalman mở rộng (IEKF-Iterated Extended Kalman Filter) cho các điểm đặc trưng và chuyển động. Davison 2007 [6] là hệ thống định vị sử dụng 01 camera hoạt động trong thời gian thực, đây là hệ thống định vị tốt nhất chỉ sử dụng 01 camera. Trong phương pháp này, khung làm việc tổng quát bao gồm những vị trí của camera và một bản đồ 3D được đánh dấu rải rác các điểm quan trọng được tính toán đồng thời sử dụng một bộ lọc

Kalman mở rộng (EKF-Extended Kalman Filter). Và từ đây chiến lược EKF-SLAM được sử dụng rộng rãi và được cải thiện đáng kể trong các lĩnh vực khác nhau như William 2007 [7] giải quyết vấn đề tái định vị tự động, Clemente năm 2007 [8] lập bản đồ ứng dụng với quy mô lớn,...

Ethan và Drummond 2007 [9] đề xuất hệ thống định vị trong phạm vi nhỏ theo hướng tiếp cận 01 camera với quy mô thử nghiệm nhỏ được mô hình hóa thành các nút khác nhau kết hợp trong một đồ thị lớn bằng kỹ thuật tối ưu hóa phi tuyến tính. Dellaert 2006 [10], Strasdat năm 2010 [11] sử dụng kỹ thuật tối ưu hóa phi tuyến tính như BA (Bundle Adjustment) hoặc Smoothing và Mapping (SAM) được cấp trên về độ chính xác để lọc các phương pháp dựa trên và cho phép để theo dõi hàng trăm đặc trưng giữa khung hình liên tiếp. Williams 2008 [12] trình bày đánh giá so sánh vai trò của thủ tục vòng lặp đóng kín (loop closure detection) xem xét các vị trí đã đi qua hay chưa trong các hệ thống SLAM sử dụng 01 camera duy nhất (Monocular SLAM) đối với các bài toán định vị hình ảnh sử dụng các kỹ thuật đối sánh ảnh như sau: bản đồ với bản đồ (Map-to-Map) trong Clemente[8], ảnh với bản đồ (Image-to-Map) trong Williams [13], ảnh với ảnh (Image-to-Image) trong Cummins [14], [15]. Môi trường thử nghiệm được tiến hành trong nhà và ngoài trời với kịch bản một vòng và nhiều vòng quỹ đạo di chuyển. Đường ROC (Image-to-Image) là hoàn chỉnh nhất cho kết quả tốt nhất, ROC trong (Map-to-Map) cho thấy có ít điểm hơn tuy nhiên với các ngưỡng khác nhau thì độ chính xác giảm mạnh, ROC trong (Image-to-Map) cho kết quả chấp nhận được với các điểm rời rạc liên tục.

Perera 2011 [16] đề xuất thuật toán theo dõi giám sát các đối tượng chuyển động trong môi trường, vấn đề khó khăn của việc phát hiện các điểm di chuyển từ một camera chuyển động được giải quyết bởi các ràng buộc epipolar bằng cách sử dụng các thông tin đo lường đã có sẵn với các thuật toán monoSLAM. Tuy nhiên để xác định được ngưỡng chính xác để phân loại các điểm di chuyển được thực hiện thủ công qua nhiều lần thử nghiệm, tùy thuộc vào số lượng điểm dự kiến sẽ tương xứng về điểm phát hiện được trong môi trường.

B. Hướng tiếp cận sử dụng từ 02 camera trở lên

Ozawa 2005 [17] đề xuất hệ thống trực tuyến cho lập kế hoạch bước chân của robot bằng việc tái tạo ra một bản đồ 3D sử dụng kỹ thuật đo hành trình bằng thông tin thị giác (visual odometry). Hệ thống gồm hai thành chính: thứ nhất xây dựng lại bản đồ 3D từ một chuỗi hình ảnh thu thập từ camera-stereo để mô tả chi tiết thế giới thực và lập ra một kế hoạch di chuyển các bước chân của robot trên bản đồ 3D tái tạo; thứ hai phương pháp đo hành trình thị giác (visual odometry) sẽ kết nối với chuỗi hình ảnh 3D để có được mô hình chuyển động của camera theo hướng tiếp cận 6DOF (sáu bậc tự do) và thông tin môi trường 3D dày đặc. Để làm được điều này, một số kỹ thuật được áp dụng như tính toán chiều sâu của ảnh thu nhận, tính toán các luồng dữ liệu 3D dựa vào việc theo vết các hình ảnh đặc trưng ban đầu, sử dụng kỹ thuật RANSAC[18] ước lượng chuyển động camera theo mô hình 6DOF. Tiếp theo là sử dụng dữ liệu kết quả trên bản đồ 3D để thực hiện một trình tự tối ưu các địa điểm mà bước chân phải qua, các kế hoạch di chuyển của bước chân được cung cấp một bản đồ độ cao của địa hình và một tập hợp rời rạc dự đoán những bước chân tiếp theo mà robot có thể thực hiện.

Michel 2007 [19] trình bày một đề xuất theo vết robot chuyển động và leo cầu thang trong môi trường 3D. Bằng việc sử dụng chuỗi hình ảnh được thu thập từ camera-stereo trong việc phân tích và theo vết các mô hình của một đối tượng đã biết và có thể khôi phục lại các tư thế trên robot và định vị robot với các đối tượng đó. Hiệu năng hoạt động trong thời gian thực (Real-time) dựa vào tài nguyên tính toán của GPU (Graphic Processing Units) cho các nhận thức, cho phép gia tăng khả năng theo viết đối tượng trong các loại camera sử dụng khi điều hướng robot. Những hạn chế của phương pháp này phụ thuộc vào đối tượng 3D cần theo vết và mô hình 3D tương đối nhỏ, tuy nhiên nó rất hữu ích cho các kịch bản robot leo cầu thang.

Khác với các nghiên cứu trên, trong bài báo này chúng tôi tập trung vào giải quyết vấn đề định vị sử dụng thông tin hình ảnh bằng việc sử dụng một khung làm việc tổng quát về việc mô hình hóa môi trường và định vị được trình bày chi tiết ở phần dưới đây:

III. GIẢI THUẬT ĐỊNH VỊ SỬ DỤNG THÔNG TIN HÌNH ẢNH

Phần này mô tả khung làm việc để xây dựng hệ thống định vị trong các môi trường trong nhà. Khác với các hệ thống dẫn đường thông thường, tiện ích hệ thống đề xuất chỉ dữ liệu hình ảnh, mà không đòi hỏi dữ liệu định vị thông thường như GPS, WIFI, LIDAR,...

Trước tiên, chúng tôi sẽ trình bày ý tưởng của giải thuật FAB-MAP gốc, sau đó chúng tôi sẽ trình bày các cải thiện nâng cao độ chính xác khi thực hiện trong nhà gọi tắt là FAB-MAP*. Các chi tiết của hệ thống đề xuất của chúng tôi được mô tả trong [20]. Cải tiến lớn trong nghiên cứu này là, chúng tôi đã đề xuất khung làm việc tổng quát cho robot di động, nơi giải thuật định vị FAB-MAP* là một yếu tố quan trọng nhằm nâng cao độ chính xác định vị giúp cho robot hiểu được môi trường đưa ra các quyết định phù hợp khi thực hiện các nhiệm vụ trợ giúp dẫn đường.

A. Giải thuật FAB-MAP [14] (Fast Appearance Based - MAPping) gốc

Giải thuật FAB-MAP [14] do Cummins và đồng nghiệp đề xuất năm 2008 với ý tưởng xác định vị trí của camera bằng cách tính xác suất lớn nhất mà quan sát hiện tại tương ứng với một vị trí mà nó đã đi qua. Mỗi vị trí được biểu diễn bởi một véctơ nhị phân với các giá trị 0 hoặc 1 biểu thị sự vắng mặt hay xuất hiện của một từ không trong bộ từ điển đã được xây dựng từ trước bởi kỹ thuật túi từ (Bag of Word) [21].

Một ưu điểm nổi trội của FAB-MAP là sử dụng mô hình cây nhị phân Chow Liu [22] để tính toán mối quan hệ đồng xuất hiện của các từ, vì thế cho phép xác định vị trí một cách chính xác hơn. Bản chất của FAB-MAP là xác định vị trí của camera trong quá trình di chuyển. Nó cập nhật bản đồ nếu vị trí hiện tại trùng với một vị trí đã đi qua một thủ tục xác định vị trí đã đi qua loop closure detection) hoặc tạo vị trí mới. FAB-MAP gốc chỉ làm việc trên bản đồ cục bộ (tính từ thời điểm ban đầu đến vị trí hiện tại) chi tiết gồm có các bước sau:

Xây dựng bộ từ điển và biểu diễn quan sát: Mỗi khung hình sẽ được biểu diễn bởi mô hình túi từ [21]. Ý tưởng của mô hình này là coi bức ảnh như một tài liệu và biểu diễn tài liệu này bởi tập các từ (Words). Trong phần dưới đây trình bày 02 bước chính của kỹ thuật túi từ: i) xây dựng bộ từ điển; ii) biểu diễn ảnh dựa trên bộ từ điển.

- *Xây dựng bộ từ điển*: Giả sử có một tập các ảnh mẫu $I_1, I_2,...,I_K$. Các bước của xây dựng từ điển từ tập ảnh này như sau:
  - Trích chọn đặc trưng trên từng ảnh I_i, $_{(i = 1...k)}$. Do ưu điểm tính toán nhanh, chúng tôi chọn đặc trưng SURF [23], mỗi đặc trưng này là một véctơ 128 chiều.
  - Phân cụm các đặc trưng trong không gian đặc trưng sử dụng phương pháp k-Means [24]. Số lượng phân cụm ***K*** được định nghĩa, mỗi cụm sẽ được đặc trưng bởi tâm và độ rộng của 2 hướng chính.
- *Biểu diễn ảnh dựa trên bộ từ điển I(x, y) gồm các bước sau:*
  - Trích chọn đặc trưng SURF trong ảnh *I(x, y).*
  - Gán mỗi đặc trưng vào cụm mà khoảng cách từ nó đến tâm cụm là ngắn nhất.
  - Biểu diễn ảnh *I(x, y)* bởi một vector nhị phân $Z_k = \{z_1, z_2, ..., z_{|v|}\}$ có độ dài $|v|$, trong đó $|v|$ là số từ trong từ điển. Thành phần thứ z_q nhận một trong hai giá trị *{0, 1}* tương ứng với sự có mặt hay vắng mặt của một từ *q* trong ảnh.

Cây nhị phân Chow Liu [22]: Các tác giả trong FAB-MAP đề xuất thêm khái niệm "từ đồng xuất hiện" nhằm tạo mối liên kết giữa các từ trong bộ từ điển bằng việc xây dựng cấu trúc cây nhị phân Chow Liu [22], thực chất đây là một cây khung nhỏ nhất trong bài toán đồ thị có hướng nhằm tạo ra mối liên hệ giữa các nút. Do vậy, sau khi đã xây dựng từ điển gồm **K** từ, các từ có thể có mỗi liên hệ với nhau.

Cụ thể có những cặp từ xuất hiện đồng thời. Để tính đến mối quan hệ này, xác suất *P(Z)* với $Z = (z_1, z_2,...,z_K)$ của ***K*** biến rời rạc cần phải xác định. Nếu như *P(Z)* là một phân bố tổng quát không có cấu trúc đặc biệt, không gian cần thiết để biểu diễn cấu trúc này là lũy thừa bậc ***K***. Để đơn giản, các nhà khoa học thường xấp xỉ *P(Z)* bởi một cấu trúc *Q(Z)* có cấu trúc đặc biệt gần giống với phân bố *P(Z)*. Cụ thể tối thiểu khoảng cách Kullback-Leibler:

$$D_{KL}(P,Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \tag{1}$$

Ý tưởng chung là xấp xỉ một phân bố rời rạc *P(Z)* bởi mạng Bayes có dạng cấu trúc cây $Q(Z)_{opt}$. Cấu trúc của $Q(Z)_{opt}$ được xác định bằng cách xem xét một đồ thị **G**. Đối với một phân bố trên n biến, ***G*** là đồ thị đầy đủ với *n* nút và $\frac{n(n-1)}{n}$ cạnh, trong đó mỗi cạnh (z_i, z_j) có trọng số là thông tin tương hỗ $I(z_i,z_j)$ giữa biến *i* và *j* xác định bởi công thức:

$$I(z_i, z_j) = \sum_{z_i, z_j \in \Omega} P(z_i, z_j) \log \frac{p(z_i, z_j)}{p(z_i)p(z_j)} \tag{2}$$

Chow Liu đã chứng minh được cây khung có trọng số lớn nhất của đồ thị ***G*** sẽ có cấu trúc tương tự với $Q(Z)_{opt}$.

Cập nhật vị trí đã đi qua hoặc tạo mới trên bản đồ: Giả sử tại thời điểm k, bản đồ môi trường đã được xây dựng gồm n_k vị trí: $L_{nk} = \{L_1,L_2,...,L_{nk}\}$. Camera thu nhận khung hình I_k.

Sử dụng bộ từ điển đã được xây dựng, biểu diễn khung hình I_k bởi véctơ Z_k như đã trình bày ở phần trên. Gọi Z_k là véctơ quan sát từ lúc bắt đầu đến thời điểm *k*: $Z_k = \{Z_1,Z_2,...,Z_k\}$, Tính xác suất mà quan sát I_k có thể ở một trong số các vị trí $L_{n_k} = \{L_1, L_2, ... L_{n_k}\}$ theo công thức dưới đây:

$$P(L_i|Z^k) = \frac{p(Z_k|L_i, Z^{k-1})p(L_i|Z^{k-1})}{p(Z_k|Z^{k-1})} \tag{3}$$

Trong đó: $i = \overline{1, n_k}$; $p(L_i|Z^k)$ tập quan sát hiện thời; $p(Z_k|L_i)$ khả năng quan sát; $p(L_i|Z^{k-1})$ tập quan sát trước; $p(Z_k|Z^{k-1})$ toàn bộ quan sát tới vị trí thứ *k*.

Dựa trên giải thuật FAB-MAP gốc, chúng tôi đề xuất một số cải tiến chính như sau:

- *Đề xuất kỹ thuật xác định cảnh phân biệt để giảm các quan sát trùng lặp.*
- *Chuyển pha trực tuyến của FAB-MAP về hoạt động ngoại tuyến nhằm xây dựng các vị trí quan trọng (đánh dấu) trên hành trình của robot.*
- *Định vị vị trí robot trên bản đồ môi trường đã định nghĩa trước các vị trí quan trọng.*

Các cải tiến này được trình bày chi tiết trong các phần dưới đây. Chúng tôi đặt tên giải thuật định vị robot cải tiến này là **FAB-MAP***.

B. Giải thuật FAB-MAP xây dựng cơ sở dữ liệu vị trí đặc tả môi trường*

Trong nghiên cứu này, chúng tôi dựa trên ý tưởng của FAB-MAP để tính xác suất mà quan sát hiện tại của robot trùng với một quan sát tại vị trí nào đó đã được huấn luyện trong CSDL. Giải thuật FAB-MAP* làm việc trên toàn bộ các vị trí của bản đồ tổng thể đã được xây dựng từ trước. Để xây dựng bộ từ điển, FAB-MAP* sử dụng toàn bộ số khung hình thu nhận được để huấn luyện. Tuy nhiên với môi trường trong nhà, các khung cảnh thường lặp đi lặp lại.

Để loại bỏ tính lặp của các mẫu, chúng tôi đề xuất chỉ sử dụng các khung cảnh phân biệt, các khung cảnh này được lựa chọn bằng cách sử dụng khoảng cách euclid của hai véctơ đặc trưng GIST [25] trích chọn từ hai ảnh liên tiếp. Cách làm này cho phép giảm thiểu các từ bị lặp, từ đó tăng hiệu năng (độ chính xác và độ triệu hồi) của giải thuật định vị.

Xác định cảnh phân biệt để giảm các quan sát trùng lặp: Bài toán xác định khung cảnh phân biệt được mô tả như sau: Giả thiết có một chuỗi N khung hình liên tiếp $I = \{I_1,I_2,...,I_N\}$. Xác định tập con của $I_d \in I$ với $I_d = \{I_{i1},I_{i2},...,I_{id}\}$ trong đó các khung cảnh I_{ij} là phân biệt. Để xác định I_{ij} với I_{ik} là phân biệt, có thể kiểm chứng bằng hàm khoảng cách $D(I_{ij},I_{ik})$:

$$D(I_{ij}, I_{ik}) = \boldsymbol{ED}\ (Gist(I_{ij}), Gist(I_{ik})) \quad (4)$$

Trong đó: ***ED*** là khoảng cách giữa hai véctơ trong không gian. Việc xác định các khung cảnh riêng biệt thực hiện bằng giải thuật sau đây. Đầu vào là một chuỗi các hình ảnh liên tiếp thu thập từ camera: $I = \{I_1,I_2...I_N\}$ gồm các bước:

- *Bước 1: Tính toán sai khác giữa hai khung hình liên tiếp I_i và I_{i-1}: Sai khác này được định nghĩa là khoảng cách Euclid D_i giữa hai véctơ đặc trưng GIST tương ứng F_i,F_{i-1}.*
- *Bước 2: Kiểm tra nếu $D_i > \theta_{Gist}$ thì I_i được lựa chọn là khung hình phân biệt, trong đó θ_{Gist} là ngưỡng xác định trước bằng thực nghiệm quyết định số lượng khung hình giữ lại.*

1. Chuyển pha trực tuyến của FAB-MAP về hoạt động ngoại tuyến: Sau khi đã xác định các cảnh phân biệt, các ảnh này được đưa vào pha ngoại tuyến để xây dựng từ điển và cây Chow Liu trước khi đưa vào pha trực tuyến của FAB-MAP gốc tạo ra các vị trí trên bản đồ. So với FAB-MAP, đầu vào của FAB-MAP* là tập các ảnh đã thu thập từ trước trên hành trình khai phá đường đi trong môi trường.

Công việc huấn luyện này có thể chạy một lần hoặc nhiều lần với các dữ liệu đường đi khác nhau để làm giàu số vị trí trên bản đồ. Một thủ tục lặp có tên *"**Loop Closure Detection**"* có nhiệm vụ đánh chỉ mục cho các vị trí mới phát hiện trùng khớp với vị trí trên bản đồ, được thực hiện liên tục và kết thúc khi không còn phát hiện các vị trí mới.

2. Định vị vị trí robot trên bản đồ môi trường: Sau khi đã xây dựng bản đồ topo số liệu ở pha ngoại tuyến, việc định vị ở pha trực tuyến. Ở pha trực tuyến, camera thu nhận ảnh I_k, quan sát từ đầu đến thời điểm k là Z_k như định nghĩa ở phần trên. Thực hiện tính xác suất mà quan sát Z_k có thể ở vị trí L_i trên bản đồ $L_N = \{L_1,L_2,...,L_N\}$ với mọi giá trị $i = \overline{1,N}$, trong đó N là tổng số vị trí đã học trong môi trường xác định bằng công thức:

$$P(L_i|Z^k) = \frac{p(Z_k|L_i, Z^N)p(L_i|Z^N)}{p(Z_k|Z^N)} \quad (5)$$

So với công thức 3, công thức này khác ở chỗ Z_N được thay bởi Z_{k-1} do lúc này bản đồ toàn bộ môi trường đã được xây dựng. Chúng tôi tiến hành đánh giá quan sát hiện tại tại vị trí L_i trên bản đồ của một xác suất khi được đưa ra các quan sát tất cả lên đến một vị trí k. Z_k chứa toàn bộ các từ xuất hiện toàn bộ quan sát tới vị trí thứ $k-1$; Z_k tập các từ tại vị trí thứ k.

Trong hệ thống này, tại vị trí thứ i xác định một tham số $k*$ là ngưỡng $argmax(p(Z_k|L_i))$ là đủ lớn (ngưỡng này được xác định trước từ thực nghiệm $\theta_{FAB-MAP*} = 0.4$).

IV. KẾT QUẢ ĐÁNH GIÁ THỰC NGHIỆM

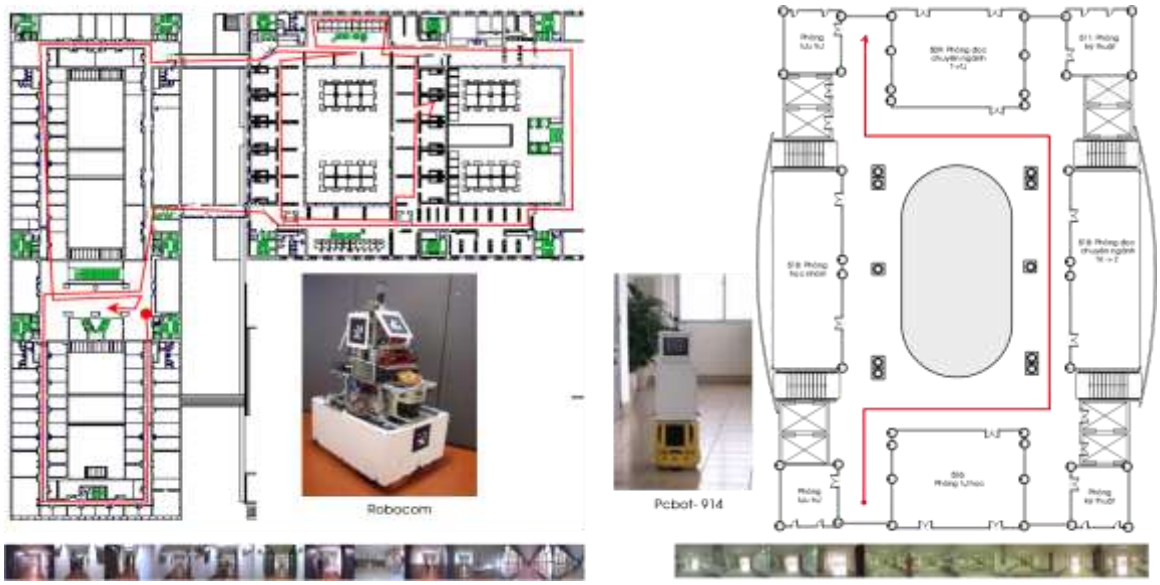
A. Thu thập dữ liệu

Chúng tôi đề xuất tiến hành đánh giá tại 02 môi trường khác nhau: *(i) thư viện Bicocca (Italy) và (ii) thư viện Tạ Quang Bửu (Việt Nam)* được kết quả chi tiết có trong bảng dưới đây:

Bảng 1. Dữ liệu huấn luyện và đánh giá tại 02 môi trường

Môi trường thực nghiệm	**Huấn luyện** *(Ảnh)*	**Ảnh thử nghiệm** *(Ảnh)*
E1: Thư viện Bicocca	41 195	44 195
E2: Thư viện Tạ Quang Bửu	10 650	10 175

Đường đi của robot (Robocom và Pcbot-914) thu thập từ 02 môi trường minh họa ở hình 1 dưới đây:



a) Môi trường E1: Thư viện Bicocca (Italy) *b) Môi trường E2: Thư viện Tạ Quang Bửu (Việt Nam)*

Hình 1. Môi trường thử nghiệm giải thuật định vị FAB-MAP*

B. Kết quả đánh giá

Đối với giải thuật định vị, cần đánh giá khả năng định vị đúng một vị trí trên bản đồ với một quan sát đưa vào I_k nào đó. Trong số **N** vị trí trên bản đồ đã xây dựng, giả sử L_k^* là vị trí có $P(L_{k*}|Z^k)$ lớn nhất. Để đánh giá khả năng định vị, chúng tôi sử dụng độ triệu hồi *R(Recall)* và độ chính xác *P(Precision)* có trong [26] được tính toán bằng công thức sau:

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

Trong đó:

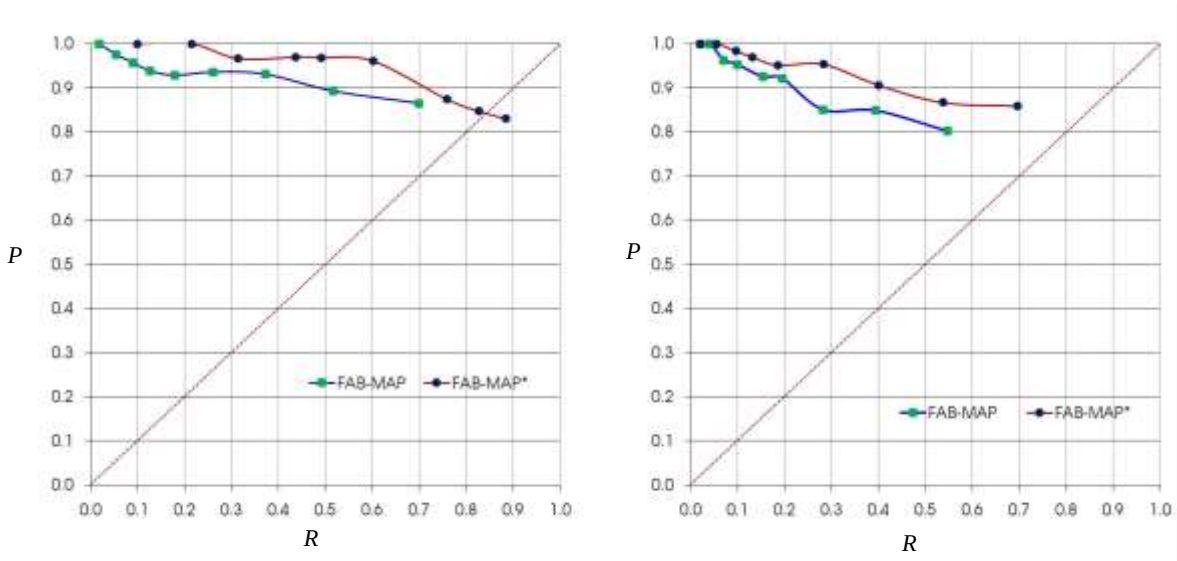
- *Vị trí không nhận dạng (ký hiệu FN): Nếu $P(L_{k*}|Z^k) < \theta_{FAB\_MAP}$ kết luận đây là vị trí không có trên bản đồ.*
- *Vị trí đúng (ký hiệu TP): Nếu $P(L_{k*}|Z^k) > \theta_{FAB\_MAP}$ và đo khoảng cách giữa L_{k*} và vị trí thực trên thực địa, nếu khoảng cách này nhỏ hơn một giá trị ngưỡng cho trước (trong thực nghiệm $\varepsilon = 0.4m$), khi đó kết luận L_{k*} là định vị đúng trên bản đồ.*
- *Vị trí sai (ký hiệu FP): Nếu $P(L_{k*}|Z^k) > \theta_{FAB\_MAP}$ và khoảng cách giữa L_{k*} và vị trí thực trên thực địa lớn hơn ngưỡng ε khi đó kết luận L_{k*} là định vị sai trên bản đồ.*

Bảng 2 trình bày chi tiết kết quả định vị giải thuật FAB-MAP*. Có thể nhận thấy trong mọi trường hợp của θ_{FAB_MAP}, khi sử dụng đặc trưng GIST trong việc phân loại cảnh luôn cho kết quả định vị tốt hơn. Kết quả này minh chứng cho việc đề xuất sử dụng đặc trưng GIST để phân tách khung cảnh có cấu trúc lặp, giống nhau trong môi trường trong nhà.

Bảng 2. Dữ liệu huấn luyện và đánh giá tại 02 môi trường

θ_{FAB_MAP}	**Thư viện Bicocca (E1)**				**Thư viện Tạ Quang Bửu (E2)**			
	FAB-MAP		**FAB-MAP***		**FAB-MAP**		**FAB-MAP***	
	Recall	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*
0.9	1.79%	100.00%	9.94%	100.00%	4.13%	98.24%	8.29%	100%
0.8	5.52%	97.56%	21.55%	100.00%	6.90%	97.41%	15.47%	100%
0.7	9.10%	95.65%	31.49%	96.61%	8.31%	96.85%	24.86%	100%
0.6	12.69%	93.88%	43.65%	96.94%	10.46%	95.11%	33.15%	98.24%
0.5	17.93%	92.86%	49.17%	96.84%	12.42%	94.07%	40.88%	98.37%
0.4	**26.07%**	**93.56%**	**60.22%**	**96.08%**	**20.81%**	**90.34%**	**53.59%**	**95.98%**
0.3	37.24%	93.10%	76.01%	87.41%	37.25%	88.22%	74.03%	94.03%
0.2	51.72%	89.29%	82.68%	84.76%	42.55%	86.96%	81.22%	93.04%
0.1	69.93%	86.51%	88.45%	83.04%	56.97%	85.71%	89.50%	92.98%

Hình 2 biểu diễn kết quả độ triệu hồi và độ chính xác với tập ngưỡng $\theta_{FAB\_MAP} \in \{0.1,...,0.9\}$. Kết quả này cho thấy khi đó ngưỡng θ_{FAB_MAP} tăng dần (ràng buộc càng chặt) thì độ triệu hồi giảm nhanh và độ chính xác tăng và ngược lại.



a) Môi trường E1: Thư viện Bicocca (Italy) *b) Môi trường E2: Thư viện Tạ Quang Bửu (Việt Nam)*

Hình 2. Biểu đồ so sánh giải thuật định vị FAB-MAP* tại E1 và E2

Hình dưới đây là một số hình ảnh minh họa đánh giá với ngưỡng $\theta_{FAB\_MAP}=0.4$ đạt kết quả cao trên 02 CSDL thu thập:

a) Môi trường E1: Thư viện Bicocca (Italy)

b) Môi trường E2: Thư viện Tạ Quang Bửu (Việt Nam)

Hình 3. Một số hình ảnh minh họa định vị robot trên 02 môi trường thử nghiệm

V. KẾT LUẬN

Trong bài báo này, chúng tôi đã tóm lược mô hình kết hợp định vị sử dụng thông tin hình ảnh đối với hai bài toán truyền thống là xây dựng bản đồ môi trường và định vị vị trí. Xây dựng bản đồ môi trường trong nhà bằng việc tạo ra các điểm đánh dấu trong môi trường đơn giản và nhanh chóng nhằm làm tăng độ chính xác của bản đồ môi trường được xây dựng. Biểu diễn tại các vị trí quan trọng trên bản đồ môi trường bằng mô hình xác xuất có điều kiện giữa quan sát hiện thời với tập các quan sát từ trước tới thời điểm hiện tại bằng giải thuật định vị FAB-MAP kết hợp đề xuất sử dụng đặc trưng GIST trong việc phân tách khung cảnh giống nhau (gọi tắt là FAB-MAP*). Thực hiện đánh giá giải thuật FAB-MAP* trên một số CSDL lớn trên thế giới, kết quả cho thấy giải thuật đề ra đáng tin cậy, áp dụng cho các bài toán định vị robot trong môi trường nhỏ hẹp.

VI. LỜI CẢM ƠN

Cảm ơn đề tài "*Trợ giúp định hướng người khiếm thị sử dụng công nghệ đa phương thức*" mã số: **ZEIN2012RIP19** - Hợp tác quốc tế các trường Đại học tại Việt - Bỉ (VLIR) đã hỗ trợ trong quá trình thực hiện bài báo này.

TÀI LIỆU THAM KHẢO

[1] E. North, J. Georgy, M. Tarbouchi, U. Iqbal, and A. Noureldin, "*Enhanced mobile robot outdoor localization using ins/gps integration*" in International Conference on Computer Engineering and Systems, 2009, pp. 127–132.

[2] X. Yuan, C.-X. Zhao, and Z.-M. Tang, "Lidar scan-matching for mobile robot localization," Information Technology Journal, vol. 9, no. 1, pp. 27–33, 2010.

[3] J. Biswas and M. Veloso, *"Wifi localization and navigation for autonomous indoor mobile robots"* in International Conference on Robotics and Automation (ICRA), 2010, pp. 4379–4384.

[4] T. J. Broida, S. Chandrashekhar, and R. Chellappa, *"Recursive 3-d motion estimation from a monocular image sequence"*, IEEE Transactions on Aerospace and Electronic Systems, vol. 26, no. 4, pp. 639–656, 1990.

[5] T. Broida and R. Chellappa, *"Estimating the kinematics and structure of a rigid object from a sequence of monocular images"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pp. 497–513, 1991.

[6] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, *"Monoslam: Real-time single camera SLAM"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 1052–1067, 2007.

[7] B. Williams, G. Klein, and I. Reid, *"Real-time slam relocalisation"*, in Computer Vision, IEEE 11th International Conference on ICCV. IEEE, 2007, pp. 1–8.

[8] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós, *"Mapping large loops with a single hand-held camera"* in Robotics: Science and Systems, vol. 2, 2007, p. 11.

[9] E. Eade and T. Drummond, *"Monocular slam as a graph of coalesced observations"* in Computer Vision, IEEE 11th International Conference on ICCV, IEEE, 2007, pp. 1–8.

[10] F. Dellaert and M. Kaess, *"Square root sam: Simultaneous localization and mapping via square root information smoothing"* The International Journal of Robotics Research, vol. 25, no. 12, pp. 1181–1203, 2006.

[11] H. Strasdat, J. Montiel, and A. J. Davison, *"Scale drift-aware large scale monocular SLAM"* in Robotics: Science and Systems, vol. 2, no. 3, 2010, p. 5.

[12] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós, *"A comparison of loop closing techniques in monocular slam"* Robotics and Autonomous Systems, vol. 57, no. 12, pp. 1188–1197, 2009.

[13] ——, *"An image-to-map loop closing method for monocular SLAM"* in Intelligent Robots and Systems, International Conference on IEEE/RSJ IEEE, 2008, pp. 2053–2059.

[14] M. Cummins and P. Newman, *"Fab-map: Probabilistic localization and mapping in the space of appearance"*, The International Journal of Robotics Research, vol. 27, no. 6, pp. 647–665, 2008.

[15] ——, *"Accelerated appearance-only SLAM"* in Robotics and automation, IEEE international conference on ICRA IEEE, 2008, pp. 1828–1833.

[16] S. Perera and A. Pasqual, *"Towards realtime handheld monoslam in dynamic environments"*, in Advances in Visual Computing. Springer, 2011, pp. 313–324.

[17] R. Ozawa, Y. Takaoka, Y. Kida, K. Nishiwaki, J. Chestnutt, J. Kuffner, S. Kagami, H. Mizoguch, and H. Inoue, *"Using visual odometry to create 3d maps for online footstep planning"* in Systems, Man and Cybernetics, International Conference on IEEE, 2005, vol. 3, pp. 2643–2648.

[18] D. Nistér, *"Preemptive ransac for live structure and motion estimation"*, Machine Vision and Applications, vol. 16, no. 5, pp. 321–329, 2005.

[19] P. Michel, J. Chestnutt, S. Kagami, K. Nishiwaki, J. Kuffner, and T. Kanade, *"Gpu-accelerated real-time 3d tracking for humanoid locomotion and stair climbing"* in Intelligent Robots and Systems, IEEE/RSJ International Conference on IROS IEEE, 2007, pp. 463–469.

[20] Q.-H. Nguyen, H. Vu, T.-H. Tran, and Q.-H. Nguyen, "*Developing a way-finding system on mobile robot assisting visually impaired people in an indoor environment,"* Multimedia Tools and Applications, pp. 1–25, 2016.

[21] A. Bosch, X. Mun˜oz, and R. Martí, *"Which is the best way to organize/classify images by content?"*, Image and vision computing, vol. 25, no. 6, pp. 778–791, 2007.

[22] C. Chow and C. Liu, *"Approximating discrete probability distributions with dependence trees"*, IEEE Transactions on Information Theory, vol. 14, no. 3, pp. 462–467, 1968.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, *"Surf: Speeded up robust features,"* Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359, 2006.

[24] J. A. Hartigan and M. A. Wong, *"Algorithm as 136: A k-means clustering algorithm"*, Applied statistics, pp. 100–108, 1979.

[25] A. Oliva and A. Torralba, *"Modeling the shape of the scene: A holistic representation of the spatial envelope"*, International journal of computer vision, vol. 42, no. 3, pp. 145–175, 2001.

[26] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, *"The pascal visual object classes challenge"*, in International Journal of Computer Vision, vol. 88, no. 2, 2009, pp. 303–338.

PERFORMANCE EVALUATION OF FAB-MAP* FOR ROBOT LOCALIZATION IN INDOOR ENVIRONMENT USING MONOCULAR CAMERA

Nguyen Quoc Hung, Vu Hai, Tran Thanh Hai, Nguyen Quang Hoan

ABSTRACT— *This paper present FAB-MAP* algorithm localization robots use visual information in an indoor environment with the main idea is to locate the robot by the operation conditional probabilities between observations present a collection of observations that robots move through, these observations from previous training based on specific classification trees frame the scene and the smallest coalition of environmental Dictionary copper appears. Thereby determining the exact location on a map built environment before. In this paper we focus implement performance evaluation FAB-MAP algorithm * on the database collected at Ta Quang Buu Library (Vietnam) and Milano-Bicocca (Italy). The results show that the algorithm positioning FAB-MAP * feasible in problem locating objects in the home, as a basis for building applications for the robot SLAM problems in the future.*