

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

VŨ TUẤN ANH

**THEO DÕI ĐỐI TƯỢNG DỰA TRÊN GIẢI THUẬT DI
TRUYỀN VÀ TỐI ƯU HOÁ BẦY ĐÀN**

Hà Nội – 11/2016

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

VŨ TUẤN ANH

**THEO DÕI ĐỐI TƯỢNG DỰA TRÊN GIẢI THUẬT DI
TRUYỀN VÀ TỐI ƯU HOÁ BẦY ĐÀN**

Ngành: Công nghệ thông tin

Chuyên ngành: Công nghệ phần mềm

Mã Số: 60 48 01 03

LUẬN VĂN THẠC SĨ

NGƯỜI HƯỚNG DẪN: PGS. TS. PHẠM NGỌC HÙNG

Hà Nội – 11/2016

MỤC LỤC

LỜI CẢM ƠN	ii
LỜI CAM ĐOAN	iii
DANH MỤC HÌNH VẼ	iv
DANH MỤC THUẬT NGỮ	v
CHƯƠNG 1: ĐẶT VẤN ĐỀ	1
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT	6
2.1. Phân hoạch mờ	6
2.2. Giải thuật di truyền.....	9
2.3. Giải thuật tối ưu bầy đàn	14
CHƯƠNG 3: ÁP DỤNG GIẢI THUẬT DI TRUYỀN VÀ TỐI ƯU BẦY ĐÀN TRONG BÀI TOÁN THEO DÕI ĐỐI TƯỢNG.....	18
3.1. Histogram màu	19
3.2. Phát hiện đối tượng dựa trên giải thuật di truyền và tối ưu bầy đàn..	20
3.3. Theo dõi đối tượng	26
CHƯƠNG 4: CÀI ĐẶT THỬ NGHIỆM.....	32
4.1. Công cụ hỗ trợ	32
4.2. Dữ liệu thử nghiệm.....	34
4.3. Kết quả thử nghiệm.....	35
CHƯƠNG 5: KẾT LUẬN	41
TÀI LIỆU THAM KHẢO	43

LỜI CẢM ƠN

Trước tiên tôi xin dành lời cảm ơn chân thành và sâu sắc đến thầy giáo, PGS. TS. Phạm Ngọc Hùng – người đã hướng dẫn, khuyến khích, chỉ bảo và tạo cho tôi những điều kiện tốt nhất từ khi bắt đầu cho tới khi hoàn thành công việc của mình.

Tôi xin dành lời cảm ơn chân thành tới các thầy cô giáo khoa Công nghệ thông tin, trường Đại học Công nghệ, ĐHQGHN đã tận tình đào tạo, cung cấp cho tôi những kiến thức vô cùng quý giá và đã tạo điều kiện tốt nhất cho tôi trong suốt quá trình học tập, nghiên cứu tại trường.

Đồng thời tôi xin cảm ơn tất cả những người thân yêu trong gia đình tôi cùng toàn thể bạn bè những người đã luôn giúp đỡ, động viên tôi những khi vấp phải những khó khăn, bế tắc.

Cuối cùng, tôi xin chân thành cảm ơn các đồng nghiệp của tôi tại Viện hàng không vũ trụ đã giúp đỡ, tạo điều kiện thuận lợi cho tôi học tập và nghiên cứu chương trình thạc sĩ tại Đại học Công nghệ, ĐHQGHN.

LỜI CAM ĐOAN

Tôi xin cam đoan rằng luận văn thạc sĩ công nghệ thông tin “Theo dõi đối tượng dựa trên giải thuật di truyền và tối ưu hoá bầy đàn” là công trình nghiên cứu của riêng tôi, không sao chép lại của người khác. Trong toàn bộ nội dung của luận văn, những điều đã được trình bày hoặc là của chính cá nhân tôi hoặc là được tổng hợp từ nhiều nguồn tài liệu. Tất cả các nguồn tài liệu tham khảo đều có xuất xứ rõ ràng và hợp pháp.

Tôi xin hoàn toàn chịu trách nhiệm và chịu mọi hình thức kỷ luật theo quy định cho lời cam đoan này.

Hà Nội, ngày 15 tháng 10 năm 2016

Vũ Tuấn Anh

DANH MỤC HÌNH VẼ

Hình 2.1. Sơ đồ chung giải thuật di truyền.....	10
Hình 3.1. Các bước giải thuật theo dõi đối tượng	18
Hình 3.2. Chuỗi chứa c vectơ	21
Hình 3.3. Minh hoạ lại ghép một điểm cắt	23
Hình 3.4. Bố trí lân cận của một điểm ảnh	25
Hình 3.5. Sơ đồ từng bước giải thuật theo vết đối tượng	29
Hình 4.1. Giao diện làm việc với Matlab	33
Hình 4.2. Cấu trúc công cụ thử nghiệm GAObjectTracking.....	34
Hình 4.3. Khung hình thử nghiệm	35
Hình 4.4. Ảnh phân vùng màu	35
Hình 4.5. Ảnh nhị phân đường	36
Hình 4.6. Ảnh nhị phân đường sau khi lọc	36
Hình 4.7. Ảnh nhị phân phương tiện giao thông	37
Hình 4.8. Ảnh nhị phân phương tiện giao thông sau khi lọc.....	37
Hình 4.9. Bao của đối tượng trích xuất được	38
Hình 4.10. Biểu diễn trên ảnh gốc	38

DANH MỤC THUẬT NGỮ

Từ viết tắt	Từ đầy đủ	Ý nghĩa
GA	Genetic Algorithm	Giải thuật di truyền
PSO	Particle Swarm Optimization	Tối ưu bầy đàn
GMM	Gaussian Mixen Model	Mô hình trộn Gauss
SVM	Support Vector Machine	Máy hỗ trợ vectơ
NN	Neural Network	Mạng nơ-ron
MOTA	Multiple Object Tracking Accuracy	Độ chính xác theo dõi đa đối tượng
UAV	Unmanned Aerial Vehicle	Máy bay không người lái
AI	Artificial intelligence	Trí tuệ nhân tạo

CHƯƠNG 1: ĐẶT VẤN ĐỀ

Trong thập niên đầu của thế kỷ 21, học máy được nghiên cứu và phát triển mạnh mẽ, đánh dấu bước ngoặt quan trọng thay đổi nền tảng nghiên cứu của Trí tuệ nhân tạo. Học máy liên quan đến việc xây dựng các chương trình máy tính có thể tự động thu thập tri thức, cải thiện khả năng của mình thông qua các kinh nghiệm, và việc nghiên cứu các nguyên lý của quá trình học [1]. Các kết quả và công nghệ của học máy được thể hiện qua các ứng dụng đa dạng trong thực tế trong các lĩnh vực như: xử lý ngôn ngữ tự nhiên, thị giác máy tính, tìm kiếm và nhận dạng, robotics, khai phá dữ liệu, v.v.

Thị giác máy tính, một lĩnh vực nghiên cứu liên ngành, liên quan đến việc nghiên cứu các lĩnh vực khoa học và công nghệ về các hệ thống máy móc có khả năng nhìn và hiểu như hệ thống thị giác con người [2]. Đây là một lĩnh vực được quan tâm nghiên cứu rộng rãi trong một vài thập niên gần đây bởi những ứng dụng thực tế đa dạng của nó. Một số ứng dụng có thể kể đến là: tự động hóa trong dây chuyền sản xuất công nghiệp, viễn thám, giám sát giao thông, bảo mật bằng sinh trắc học, y học, an ninh, web 3D, giải trí, v.v.

Vấn đề phát hiện, nhận dạng, phân tách và hiểu ngữ nghĩa của đối tượng trong ảnh/video đã được nghiên cứu rộng rãi trong lĩnh vực thị giác máy tính hàng thập kỷ qua [2]. Các nghiên cứu được nhanh chóng phát triển nhờ những tiến bộ trong một số lĩnh vực liên quan như: việc phát triển các mô hình toán học phức tạp, các nghiên cứu chuyên sâu về nhận thức tri giác (cognitive vision), năng lực của các hệ thống tính toán, các giải thuật thông minh, cũng như đòi hỏi của kiểm thử trên các bộ dữ liệu lớn.

Tuy nhiên vấn đề này vẫn còn khá mới mẻ ở Việt Nam bởi thiếu các thiết

bị hỗ trợ và nghiên cứu làm chủ công nghệ. Và đây cũng là một hướng phát triển mở nhiều hứa hẹn và đồng thời cũng nhiều thách thức. Hiện nay ở Việt Nam các hệ thống theo dõi – giám sát hầu hết là không tự động, chủ yếu vẫn dựa vào con người. Tuy nhiên trong tương lai không xa, khi kinh tế và khoa học kỹ thuật phát triển thì các hệ thống giám sát này cũng sẽ phát triển theo. Với mong muốn tham gia vào hướng nghiên cứu còn mới này và giúp các hệ thống giám sát đạt hiệu quả cao hơn và giảm được chi phí con người chúng tôi thực hiện đề tài “*Theo dõi đối tượng dựa trên giải thuật di truyền và tối ưu hoá bầy đàn*”.

Vấn đề phát hiện đối tượng đang được nghiên cứu và có nhiều ứng dụng trong cuộc sống. Các đối tượng được phát hiện nhờ những thông tin trong một khung hình ảnh. Có rất nhiều hướng tiếp cận để giải quyết vấn đề trên. Các tác giả Alper Yilmaz, Omar Javed và Mubarak Shah đã phân loại các hướng tiếp cận này được trình bày trong [3]. Có thể phân loại các giải thuật phát hiện đối tượng thành các hướng tiếp cận như: phát hiện điểm quan trọng (interest point detector) [4] [5], phân đoạn ảnh (segmentation) [6] [7] [8], mô hình nền (background modeling) [9] [10] [11] và phân loại có giám sát (supervised classifier) [12] [13]. Việc lựa chọn phương pháp áp dụng phải dựa vào tình huống cụ thể, đối với trường hợp có ảnh nền không thay đổi việc phát hiện đối tượng chuyển động có thể bằng các phương pháp trừ nền. Các giải thuật này sẽ được trình bày sau đây. Hướng giải quyết là xây dựng mô hình nền, sau đó sử dụng mô hình này cùng với khung hình hiện tại để rút ra được các vật thể chuyển động. Để có thể tiếp cận cần phải xây dựng được mô hình nền. Có nhiều phương pháp được xây dựng dựa trên mô hình nền bởi các tác giả. Anurag Mittal [12] dùng mô hình ước lượng mật độ nhân thích ứng (Adaptive Kernel Density Estimation) cho kết quả tốt tuy nhiên khó khăn về không gian lưu trữ, tính toán phức tạp, tốc độ không đáp ứng thời gian thực. Stauffer sử dụng mô hình trộn Gaussian (Mixture of Gaussian) [14]

để xây dựng mô hình nền, nhằm phát hiện được các đối tượng chuyển động, xác định xem những đối tượng này có đúng là những đối tượng ta cần phát hiện hay không. Đây là các khó khăn cần khắc phục.

Trong các lĩnh vực về phát hiện phần đầu của người thì Wei Qu, Nidhal Bouaynaya và Dan Schonfeld [15] đề ra hướng tiếp cận bằng cách kết hợp mô hình màu da cùng với mô hình màu tóc (skin and hair color model). Những màu này được phát hiện dựa vào mô hình Gauss. Sau đó bằng cách áp dụng phương pháp so khớp mẫu (template matching) để đạt được mục đích phát hiện phần đầu người đáp ứng thời gian thực. Khó khăn trong hướng tiếp cận này thường gặp ở việc thu thập dữ liệu huấn luyện màu da và màu tóc, độ chính xác dễ bị ảnh hưởng bởi độ sáng của môi trường.

Việc phát hiện đối tượng có thể được thực hiện bằng các phương pháp học máy. Các phương pháp này có thể kể đến như: mạng nơ-ron (Neural Network), cây quyết định (Decision Tree), máy hỗ trợ vector (Support Vector Machine - SVM). Điểm chung của các phương pháp này đều phải trải qua giai đoạn huấn luyện trên một tập dữ liệu. Tập dữ liệu này phải đủ lớn, bao quát hết được các trạng thái của đối tượng. Sau đó các đặc trưng sẽ được rút trích ra trên bộ dữ liệu huấn luyện này. Việc lựa chọn đặc trưng sử dụng đóng vai trò quan trọng ảnh hưởng đến hiệu quả của các phương pháp học máy. Một số đặc trưng thường được sử dụng như: đặc trưng về màu sắc, đặc trưng về góc cạnh, đặc trưng histogram, v.v. Sau khi đã có được đặc trưng, ta sẽ đánh nhãn lớp cụ thể cho các đặc trưng đó để sử dụng trong việc huấn luyện. Trong quá trình huấn luyện, các phương pháp học máy sẽ sinh ra một hàm để ánh xạ những đặc trưng đầu vào tương ứng với nhãn lớp cụ thể. Sau khi đã huấn luyện xong thì các phương pháp học máy trên sẽ được dùng để phân lớp cho những đặc trưng mới. Đặc điểm của phương pháp này là độ chính xác cao. Tuy nhiên nó gặp phải khó khăn trong việc thu thập dữ liệu huấn luyện ban

đầu, tốn thời gian và chi phí cho quá trình học máy.

Luận văn này nhằm mục đích nghiên cứu, xây dựng giải thuật theo dõi tự động các đối tượng có trong video. Giải thuật theo dõi cần có độ chính xác tốt, đồng thời chi phí tính toán thấp phục vụ các ứng dụng thời gian thực. Do đó, luận văn tập trung đi sâu vào việc khảo sát các đặc trưng của video, đặc trưng ảnh, đặc trưng của đối tượng chuyển động, đặc trưng nền, v.v. từ đó áp dụng các thuật toán phù hợp, kết hợp với các thuật toán học máy phù hợp để đưa ra kết quả tối ưu, rút ngắn thời gian tính toán và chi phí bộ nhớ, để từ đó hệ thống phù hợp với thời gian thực hơn. Đầu vào của bài toán theo dõi đối tượng là các khung hình video. Qua quá trình xử lý phát hiện đối tượng chuyển động (Object Detection) sẽ đưa ra các đối tượng trong khung hình. Khối phát hiện đối tượng chuyển động có thể coi là quyết định độ chính xác của hệ thống giám sát thông minh bằng hình ảnh, vì hiệu quả, tính chính xác của khối xử lý này sẽ ảnh hưởng đến đầu vào và đầu ra của khối xử lý tiếp theo. Luận văn này sẽ đưa ra các kỹ thuật tối ưu hiệu quả như giải thuật di truyền và tối ưu bầy đàn để tăng độ chính xác và hiệu quả của bước phát hiện đối tượng. Và cuối cùng là quá trình xử lý để theo dõi đối tượng (Object Tracking) đó là việc tìm ra đường chuyển động của đối tượng, dự đoán chuyển động, xử lý nhập nhằng trong chuyển động.

Hiện nay, trên thế giới các hệ thống theo dõi - giám sát thông minh bằng hình ảnh đã được phát triển và đã chứng minh được hiệu quả nhất định trên một số lĩnh vực như giám sát hoạt động con người, giám sát giao thông, v.v. Từ các hình ảnh thu được từ những nơi được quan sát, ta có thể phát hiện được chuyển động của các đối tượng trong các khung hình, xác định được đối tượng đó là người, phương tiện hay vật thể gì. Nhiều hệ thống đã được nghiên cứu và phát triển. Chẳng hạn, với bài toán giám sát giao thông có thể cho chúng ta biết được số lượng phương tiện lưu thông qua đoạn đường được theo dõi, đưa ra thông tin về tốc độ chuyển động,

đường đi của đối tượng được theo dõi v.v. Tuy nhiên, các hệ thống vẫn gặp phải một số tồn tại như hiệu quả của việc quan sát luôn phụ thuộc vào điều kiện môi trường quan sát, kiểu chuyển động của đối tượng hay các lý do khác quan khác. Vì vậy, với khả năng cá nhân, tôi mong muốn làm chủ các công nghệ theo dõi đối tượng, từ đó xây dựng các ứng dụng phù hợp với môi trường Việt Nam, phục vụ an ninh - quốc phòng, đem lại các lợi ích về kinh tế cho đất nước.

Luận văn này được cấu trúc các phần như sau. Chương 2 tiếp theo là một định nghĩa cơ bản được sử dụng trong luận văn, bao gồm: lý thuyết trích xuất đặc trưng, giải thuật phân hoạch mờ, giải thuật di truyền và giải thuật tối ưu bầy đàn. Chương 3 trình bày cách tiếp cận giải quyết bài toán theo dõi đối tượng của luận văn. Cách tiếp cận này được ứng dụng giải quyết với đối tượng cụ thể là phương tiện giao thông chụp từ ảnh UAV, các kết quả thử nghiệm chỉ ra ở chương 4. Cuối cùng là các kết luận, định hướng mở rộng được đưa ra ở chương 5 và danh sách các tài liệu tham khảo.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1. Phân hoạch mờ

Ngày nay, xã hội đã phát triển đồng thời sự tiến bộ của khoa học công nghệ, các suy luận lôgic nguyên thủy (hay lôgic rõ) với hai giá trị đúng sai hay 1, 0 riêng biệt đã không giải quyết được hết các bài toán phức tạp nảy sinh trong thực tế. Ví dụ, một bộ phim thể nào được gọi là hay hay không hay, một bức ảnh đẹp hay không đẹp. Những bài toán như vậy ngày xuất hiện càng phổ biến trong lĩnh vực điều khiển tự động cũng như khoa học máy tính. Năm 1965, giáo sư Lotfi Zadeh trường đại học California – Hoa Kỳ đã đưa ra lý thuyết mờ [16], một cách tiếp cận mới mang lại nhiều hiệu quả trong thực tiễn và đang được tiếp tục phát triển. Công trình này thực sự đã khai sinh một ngành khoa học mới là lý thuyết tập mờ (fuzzy set theory) và nhanh chóng được các nhà nghiên cứu chấp nhận và phát triển. Lý thuyết tập mờ ngày càng phong phú và hoàn chỉnh đã tạo nền tảng vững để phát triển lôgic mờ.

Theo như khung mẫu mờ của George và Bo năm 1995 [17], phân cụm mờ (fuzzy clustering) được xem là cách phân hoạch không gian dữ liệu thành một số tập mờ và gán mỗi điểm dữ liệu như một thành viên một cụm. Coi tập vector V biểu diễn bởi n vector trong không gian L chiều số thực R^L , $V = \{v_1, v_2, \dots, v_n\}$, $v_j = (v_{j1}, v_{j2}, \dots, v_{jL}) \in R^L$ với $j = 1 \dots n$, một phân hoạch mờ (fuzzy c-partition) được biểu diễn bởi một ma trận phân hoạch mờ $P = [p_{ij}]$, $i = 1 \dots c$, trong đó c là số nguyên dương chỉ số lượng cụm của phân hoạch, và $p_{ij} \in [0,1]$ là giá trị thành viên mờ thể hiện khả năng v_j thuộc về cụm thứ i và thoả mãn điều kiện (2.1) và (2.2).

$$\sum_{i=1}^c p_{ij} = 1 \text{ với tất cả } j = 1 \dots n \quad (2.1)$$

$$0 < \sum_{j=1}^n p_{ij} \leq n \text{ với tất cả } i = 1 \dots c \quad (2.2)$$

Trước khi sử dụng phân hoạch mờ để thiết kế một giải thuật phân cụm, hai vấn đề sau cần được giải quyết. Vấn đề thứ nhất là cách xác định số lượng cụm cho giải thuật phân cụm. Vấn đề nữa là cách tính ma trận phân hoạch mờ. Thông thường thì số cụm được xác định trước bởi người dùng theo kinh nghiệm. Vấn đề thứ hai được giải quyết bởi độ đo tương đồng như sau.

Cho một tập vector $V = \{v_1, v_2, \dots, v_n\}$, ma trận phân hoạch mờ được tính như công thức (2.3).

$$p_{ij} = \frac{\mu(u_i, v_j)^{\frac{1}{m-1}}}{\sum_{k=1}^c \mu(u_k, v_j)^{\frac{1}{m-1}}} \quad (2.3)$$

trong đó, $m \in (1, \infty)$ là trọng số mũ của mỗi thành phần mờ, u_i là vector trọng tâm của cụm i , và $\mu(u_i, v_j)$ là độ đo tương đồng giữa vector u_i và v_j được định nghĩa như công thức (2.4).

$$\mu(u_i, v_j) = \exp(-k_1 d(u_i, v_j)) \cos(k_2 \theta(u_i, v_j)) \quad (2.4)$$

trong đó, k_1 và k_2 là các tham số trọng số và $d(u_i, v_j)$ và $\theta(u_i, v_j)$ lần lượt là khoảng cách và góc giữa u_i và v_j được định nghĩa như các công thức (2.5) và (2.6) tương ứng.

$$d(u_i, v_j) = \left(\sum_{l=1}^L |u_{il} - v_{jl}|^2 \right)^{1/2} \quad (2.5)$$

$$\theta(u_i, v_j) = \arccos \left(\frac{\sum_{l=1}^L u_{il} v_{jl}}{\sqrt{\sum_{l=1}^L u_{il}^2 \sum_{l=1}^L v_{jl}^2}} \right) \quad (2.6)$$

Giải thuật phân hoạch mờ được thực hiện lần lượt theo các bước như sau:

- Bước 1: Khởi tạo ma trận $P = [p_{ij}]$, $P(0)$
- Bước 2: Tại lần lặp thứ k : tính toán vector trọng tâm $U(k) = [u_i]$ với $P(k)$

$$u_i = \frac{\sum_{j=1}^n p_{ij}^m \cdot v_j}{\sum_{j=1}^n p_{ij}^m} \quad (2.7)$$

- Bước 3: Cập nhật $P(k)$ và $P(k+1)$ theo công thức (2.3)
- Bước 4: Kiểm tra xem các trọng tâm đã hội tụ chưa, nếu không quay lại bước 2, nếu đã thoả mãn ta kết thúc tính toán.

$$\|P^{(k+1)} - P^{(k)}\| < \varepsilon \left(\max_{ij} \|p_{ij}^{k+1} - p_{ij}^k\| \right) \quad (2.8)$$

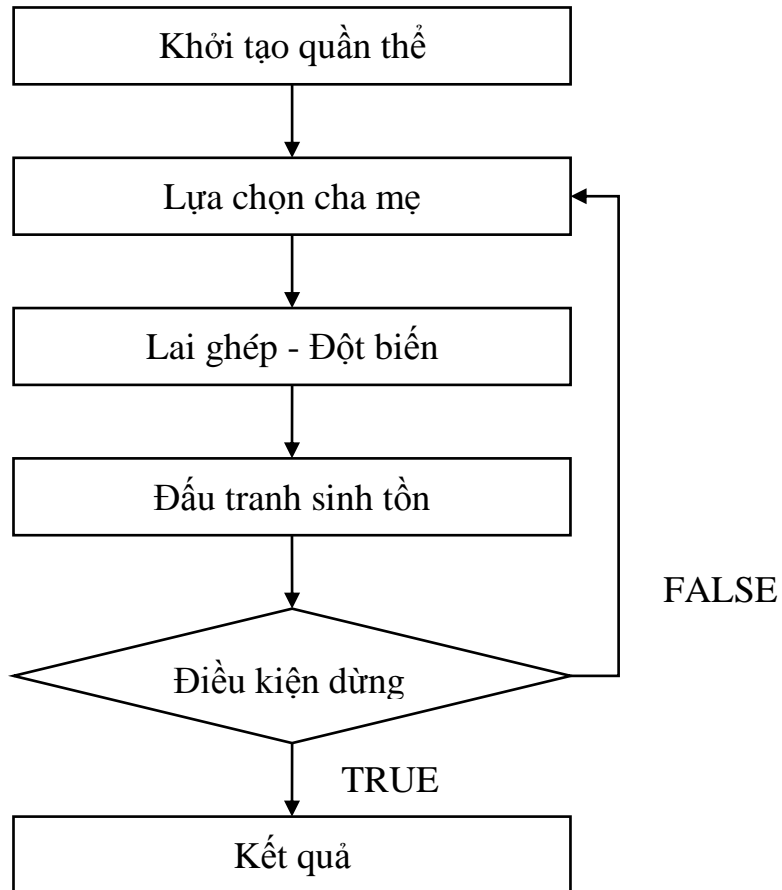
Ưu và nhược điểm

Thông thường, phân hoạch mờ cho kết quả tốt nhất cho dữ liệu chồng chéo nhiều và tương đối tốt hơn thuật toán phân cụm k-means. Không giống như k-means, dữ liệu điểm duy nhất phải thuộc về một cụm duy nhất, ở mỗi điểm được phân vào cụm dựa vào kết quả tính toán hàm thành viên. Vì vậy, một điểm có thể thuộc về nhiều hơn một cụm với giá trị mờ nào đó, giúp tránh được các sai số tích lũy trong tính toán. Tuy nhiên, phân hoạch mờ cũng còn tồn tại một số nhược điểm như: cần tiên nghiệm số lượng các cụm, ε càng thấp kết quả nhận được càng tốt nhưng chi phí tính toán càng nhiều, khoảng cách Euclide các yếu tố cơ bản có thể không đồng đều.

2.2. Giải thuật di truyền

Ý tưởng về giải thuật di truyền đã được một số nhà sinh vật học đưa ra từ những năm 50-60 của thế kỉ 20. A.S. Fraser là người tiên phong nêu lên sự tương đồng giữa sự tiến hóa của sinh vật và chương trình tin học giả tưởng về giải thuật di truyền (Genetic Algorithms - GA). Tuy nhiên, chính John Henry Holland mới là người triển khai ý tưởng và phương pháp giải quyết vấn đề dựa theo sự tiến hóa. Các nguyên lý cơ bản của giải thuật di truyền được tác giả Holland công bố lần đầu tiên vào năm 1962. Sau đó, các nền tảng toán học của giải thuật lần đầu tiên được công bố vào năm 1975 trong cuốn sách “Adaptation in Natural and Artificial System” [18]. Có thể nói Holland là người đi tiên phong nghiên cứu trong lĩnh vực giải thuật di truyền cùng với những tác giả Goldbeg, Beglay... Dựa trên lý thuyết cơ bản về GA của Holand, Keneth De Jong đã triển khai và chứng minh những thành quả do ông thực hiện đã góp phần quan trọng trong việc tạo ra nền tảng toán học cho lý thuyết GA.

Giải thuật di truyền là một giải thuật dựa trên cơ chế của chọn lọc tiến hoá trong tự nhiên. Trong mọi thế hệ, một tập mới các sinh vật được tạo ra bằng cách lai ghép những nhân tố thích nghi nhất với môi trường của những sinh vật trong thế hệ cũ cùng với sự xuất hiện đột biến ngẫu nhiên của các cá thể trong thế hệ m. Hình 2.1 mô tả sơ đồ chung thực hiện giải thuật di truyền, bao gồm các bước chính như sau.



Hình 2.1. Sơ đồ chung giải thuật di truyền

- i. Khởi tạo một quần thể ban đầu (tập lời giải ban đầu của bài toán).
- ii. Tạo ra quần thể mới bằng các phép toán di truyền: lai ghép chéo (crossover) từ các cá thể hiện tại có chọn lọc (selection), đột biến (mutation) các cá thể trong quần thể mới theo một xác suất nhất định.
- iii. Đấu tranh sinh tồn: Đánh giá độ thích nghi thông qua giá trị hàm mục tiêu (fitness) của mỗi cá thể trong quần thể. Các cá thể trong quần thể mới sinh ra được thay thế cho các cá thể trong quần thể cũ dựa trên đánh giá hàm thích nghi.
- iv. Nếu điều kiện dừng thỏa mãn thì giải thuật dừng lại và trả về cá thể tốt nhất cùng với giá trị hàm mục tiêu của nó, nếu không thì quay lại bước 2.

Đây là sơ đồ chung nhất áp dụng cho rất nhiều lớp bài toán sử dụng giải thuật di truyền. Một số khái niệm về giải thuật di truyền sẽ được trình bày ở phần tiếp theo của chương.

So sánh giải thuật di truyền với các phương pháp truyền thống

Chúng ta xét bài toán đơn giản sau đây: tối ưu hoá hàm $y = f(x)$ trên khoảng xác định D . Khi dùng phương pháp truyền thống có một số cách giải sau đây:

- *Phương pháp liệt kê*: Duyệt tất cả các điểm nằm trong vùng khảo sát D để tìm ra điểm cực trị của nó. Phương pháp này không thích hợp khi dữ liệu đầu quá lớn. Trong trường hợp này miền D có lực lượng lớn hơn đếm được.
- *Phương pháp giải tích*: Tìm điểm cực trị bằng cách giải tập các phương trình khi cho Gradient bằng 0. Để xét được Gradient phải tính đạo hàm của hàm số. Điều này không giải quyết được trong trường hợp hàm số không liên tục hoặc không có đạo hàm. Ngoài ra, đối với hàm nhiều cực trị thì có thể phương pháp này bỏ mất cực trị, cực trị tìm được chỉ mang tính chất địa phương.
- *Phương pháp tìm kiếm ngẫu nhiên*: là phương pháp kết hợp giữa phương pháp tính toán giải tích và sơ đồ liệt kê. Tuy nhiên giải thuật tìm kiếm ngẫu nhiên cũng phải bị suy yếu bởi thiếu tính hiệu quả.

Đối với giải thuật di truyền, các thông số của bài toán tìm kiếm phải được mã hoá thành một chuỗi hữu hạn các ký tự trên một tập hữu hạn các ký tự. Chuỗi này tương tự như các chuỗi gen của các cơ thể sinh vật. Có rất nhiều cách để mã hóa tập thông số. Một cách đơn giản là chúng ta có thể mã hoá thành các chuỗi bit trên tập ký tự $\{0,1\}$. Mỗi một chuỗi đại diện cho một điểm tìm kiếm trong không gian. GA xuất phát với một quần thể các chuỗi được khởi tạo một cách ngẫu nhiên sau đó sẽ sản sinh các quần thể tiếp theo thông qua việc sử dụng lựa chọn ngẫu nhiên như một công

cụ. Nhờ đó giải thuật di truyền tìm kiếm trên nhiều điểm song song có khả năng leo lên nhiều cực trị cùng một lúc. Thông qua các toán tử của mình, giải thuật trao đổi thông tin giữa các cực trị với nhau, từ đó làm giảm thiểu khả năng giải thuật kết thúc tại các cực trị địa phương và bỏ qua mất cực trị toàn cục.

Đây là các đặc trưng của giải thuật di truyền so với các phương pháp truyền thống:

- Các giải thuật di truyền làm việc với sự mã hoá của tập thông số chứ không làm việc với các giá trị của các thông số.
- Các giải thuật di truyền tìm kiếm từ một quần thể các điểm chứ không phải từ một điểm.
- Các giải thuật di truyền chỉ sử dụng thông tin về các tiêu chuẩn tối ưu của hàm mục tiêu chứ không dùng các thông tin hỗ trợ nào khác.
- Các giải thuật di truyền sử dụng các luật chuyển đổi mang tính xác suất chứ không phải là các luật chuyển đổi mang tính xác định.
- Các giải thuật di truyền thường dễ cài đặt, áp dụng. Tuy nhiên không phải lúc nào cũng cho lời giải chính xác. Một số giải thuật di truyền có thể cung cấp lời giải tiềm năng cho một bài toán xác định để người sử dụng lựa chọn.

Các ứng dụng của giải thuật di truyền.

Ban đầu giải thuật di truyền ra đời được áp dụng cho tối ưu hoá và học máy là chủ yếu. Đến nay nó đã phát triển mạnh và có nhiều ứng dụng thực tế, đặc biệt là các bài toán về trí tuệ nhân tạo. Ví dụ: ta có thể tối ưu công việc dự báo thời tiết sao cho chính xác nhất dựa trên các thông số khí tượng do được. Năm 1967, Beglay [19] xây dựng máy chơi cờ Hexapawn dựa trên thuật giải di truyền và nhận ra rằng thuật giải Di truyền có thể thực hiện tốt mà không phụ thuộc độ phức tạp của trò chơi.

Tối ưu hoá và học máy

Trong lĩnh vực tối ưu hóa có nhiều bài toán được áp dụng giải thuật di truyền và đã thành công như tối ưu hoá hàm một biến, tối ưu hóa hàm nhiều biến, hay như bài toán người du lịch, bài toán hộp đen, các bài toán kinh doanh, nhận dạng điều khiển hệ thống, v.v. Sau đây sẽ giới thiệu một số bài toán tối ưu hóa.

David E. Goldberg [20] đã ứng dụng giải thuật di truyền để tối ưu hóa bài toán điều khiển hệ thống đường ống dẫn khí thiên nhiên. Trong bài toán này, mục tiêu là cực tiểu hóa năng lượng do quá trình nén, phụ thuộc vào áp suất tối đa và áp suất tối thiểu và các ràng buộc tỉ lệ áp suất.

Trong tối ưu hoá kết cấu [20], mục tiêu của bài toán là cực tiểu hóa trọng lượng của kết cấu, phụ thuộc vào các ràng buộc về ứng suất lớn nhất và ứng suất nhỏ nhất của mỗi thanh. Một bộ mã cho khung kết cấu theo ma trận tiêu chuẩn được dùng để phân tích mỗi thiết kế tạo ra bởi giải thuật di truyền.

Trong lĩnh vực học máy, giải thuật di truyền được sử dụng cho việc tìm hiểu các quy luật có cấu trúc như cấu trúc IF-THEN trong môi trường nhân tạo [20].

Ghi ảnh y học với giải thuật di truyền

Giải thuật di truyền đơn giản đã được sử dụng để thực hiện ghi hình ảnh, như là bộ phận của hệ thống lớn có tên là Digital Subtraction Angiography (DSA) [21]. Trong DSA, bác sĩ sẽ cố gắng xem xét bên trong của một động mạch khả nghi bằng cách so sánh hình ảnh x-quang, một được chụp trước khi tiêm thuốc đã nhuộm màu vào động mạch, một và một được chụp sau khi tiêm thuốc. Cả hai hình được số hóa và được trừ nhau theo từng điểm một, với kết quả mong muốn cuối cùng nhận được một hình ảnh sai khác phác họa rõ ràng hình ảnh bên trong động mạch chủ. Tuy nhiên sự chuyển động nhẹ của bệnh nhân có thể tạo ra hai

hình ảnh kế nhau, làm rối loạn phần hình ảnh sai khác. Kết quả là, các hình ảnh phải được xếp kế nhau, để tính toán phần hình ảnh sai khác. Giải thuật di truyền được dùng để tìm kiếm các hệ số biến đổi để tìm kiếm các hệ số giúp cực tiểu hóa sự sai biệt hình ảnh trước và sau khi tìm, trên cơ sở các sai khác hình ảnh tuyệt đối.

Một số ứng dụng khác có thể kể đến của giải thuật di truyền như: thiết kế mạng nơ-ron, kiến trúc lẫn trọng số, quỹ đạo cho người máy; mô phỏng các hệ phi tuyến động - phỏng đoán, phân tích dữ liệu; tìm dạng của các phân tử protein; cải tiến chương trình LISP (lập trình gen). Có thể thấy giải thuật di truyền đã được áp dụng rộng rãi trong lĩnh vực khoa học máy tính và giải quyết hiệu quả nhiều bài toán thiết thực.

2.3. Giải thuật tối ưu bầy đàn

Giải thuật tối ưu hóa theo bầy đàn (Particle Swarm Optimization – PSO) là một kỹ thuật tối ưu hóa ngẫu nhiên dựa trên một quần thể được phát triển bởi Eberhart và Kennedy, phỏng theo hành vi của các bầy chim hay các đàn cá. Cũng giống như giải thuật di truyền, PSO tìm kiếm giải pháp tối ưu bằng việc cập nhật các thế hệ. Tuy nhiên, không giống như giải thuật di truyền, PSO không có các thao tác tiến hóa như là lai ghép hay đột biến.

Năm 1987, quan sát quá trình chuyển động của các theo bầy đàn (bầy chim, đàn cá), Reynolds [22] đưa ra nhận ra ba quy luật: Tách biệt; Sắp hàng và Liên kết. Từ nghiên cứu của Reynolds, Eberhart và Kennedy [23] đưa thêm giả thuyết về quá trình tìm về tổ của bầy đàn theo các quy luật:

- i. Tất cả các phần tử trong bầy đàn đều có xu hướng chuyển động về tổ.
- ii. Mỗi phần tử đều ghi nhớ vị trí gần tổ nhất nó đã đạt tới.

Tương tự như vậy, hai ông đưa giả thuyết về quá trình tìm mồi của bầy

đàn trong một vùng không gian mà các phần tử trong bầy đàn đều biết thông tin về thức ăn cách bao xa và lưu giữ vị trí gần thức ăn nhất mà chúng đã đạt tới. Khi đó, cách tốt nhất để tìm thức ăn là theo sau những con phần tử đầu đàn – những con trong bầy gần chỗ thức ăn nhất. Từ đó, hai ông đề xuất thuật toán PSO phỏng theo kịch bản này và sử dụng nó để giải các bài toán tối ưu.

Trong PSO, mỗi giải pháp đơn là một phần tử (particle) trong kịch bản trên. Mỗi phần tử được đặc trưng bởi hai tham số là vị trí hiện tại của phần tử - $present[]$ và vận tốc - $v[]$. Đây là hai vectơ trên trường số R^n với n là số chiều của phần tử được xác định từ bài toán cụ thể. Đồng thời mỗi phần tử có một giá trị thích nghi (fitness value), được đánh giá bằng hàm đo độ thích nghi (fitness function). Tại thời điểm xuất phát, bầy đàn, hay chính xác là vị trí của mỗi phần tử được khởi tạo một cách ngẫu nhiên (hoặc theo một cách thức nào đó dựa vào tri thức biết trước về bài toán). Trong quá trình chuyển động, mỗi phần tử chịu ảnh hưởng bởi hai thông tin: thông tin thứ nhất, gọi là $pBest$, là vị trí tốt nhất mà phần tử đó đã đạt được trong quá khứ; thông tin thứ hai, gọi là $gBest$, là vị trí tốt nhất mà cả bầy đàn đã đạt được trong quá khứ. Trong nguyên bản do Eberhart và Kennedy đưa ra, các phần tử trong PSO sẽ duyệt không gian bài toán bằng cách theo sau các phần tử có điều kiện tốt nhất hiện thời (độ thích nghi lớn nhất). Cụ thể là sau mỗi khoảng thời gian rời rạc, vận tốc và vị trí của mỗi phần tử được cập nhật theo các công thức (2.9) và (2.10) được định nghĩa như sau.

$$v[] = v[] + c_1 \cdot rand() * (pbest[] - present[]) + c_2 \cdot rand() * (gbest[] - present[]) \quad (2.9)$$

$$present[] = present[] + v[] \quad (2.10)$$

Trong đó, $rand()$ là một số ngẫu nhiên trong khoảng $(0,1)$; c_1 , c_2 là các hệ số học, chúng thường được chọn là $c_1 = c_2 = 2$. Mã giả của thuật toán PSO được trình bày dưới đây.

Giải thuật tối ưu bầy đàn

```

ForEach particle
    Khởi tạo particle;
EndFor
Do
    ForEach particle
        Tính fitness value;
        If (fitness value <  $pBest$ ) then
             $pBest$  = the fitness value;
        EndIf
        If ( $pBest$  <  $gBest$ ) then
             $gBest$  =  $pBest$ 
        EndIf
    EndFor
    ForEach particle
        Tính vận tốc theo công thức (2.9);
        Cập nhật vị trí theo công thức (2.10);
    EndFor
While (chưa thỏa mãn điều kiện dừng);

```

Trong đó, một số điều kiện dừng phổ biến là: số lần cập nhật, số lần cập nhật bầy đàn mà không đưa lại kết quả tốt hơn, số lần cập nhật mà lượng thay đổi giữa hai lần cập nhật liên tiếp nhỏ hơn một ngưỡng nào đó... Ngoài ra điều kiện dừng có thể được xác định từ bài toán cụ thể. Phiên bản ban đầu của PSO được trình bày ở trên được gọi là phiên bản “tốt nhất toàn cục” (global best), trong đó vận tốc của mỗi phân tử đều chỉ bị ảnh hưởng bởi hai yếu tố là: yếu tố nội tại – vị trí tốt nhất nó đã từng đạt được – và yếu tố toàn cục – vị trí tốt nhất cả bầy đã đạt được. Các cải tiến của PSO đưa vào yếu tố “cục bộ”, tức là vận tốc của mỗi phân tử trong quá trình chuyển động còn bị tác động bởi vị trí tốt nhất đã đạt được trong

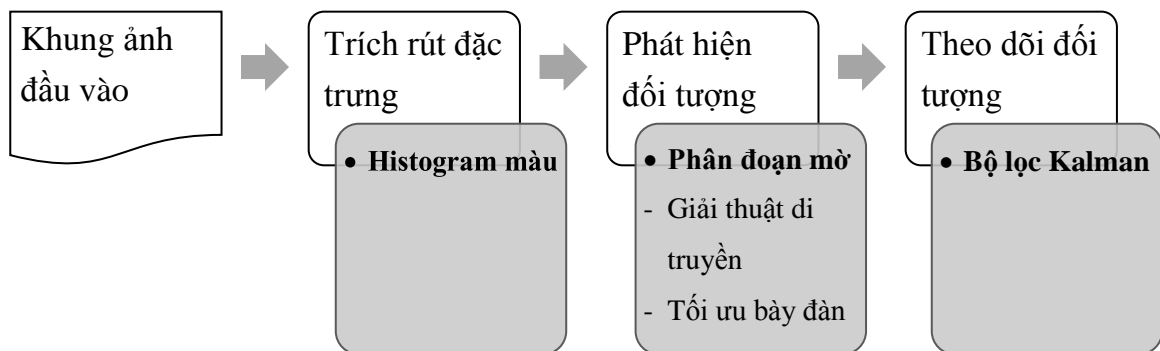
số những hàng xóm của nó $lbest[]$. Khi đó, công thức cập nhật vận tốc được định nghĩa trong (2.11).

$$\begin{aligned} v[] &= v[] + c_1.rand() * (pbest[] - present[]) \\ &\quad + c_2.rand() * (gbest[] - present[]) \\ &\quad + c_3.rand() * (lbest[] - present[]) \end{aligned} \quad (2.11)$$

Theo nghiên cứu của M. Clerc [24], việc chỉ sử dụng yếu tố cục bộ (được gọi là phiên bản “tốt nhất cục bộ” của PSO) thường đem lại hiệu quả tốt hơn so với việc sử dụng yếu tố toàn cục hoặc sử dụng cả hai yếu tố (cục bộ và toàn cục). Để thống nhất trong việc sử dụng PSO, các nhà nghiên cứu đã thống nhất đưa ra phiên bản PSO chuẩn và bản chuẩn mới nhất hiện nay là Standard PSO 2011 (SPSO-11). Thông tin về các phiên bản chuẩn, các biến thể và những ứng dụng của PSO có thể tìm thấy tại [25].

CHƯƠNG 3: ÁP DỤNG GIẢI THUẬT DI TRUYỀN VÀ TỐI ƯU BÀY ĐÀN TRONG BÀI TOÁN THEO DÕI ĐỐI TƯỢNG

Giải thuật theo dõi đối tượng bao gồm ba bước: trích rút đặc trưng, phát hiện đối tượng và theo dõi đối tượng (như Hình 3.1). Trong bước đầu tiên, đặc trưng được trích rút từ các khung ảnh của video. Luận văn này sử dụng histogram màu, một đặc trưng đơn giản và hiệu quả, cho phép trích xuất nhanh chóng các thông tin về màu sắc từ ảnh. Sau đó, ta tiến hành phát hiện, tách các đối tượng ra khỏi khung cảnh nền sử dụng phương pháp phân đoạn mờ. Đây là bước rất quan trọng, quyết định lớn tới hiệu năng hiệu quả của toàn giải thuật theo dõi. Để lựa chọn được cách phân đoạn tốt nhất, nhiều kỹ thuật tối ưu tiên tiến có thể được áp dụng. Trong luận văn này, tôi đề xuất hai hướng tối ưu sử dụng giải thuật di truyền hoặc tối ưu bày đàn, bởi không gian tìm kiếm lời giải lớn của bài toán. Hai giải thuật này đều là những kỹ thuật hiệu quả giải quyết các bài toán tối ưu. Trong chương này, tôi trình bày chi tiết cách áp dụng hai kỹ thuật này trong phát hiện đối tượng. Sau khi trích xuất được các đối tượng, tôi tiến hành theo dõi các đối tượng dựa trên bộ lọc Kalman theo giả thiết các đối tượng thay đổi nhỏ các tham số vị trí, hướng và vận tốc giữa hai khung hình liên tiếp.



Hình 3.1. Các bước giải thuật theo dõi đối tượng

3.1. Histogram màu

Histogram màu (colour histogram) là một kỹ thuật phân tích ảnh màu quan trọng bởi tính hiệu quả và độ phức tạp tính toán thấp của nó [26]. Thông thường, một histogram màu biểu diễn phân bố thống kê của các màu trong một ảnh màu trên một không gian cho trước.

Giả sử không gian màu được chia thành I giỏ (bin), histogram màu của ảnh màu có n điểm ảnh được biểu diễn như là một vectơ $H = [h_0, \dots, h_{I-1}]$, trong đó mỗi giá trị h_i thể hiện con số thống kê số lượng điểm ảnh thuộc về giỏ i và được định nghĩa như công thức (3.1).

$$h_i = \frac{n_i}{n}, i = 0, 1, \dots, I - 1 \quad (3.1)$$

trong đó n_i là số lượng điểm ảnh có màu thuộc giỏ thứ i .

Không gian màu RGB được rời rạc hoá theo các không gian R, G, B tương ứng với số lượng giỏ N_R , N_G và N_B . Do đó không gian màu RGB có tất cả $N (= N_R \times N_G \times N_B)$ bin màu. Các giỏ này được đánh thứ tự từ R qua G sau đó là từ G qua B. Theo như cách rời rạc và mã hoá như trình bày ở trên, chỉ số của mỗi giỏ có thể biểu diễn như công thức (3.2).

$$i = R + N_G \times G + N_B^2 \times B \quad (3.2)$$

trong đó, $R = 0, 1, \dots, N_R - 1$; $G = 0, 1, \dots, N_G - 1$ và $B = 0, 1, \dots, N_B - 1$.

Do đó điểm ảnh (r_p, g_p, b_p) sẽ thuộc về giỏ có chỉ số i_p được định nghĩa như công thức (3.3).

$$i_p = \left\lfloor \frac{r_p N_R}{256} \right\rfloor + N_G \times \left\lfloor \frac{g_p N_G}{256} \right\rfloor + N_B^2 \times \left\lfloor \frac{b_p N_B}{256} \right\rfloor \quad (3.3)$$

trong đó, $\lfloor \cdot \rfloor$ là phép toán lấy cận dưới trong số học.

3.2. Phát hiện đối tượng dựa trên giải thuật di truyền và tối ưu bầy đàn

Phân đoạn ảnh mờ (Fuzzy Segmentation)

Dựa trên định nghĩa về phân hoạch mờ như trình bày ở phần trước, luận văn tiến hành tiếp cận theo hướng phân đoạn màu sắc. Giải thuật phân đoạn bao gồm ba bước:

- i. Tiền phân cụm: Quá trình này bao gồm việc tìm kiếm tập vectơ trọng tâm khởi tạo U_0 và chỉ ra khoảng các vectơ trọng tâm được chọn theo một thủ tục tối ưu. Thủ tục này kết thúc bởi sử dụng kỹ thuật dựa trên histogram.
- ii. Tìm kiếm phân hoạch mờ tốt nhất. Giải thuật di truyền hoặc tối ưu bầy đàn được sử dụng để tìm một ma trận phân hoạch mờ tốt nhất.
- iii. Thủ tục giải mờ. Bước này tiến hành chuyển đổi ma trận phân hoạch mờ thành ma trận phân hoạch rõ.

Tiền phân cụm

Với một ảnh màu C cho trước, histogram màu $H(C)$ thu được bởi phương pháp trình bày như ở trên. Ta quan sát thấy rằng nếu một ảnh gồm nhiều đối tượng với màu sắc khác nhau, histogram màu của nó thường chứa các cực đại nhọn. Mỗi cực đại tương ứng với một đối tượng và các cực đại kề nhau bị chia cắt bởi các hõm. Chiều cao của cực đại chỉ ra số lượng điểm ảnh rơi vào giỏ tương ứng với vị trí của cực đại.

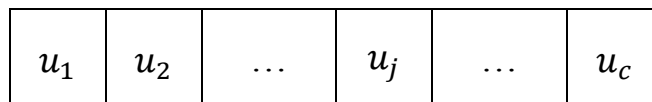
Sau khi số lượng cụm c được lựa chọn, thì c cực đại cao nhất của histogram màu được phát hiện và các giỏ tương ứng các cực đại đó xác định các khoảng trong đó các vectơ trọng tâm được tìm kiếm với mục đích tối ưu. Giá trị vectơ trọng tâm khởi tạo được lựa chọn ngẫu nhiên

trong mỗi c giỏ.

Để có được phân đoạn tốt nhất, ta có thể sử dụng giải thuật di truyền hoặc tối ưu bầy đàn để tìm được cách phân hoạch tối ưu. Giải thuật di truyền tối ưu phân đoạn mờ được thiết kế và sử dụng như trình bày sau.

Biểu diễn di truyền

Với một tập vector m -chiều cho trước, cá thể trong quần thể được biểu diễn bởi một nhiễm sắc thể (chromosome) – một chuỗi chứa c vector m -chiều số thực, nó mã hoá các vector trọng tâm tương ứng với c cụm trong phân hoạch mờ. Hình 3.2 là một ví dụ của một nhiễm sắc thể, trong đó cá thể U được mã hoá là một chuỗi c vector trọng tâm tương ứng.



Hình 3.2. Chuỗi chứa c vector

Khởi tạo quần thể

Với quá trình khởi tạo quần thể, chuỗi vector là những vector được chọn ngẫu nhiên từ mỗi giỏ trong c giỏ như phần trước đã đề cập. Lực lượng của quần thể N được lựa chọn bởi người sử dụng.

Hàm mục tiêu

Để sử dụng giải thuật di truyền, điều cần thiết là phải xác định một hàm mục tiêu. Trong luận văn này, quan hệ tương tự của tập vector trọng tâm với tất cả các vector trong tập vector được xét, được tính như công thức (3.4).

$$fitness = \frac{1}{c} \sum_{i=1}^c \sum_{j=1}^n p_{ij} \quad (3.4)$$

Mục tiêu của mỗi nhiễm sắc thể trong quần thể được tính dựa trên hàm mục tiêu. Với mỗi nhiễm sắc thể, các vector trọng tâm mã hoá trong nó

được tính đầu tiên, sau đó ma trận phân hoạch mờ tương ứng với nhiễm sắc thể được tính dựa trên công thức (2.3) như đã giới thiệu ở phần 2.1.

Các thao tác di truyền

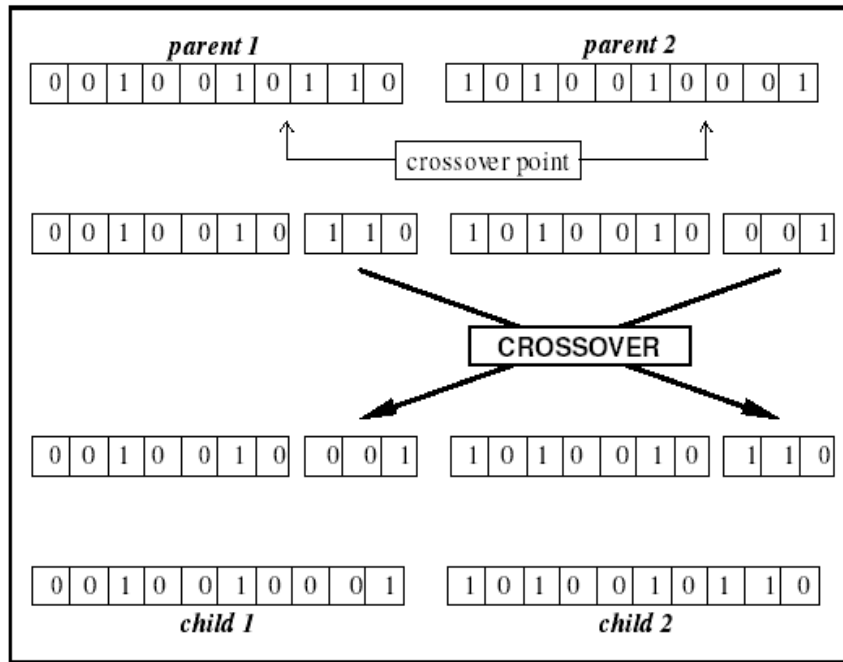
Có ba loại thao tác di truyền được sử dụng trong giải thuật di truyền bao gồm: chọn lọc, lai ghép và đột biến. Trong luận văn này, một vòng quay roulette thích hợp được sử dụng cho thao tác chọn lọc. Đây được coi là phương pháp chọn lọc đơn giản và hiệu quả, ở đây mỗi chuỗi (cá thể) trong quần thể chiếm một khe trong vòng tròn Roulette có độ rộng tỷ lệ với giá trị hàm mục tiêu của chuỗi. Mỗi lần quay vòng tròn Roulette chúng ta nhận được một chuỗi và coi như đó là cách lựa chọn chuỗi cho việc tái tạo.

Các bước thực hiện chọn lọc bao gồm:

- i. Tính tổng các giá trị mục tiêu của các cá thể trong quần thể và gán kết quả này vào biến τ (total fitness – tổng mục tiêu).
- ii. Ở thế hệ thứ n , lấy một số ngẫu nhiên giữa 0 và τ .
- iii. Trả về số cá thể đầu tiên của một dân số mới, dựa vào giá trị mục tiêu của nó.

Sau khi chọn lọc, lai ghép một điểm cắt được áp dụng (như minh họa Hình 3.3) với các cặp chuỗi được chọn để sinh ra các cặp chuỗi con. Thao tác lai ghép được áp dụng ngẫu nhiên với xác suất p_c .

Sau khi lai ghép, các xâu con được xem xét tiến hành đột biến. Trong giải thuật này, thao tác đột biến được cài đặt bằng cách thay thế các chuỗi trong thế hệ hiện tại bằng các chuỗi từ cùng giỏ trong tập chuỗi ban đầu. Thao tác đột biến được tiến hành với xác suất cố định là p_m .



Hình 3.3. Minh họa lại ghép một điểm cắt

Điều kiện dừng

Có hai điều kiện dừng được áp dụng cho giải thuật di truyền. Điều kiện thứ nhất là sau một số thế hệ không thay đổi trong quần thể lời giải thì giải thuật kết thúc. Điều kiện thứ hai là số lượng vòng lặp tối đa, sau đó giải thuật cũng kết thúc.

Tối ưu bầy đàn cho phân đoạn mờ

Ngoài giải thuật di truyền như trình bày ở trên, ta có thể sử dụng tối ưu bầy đàn (PSO) để tìm kiếm được cách phân hoạch tối ưu. Đầu tiên ta cũng tiến hành mã hoá lời giải (cách phân hoạch) tương tự như trong giải thuật di truyền, trong PSO được gọi là các phân tử (particle). Hàm mục tiêu (fitness function) của PSO cũng tương tự hàm mục tiêu của giải thuật di truyền đã trình bày ở trên. Mỗi phân tử được đặc trưng bởi hai tham số là vị trí hiện tại của phân tử $present[]$ và vận tốc $v[]$. Đồng thời mỗi phân tử có một giá trị thích nghi (fitness value), được đánh giá dựa trên hàm mục tiêu.

Tại thời điểm xuất phát, vị trí của mỗi phần tử được khởi tạo một cách ngẫu nhiên. Trong quá trình chuyển động, mỗi phần tử chịu ảnh hưởng bởi ba thông tin: thông tin thứ nhất, gọi là $pBest$, là vị trí tốt nhất mà phần tử đó đã đạt được trong quá khứ; thông tin thứ hai, gọi là $gBest$, là vị trí tốt nhất mà cả bầy đàn đã đạt được trong quá khứ; và thông tin cuối cùng là $lBest$, vị trí tốt nhất đã đạt được trong hàng xóm của nó. Các phần tử trong PSO sẽ duyệt không gian bài toán bằng cách theo sau các phần tử có điều kiện tốt nhất hiện thời (độ thích nghi lớn nhất). Cụ thể là sau mỗi khoảng thời gian rời rạc, vận tốc và vị trí của mỗi phần tử được cập nhật theo các công thức (2.10) và (2.11) như đã trình bày trong phần 2.3.

Các điều kiện dừng của giải thuật là: số lần cập nhật, số lần cập nhật bầy đàn mà không đưa lại kết quả tốt hơn, số lần cập nhật mà lượng thay đổi giữa hai lần cập nhật liên tiếp nhỏ hơn ngưỡng.

Giải mờ

Để thu được ảnh phân vùng, ta cần phải biến đổi ma trận phân hoạch mờ thành ma trận phân hoạch rõ (c-partition crisp matrix). Trong luận văn này, phương pháp giải mờ được áp dụng như trình bày dưới đây.

Đặt $P = [p_{ij}]$ với $i = 1, \dots, c$ và $j = 1, \dots, n$ là ma trận phân hoạch mờ, p_{ij} thể hiện khả năng điểm ảnh j thuộc về cụm i . Ma trận phân trăm P_p được tính như công thức (3.5).

$$p_{pij} = \frac{p_{ij}}{\sum_{j=1}^n p_{ij}} \quad (3.5)$$

Ma trận phân hoạch rõ P_c được định nghĩa là

$$p_{cij} = \begin{cases} 1, & p_{pij} = \max_{i=1..c} p_{pij} \\ 0, & \text{khác} \end{cases} \quad (3.6)$$

Rõ ràng việc biến đổi thành ma trận phân hoạch rõ, mỗi điểm ảnh thuộc

về một cụm cụ thể.

Trích xuất đối tượng

Khi ảnh phân vùng đã thu được bởi giải thuật phân vùng như trên, ảnh nhị phân đối tượng có thể được trích xuất bằng cách tô giả màu tương ứng với các vùng đối tượng. Thông thường, các đối tượng trong ảnh nhị phân bị lỗi bởi nhiễu đối tượng, những thứ có màu tương tự như các đối tượng. Để làm cho vùng đối tượng trở nên rõ ràng, ta cần phải lọc bỏ những vùng lỗi trên ảnh nhị phân đối tượng. Cuối cùng một thao tác hình thái học được áp dụng. Ví dụ, phụ thuộc vào hình dạng của đối tượng nhiễu, một tổ hợp thích hợp của các phép dilation (nở ra), erosion (xói mòn), opening (mở rộng) và closing (khép lại) sẽ được chọn.

Để trích xuất được vùng chứa đối tượng ta sử dụng đặc trưng màu như trình bày ở trên và một giải thuật trích xuất cạnh để trích xuất khung của đối tượng.

p_7	p_0	p_1
p_6	p	p_2
p_5	p_4	p_3

Hình 3.4. Bố trí lân cận của một điểm ảnh

Bố trí lân cận của một điểm ảnh được trình bày như trong hình 3.4, điểm ảnh ranh giới của một đối tượng được xác định nếu nó là một điểm ảnh viên và thoả mãn điều kiện sau $0 < N(p) < 8$, trong đó $N(p)$ là số lượng hàng xóm khác không của điểm ảnh được tính như công thức (3.7).

$$N(p) = \sum_{i=0}^7 p_i \quad (3.7)$$

3.3. Theo dõi đối tượng

Sau khi các đối tượng đã phát hiện, ta tiến hành theo vết đối tượng. Hầu hết các hệ thống hiện nay đều theo dõi đối tượng dựa vào bộ lọc Kalman. Ở đây tôi cũng sử dụng bộ lọc Kalman nhưng có một số sửa đổi. Như đã biết theo vết đối tượng có nghĩa là chúng ta phải phân tích và ghi lại quỹ đạo của đối tượng. Để có thể theo vết tốt chúng ta phải ước lượng được chuyển động của đối tượng. Mỗi đối tượng sẽ được đặc trưng bởi một số tham số sau:

- p = vị trí trong tọa độ ảnh. Tọa độ của trọng tâm của đối tượng sẽ chính là tọa độ của đối tượng.
- δp = là độ không chắc chắn của vị trí. Không bao giờ chúng ta có thể ước lượng chính xác được vì thế độ không chắc chắn này chính là độ sai số trong ước lượng.
- \vec{v} = tốc độ của đối tượng. Ở đây chúng ta sẽ dùng độ dịch chuyển thay cho tốc độ vì chúng ta đang xét là các điểm rời rạc chứ không phải các điểm liên tục.
- $\delta \vec{v}$ = độ không chắc chắn vận tốc.
- \mathcal{R} = hình chữ nhật bao quanh đối tượng.
- d = chỉ số tin cậy của đối tượng. Độ tin cậy này dùng để quyết định loại bỏ hay tiếp tục theo dõi đối tượng.

Dự đoán vị trí của đối tượng trong tương lai

Việc ước lượng vị trí của đối tượng trong mỗi khung hình là rất quan trọng. Nó giúp cho việc tính toán đơn giản hơn và việc theo vết sẽ chính xác hơn. Vị trí ước lượng sẽ được sử dụng để lựa chọn số các đối tượng cần phải xem xét trong quá trình theo vết. Việc ước lượng sẽ giảm được rất nhiều thời gian vì với mỗi đối tượng chúng ta không cần phải xem xét trên toàn bộ ảnh mà chỉ xét trong vùng được dự đoán mà thôi. Gọi Δt là

thời gian giữa hai khung hình, theo mô hình chuyển động tuyến tính thì ta có tọa độ khung hình thứ $n + 1$ như công thức (3.8).

$$p_{n+1} = p_n + v * \Delta t \quad (3.8)$$

Và độ không chắc chắn được cho là bằng độ không chắc chắn của vị trí hiện tại cộng với độ không chắc chắn của vận tốc nhân cho thời gian.

$$\delta p_{n+1} = \delta p_n + \delta v_n * \Delta t \quad (3.9)$$

Những vị trí này sẽ được chọn để ước lượng vị trí mới của đối tượng trong khung hình tiếp theo. Mỗi đối tượng trong khung hình hiện tại sẽ được dự đoán vị trí mới bằng cách rộng hình chữ nhật bao quanh đối tượng một khoảng là $v_n * \Delta t$ sau đó mở rộng ra một khoảng δp_{n+1} . Do đó nếu có vùng chuyển động mới nào thuộc vào vùng dự đoán này thì ta sẽ tiến hành so khớp.

So khớp đối tượng

Cho một vùng đối ở khung hình hiện tại, chúng ta tìm vùng khớp với vùng này ở khung hình tiếp theo bằng phương pháp so khớp tương quan ảnh. Hàm tương quan giữa vùng M ở khung hình hiện tại với vùng ứng viên ở khung hình tiếp theo và cách vùng này một khoảng d được tính như công thức (3.10).

$$C(d) = \sum_{x \in R} \frac{W(i, j) |I_n(x) - I_{n-1}(x + d)|}{\|W\|} \quad (3.10)$$

trong đó, I_n là mật độ điểm ảnh tại khung ảnh hiện tại, I_{n+1} là mật độ điểm ảnh tại khung hình tiếp theo. W là hàm trọng số và $\|W\|$ được tính bằng công thức (3.11).

$$\|W\| = \sum_{x \in R} W(x) \quad (3.11)$$

Ở đây, $W(x)$ là hàm trọng số thể hiện mức độ ưu tiên của điểm x . Các

đối tượng mà chúng ta quan sát được có thể di chuyển từ xa tới hoặc đi ra xa. Khi đó kích thước của chúng ở khung hình tiếp theo sẽ lớn hơn hoặc nhỏ hơn. Vì thế hàm trọng số này phải được tính làm sao để các điểm ở gần tâm của đối tượng sẽ có độ ưu tiên cao hơn. Và để giảm chi phí tính toán thì hàm trọng số này chỉ áp dụng đối với các điểm ảnh nào được xem là chuyển động, do đó các điểm được xem là đứng yên sẽ có hàm trọng số bằng 0. Nếu x là điểm chuyển động $W(x)$ được tính như công thức (3.12).

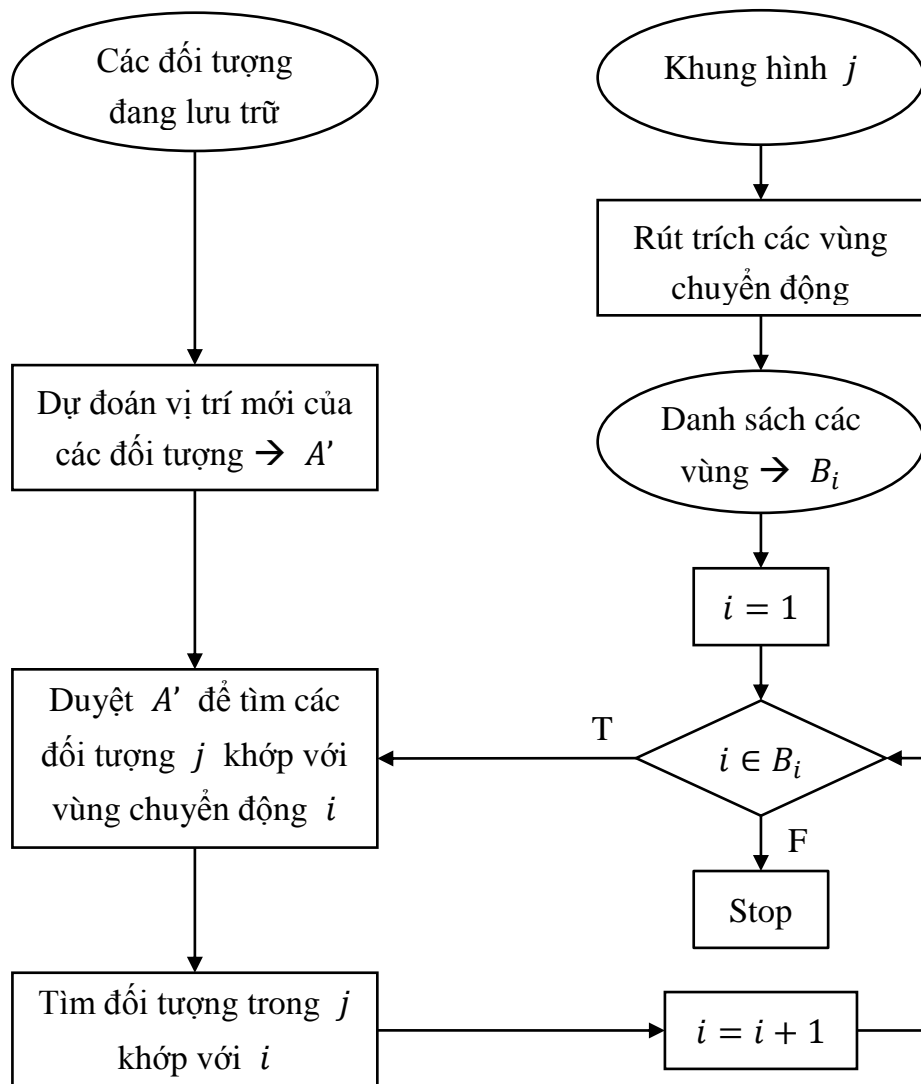
$$W(x) = \frac{1}{2} + \frac{1}{2} \left(1 - \frac{r(x)}{r_{max}} \right) \quad (3.12)$$

trong đó, $r(x)$ là khoảng cách bán kính từ x tới tâm của vùng \mathcal{R} và r_{max} là khoảng cách bán kính lớn nhất trong \mathcal{R} . Để giảm chi phí tính toán ta có thể ước lượng tâm của đối tượng chính là tâm của hình chữ nhật bao quanh đối tượng và khoảng cách lớn nhất trong vùng \mathcal{R} bằng một nửa đường chéo của hình chữ nhật bao quanh đối tượng. Khi so khớp vùng đối tượng có kích thước $m \times n$ thì chi phí tính toán sẽ là $O(m^2n^2)$. Nếu kích thước của vùng đối tượng lớn thì chi phí sẽ tăng lên rất lớn khó mà kiểm soát được. Vì thế chúng tôi đã cố định một ngưỡng mà chi tính toán sẽ không vượt qua ngưỡng đó. Ở đây chúng tôi chọn ngưỡng là 25 điểm ảnh. Kích thước của vùng ảnh theo mỗi chiều sẽ bị giảm đi một nửa nếu nó lớn hơn 25 và chúng tôi sẽ đếm số lần giảm kích thước để lựa chọn các điểm khi so khớp. Ví dụ như một vùng có kích thước 80×45 thì kích thước sẽ giảm đi hai lần theo chiều x và một lần theo chiều y . Do đó trong quá trình so khớp, theo chiều x rõ bốn điểm thì sẽ chọn một điểm để so khớp và cứ hai điểm theo chiều y thì sẽ chọn một điểm.

Theo vết đối tượng

Theo vết đối tượng thì rộng lớn hơn là so khớp đối tượng. Chúng ta phải lưu vết được quỹ đạo chuyển động của đối tượng. Ý chính ở đây là so

khớp đối tượng có được với các đối tượng mà ta thu được ở khung hình tiếp theo. Các đối tượng ở khung hình tiếp theo sau khi qua bộ phát hiện chuyển động sẽ rút trích ra được các vùng chuyển động. Chúng ta sẽ tạo một danh sách các vùng chuyển động để thuận tiện cho việc theo vết. Sau đó chúng ta ước lượng vị trí mới của các đối tượng ở khung hình hiện tại. Khi đó việc so khớp sẽ dễ dàng hơn. Chúng ta chỉ so khớp các đối tượng với vùng chuyển động mới mà vị trí ước lượng của các đối tượng này thuộc vào vùng chuyển động mới đó. Tùy trường hợp so khớp mà chúng ta có các quyết định theo vết khác nhau. Hình 3.5 là sơ đồ thuật giải từng bước theo vết đối tượng.



Hình 3.5. Sơ đồ từng bước giải thuật theo vết đối tượng

Khi so khớp các đối tượng j với đối tượng i sẽ xảy ra các trường hợp sau đây:

- Không có đối tượng nào khớp với vùng chuyển động hết. Có hai trường hợp xảy ra: một là có chuyển động mới vào vùng quan sát (khi đó ta tạo một đối tượng mới, thiết lập các tham số và cho độ tin cậy ở mức thấp nhất); hai là dự đoán sai (khi này ta có thể mở rộng vùng tìm kiếm và so khớp đối tượng)
- Chỉ có một đối tượng khớp với vùng chuyển động. Đây là trường hợp tốt nhất để theo vết. Trường hợp này ta sẽ cập nhật đối tượng theo vùng chuyển động mới này.
- Một đối tượng khớp với nhiều vùng chuyển động. Điều này có thể do đối tượng tách ra thành nhiều đối tượng. Trường hợp này ta sẽ cập nhật đối tượng hiện tại theo đối tượng khớp nhất. Các đối tượng còn lại sẽ tương ứng với đối tượng mới và chúng ta thiết lập các tham số ban đầu cho đối tượng.
- Trường hợp nhiều đối tượng cùng khớp với một vùng chuyển động. Trường hợp này xảy ra do các đối tượng giao nhau.
- Trường hợp các đối tượng không khớp với vùng chuyển động nào. Trường hợp này có thể do đối tượng ra khỏi vùng quan sát hoặc là đối tượng này đã bị đối tượng khác che khuất, vì thế độ tin cậy của đối tượng sẽ bị giảm xuống. Khi độ tin cậy giảm xuống dưới ngưỡng thì ta sẽ loại bỏ đối tượng khỏi danh sách.

Cập nhật đối tượng

Các tham số của đối tượng sẽ được cập nhật theo các tham số của đối tượng mới. Gọi d là độ dịch chuyển của đối tượng giữa hai khung hình I_n và I_{n+1} , d được tính như công thức (3.13).

$$d = \frac{p_{n+1} - p_n}{\Delta t} \quad (3.13)$$

Với p_{n+1} là vị trí mới của đối tượng.

Khi đó ta có vận tốc mới của đối tượng v_{n+1} được cập nhật như công thức (3.14).

$$v_{n+1} = \alpha * d + (1 - \alpha) * v_n \quad (3.14)$$

Và độ không chắc chắn của vận tốc δv_{n+1} được cập nhật như công thức (3.15).

$$\delta v_{n+1} = \alpha * |v_{n+1} - d| + (1 - \alpha) * v_n \quad (3.15)$$

CHƯƠNG 4: CÀI ĐẶT THỬ NGHIỆM

Phần này luận văn trình bày cách thức cài đặt và kết quả thử nghiệm đạt được. Luận văn này tập trung vào nhận dạng và theo dấu đối tượng cụ thể là các phương tiện giao thông từ ảnh thu được từ các thiết bị bay UAV, drone. Đây là đối tượng phổ biến và có nhiều ứng dụng thực tế trong an ninh và quốc phòng (ví dụ, giám sát giao thông tự động). Các thử nghiệm tương tự cũng có thể dễ dàng mở rộng ra các đối tượng khác.

4.1. Công cụ hỗ trợ

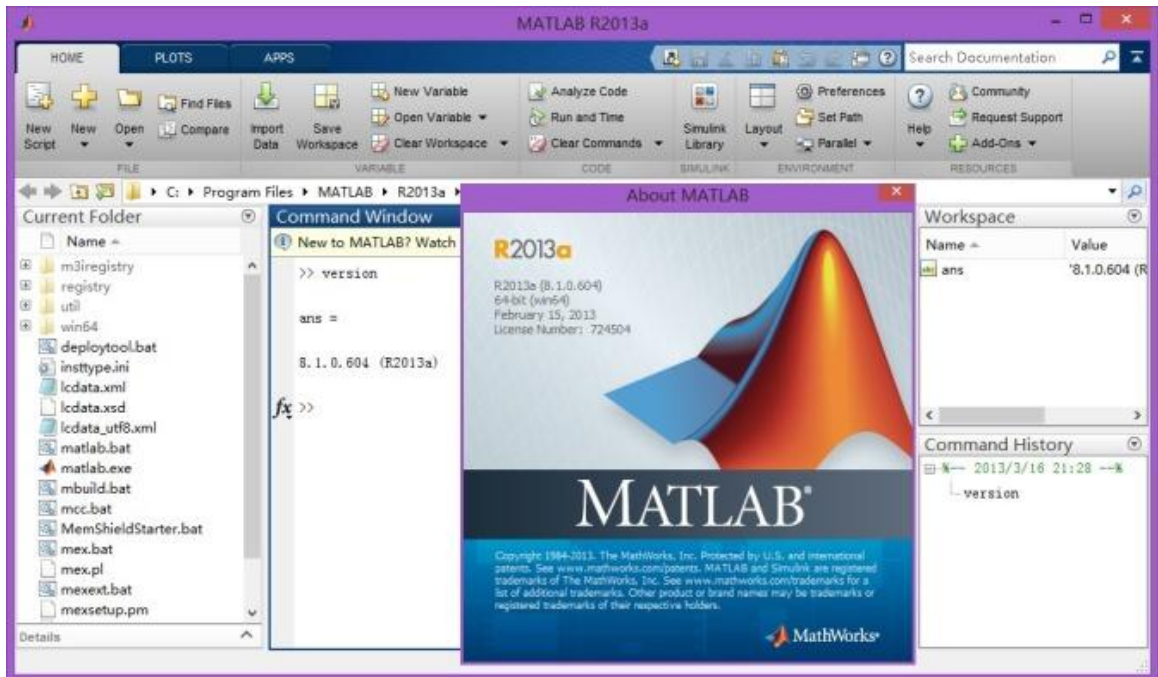
Các thử nghiệm trong luận văn này được cài đặt sử dụng ngôn ngữ và môi trường Matlab, chi tiết như trong Bảng 4.1.

Bảng 4.1. Cấu hình và môi trường thử nghiệm

Cấu hình	Chi tiết
Phần cứng	Dell Optiplex 7020, CPU Intel Core i7-4790 @ 8×3.60Ghz, RAM DDR3 16GB
Hệ điều hành	Window 7 SP1 Professional 64-bit
Ngôn ngữ & môi trường	Matlab 2013a 64-bit

MATLAB¹ là phần mềm cung cấp môi trường tính toán số và lập trình có tính trực quan cao, do công ty MathWorks thiết kế. MATLAB cho phép tính toán số với ma trận, vẽ đồ thị hàm số hay biểu đồ thông tin, thực hiện thuật toán, tạo các giao diện người dùng và liên kết với những chương trình máy tính viết trên nhiều ngôn ngữ lập trình khác.

¹ https://www.mathworks.com/products/?s_tid=gn_ps

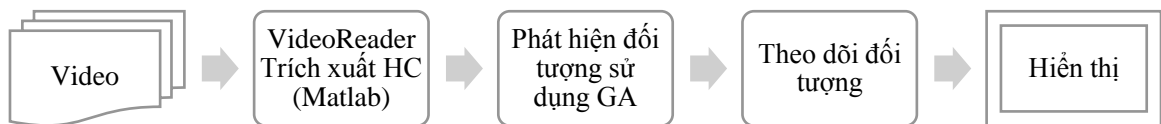


Hình 4.1. Giao diện làm việc với Matlab

Hiện nay, MATLAB có đến hàng ngàn lệnh và hàm tiện ích. Ngoài các hàm cài sẵn trong chính ngôn ngữ, MATLAB còn có các lệnh và hàm ứng dụng chuyên biệt trong các hộp công cụ (Toolbox), để mở rộng môi trường MATLAB nhằm giải quyết các bài toán thuộc các phạm trù riêng. Các công cụ khá quan trọng và tiện ích cho người dùng như toán sơ cấp, xử lý tín hiệu số, xử lý ảnh, xử lý âm thanh, ma trận thưa, logic mờ, v.v.

Trong luận văn, tôi đã cài đặt công cụ theo dõi đối tượng sử dụng giải thuật di truyền (gọi là GAObjectTracking – Hình 4.2). Đầu vào của công cụ này là đoạn video chứa các đối tượng cần theo dõi. Đầu tiên, đoạn video sẽ được giải mã, giải nén thành các khung hình liên tiếp nhau. Với bước này, tôi sử dụng hàm VideoReader() được hỗ trợ trong “Audio and Video Toolbox” của Matlab, giúp giải mã được phần lớn các định dạng video phổ biến hiện nay như: AVI (Audio Video Interleave), MP4 (MPEG-4 Part 14), WMV (Windows Media Video) và v.v. Tiếp theo, tôi tiến hành trích rút đặc trưng histogram màu dựa trên công cụ xử lý ảnh của Matlab. Sau đó, các đặc trưng được đưa vào mô-đun phát hiện đối

tượng. Như trình bày trong chương 3, tôi đề xuất hai hướng tìm phân đoạn đối tượng tối ưu là giải thuật di truyền và tối ưu bầy đàn. Trong công cụ GAObjectTracking tôi tiến hành cài đặt tối ưu phát hiện đối tượng sử dụng giải thuật di truyền. Đối với hướng sử dụng phương pháp tối ưu bầy đàn, do thời gian nghiên cứu có hạn nên tôi chưa thể cài đặt thử nghiệm thành công. Đây cũng là một trong các hướng phát triển giúp hoàn thiện đề tài trong tương lai của tôi. Đầu ra của mô-đun phát hiện đối tượng là một danh sách các đối tượng chứa trong khung hình. Danh sách này được đưa vào đối sánh và theo dõi trong mô-đun theo dõi đối tượng. Các đối tượng theo dõi được đánh dấu và hiển thị trực quan trên giao diện Matlab.



Hình 4.2. Cấu trúc công cụ thử nghiệm GAObjectTracking

4.2. Dữ liệu thử nghiệm

Dữ liệu tôi tiến hành thử nghiệm được từ thiết bị bay UAV quay một đoạn đường cao tốc. Dữ liệu được lưu vào máy dưới định dạng AVI nén, với các thông số như sau:

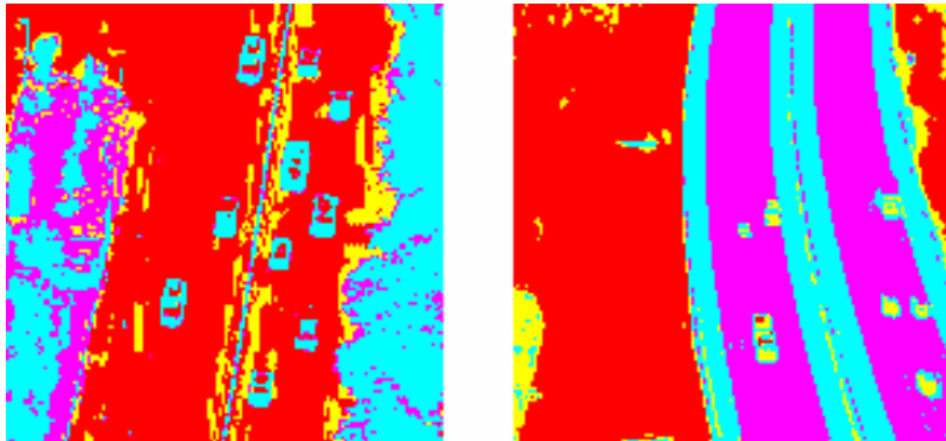
- Chuẩn nén là Cinepak Codec.
- Kích thước của AVI là 56.5MB.
- Thời gian của đoạn phim là 13 phút 07 giây.
- Tốc độ khung hình là 30fps.
- Kích thước khung hình là 320×320 điểm ảnh.

4.3. Kết quả thử nghiệm



Hình 4.3. Khung hình thử nghiệm

Trong bước phân vùng ảnh, tham số k_1 và k_2 trong công thức (2.4) để tính toán độ tương tự màu sắc được chọn lần lượt là 0.0001 và 0.2. Số lượng cụm được chọn là 4 và số vòng lặp tối đa là 50. Xác suất lai ghép và đột biến là 20%. Hình 4.4 là ảnh sau khi được phân vùng với giải thuật phân vùng cùng các tham số như trên.



Hình 4.4. Ảnh phân vùng màu

Với cách tiếp cận trích rút đối tượng như trình bày ở phần trước, tôi sử dụng phân phối để phân tách đối tượng đường và phương tiện giao thông. Trong trường hợp này, cụm đường được chọn và phương tiện giao thông có thể coi như “nhiều”. Hình 4.5 là ảnh nhị phân chứa đối tượng đường.



Hình 4.5. Ảnh nhị phân đường

Ảnh nhị phân đường trên được lọc bởi bộ lọc hình thái học mở rộng để thu được đối tượng đường liền mạch (theo kinh nghiệm thành phần cấu trúc được đặt là 5). Kết quả của thao tác hình thái học như hình 4.6.



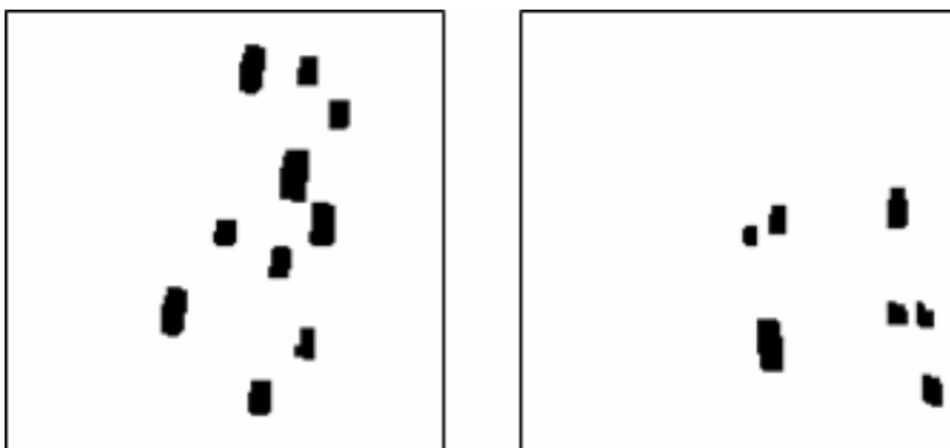
Hình 4.6. Ảnh nhị phân đường sau khi lọc

Để trích xuất được phương tiện giao thông từ ảnh hình 4.5, ta tiến hành trừ hình 4.5 cho hình 4.6. Hình 4.7 là kết quả trích xuất phương tiện giao thông bởi phép trừ.



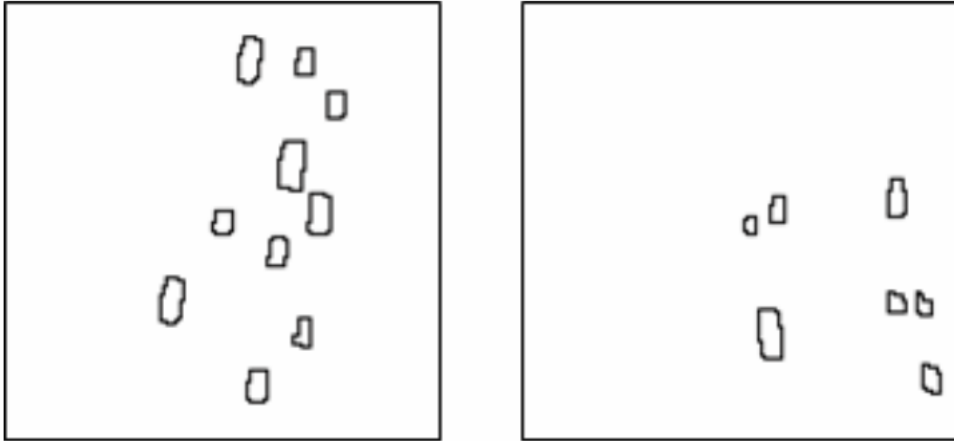
Hình 4.7. Ảnh nhị phân phương tiện giao thông

Thao tác hình thái học tiếp tục được sử dụng với ảnh phương tiện giao thông để thu được đối tượng phương tiện giao thông hoàn chỉnh như hình 4.8.



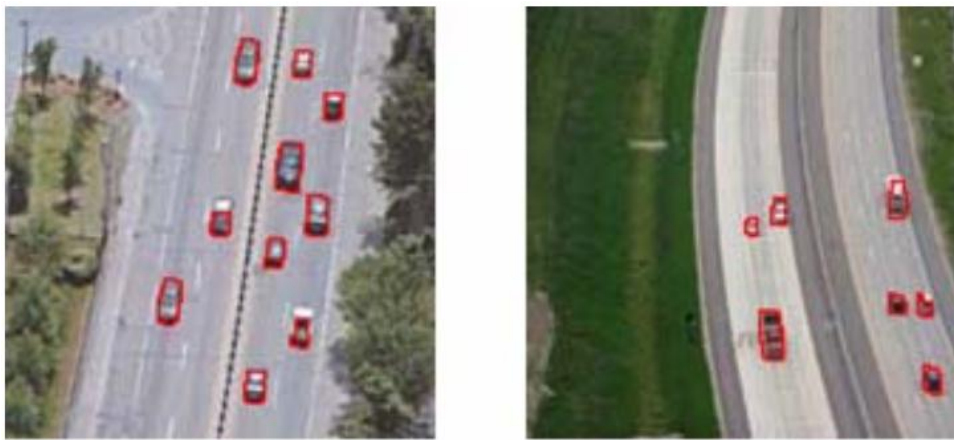
Hình 4.8. Ảnh nhị phân phương tiện giao thông sau khi lọc

Các cạnh của đối tượng phương tiện giao thông được trích xuất sử dụng giải thuật tìm đường bao trình bày ở phần 3 và kết quả như hình 4.9.



Hình 4.9. Bao của đối tượng trích xuất được

Để minh họa cho tính chính xác, đường bao của đối tượng phương tiện giao thông trích xuất được ghép vào ảnh màu gốc như Hình 4.10. Trong hình được ghép các đường màu đỏ là bao của đối tượng. Có thể quan sát từ Hình 4.10 rằng phần lớn đường bao khớp chính xác với các phương tiện giao thông.



Hình 4.10. Biểu diễn trên ảnh gốc

Để đánh giá độ chính xác của việc theo dấu đối tượng. Ta tiến hành đánh giá kết quả theo độ chính xác theo dõi đa đối tượng (Multiple Object Tracking Accuracy - MOTA) [27].

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \quad (4.1)$$

trong đó, m_t , fp_t và mme_t lần lượt là số lượt theo dõi trượt (misses), sai tích cực (false positives) và số khớp sai (mismatches) tương ứng với khung hình t , và g_t là tổng số đối tượng có trong khung hình t . Độ đo MOTA có thể coi là tổng hợp của ba độ lỗi: tỉ lệ trượt (misses rate - \bar{m}), tỉ lệ sai tích cực (false positives rate - \overline{fp}) và tỉ lệ khớp sai (mismatches rate - \overline{mme}).

$$\bar{m} = \frac{\sum_t m_t}{\sum_t g_t} \quad (4.2)$$

$$\overline{fp} = \frac{\sum_t fp_t}{\sum_t g_t} \quad (4.3)$$

$$\overline{mme} = \frac{\sum_t mme_t}{\sum_t g_t} \quad (4.4)$$

Bảng 4.2. Kết quả theo dõi đối tượng theo độ đo MOTA

STT	Đánh giá	Kết quả
1	\bar{m}	5.73%
2	\overline{fp}	5.78%
3	\overline{mme}	2.53%
4	MOTA	85.96%

Kết quả theo dõi đối tượng tính theo các độ đo MOTA, tỉ lệ trượt, tỉ lệ sai tích cực và tỉ lệ khớp sai được ghi nhận ở Bảng 4.2. Tỉ lệ trượt thể hiện phần trăm các đối tượng có trong thực tế nhưng không được theo dõi. Tỉ lệ sai tích cực thể hiện phần trăm các đối tượng xuất hiện trong theo dõi nhưng không có trong thực tế. Tỉ lệ khớp sai thể hiện phần trăm số lần đối tượng bị thay đổi sai so với khung hình trước đó. Bảng 4.2 cho thấy

chương trình thử nghiệm cho kết quả rất tốt, các lỗi theo dõi trượt, theo dõi sai tích cực và so khớp sai đều tương đối nhỏ ($<8\%$). Đặc biệt lỗi so khớp sai là rất nhỏ (2.53%), cho thấy giải thuật đã xử lý các nhập nhằng tốt giữa các đối tượng trong quá trình theo dõi. Kết quả theo dõi đối tượng với độ đo tổng hợp MOTA tương đối cao (85.96%), cho thấy chương trình thử nghiệm đã nhận dạng và theo dõi đối tượng là phương tiện giao thông với độ chính xác cao.

CHƯƠNG 5: KẾT LUẬN

Trong bối cảnh các phần mềm trí tuệ nhân tạo được phát triển mạnh mẽ, áp dụng rộng rãi trong nhiều lĩnh vực cuộc sống. Nhằm bắt được điều đó, cùng với mục tiêu làm chủ công nghệ, áp dụng trí tuệ nhân tạo vào các ứng dụng phục vụ an ninh quốc phòng tại Việt Nam, tôi đã lựa chọn đề tài ***“Theo dõi đối tượng dựa trên giải thuật di truyền và tối ưu hoá bầy đàn”***.

Trong luận văn, tôi đã trình bày đề xuất áp dụng giải thuật di truyền và tối ưu bầy đàn giải quyết bài toán theo dõi đối tượng. Giải thuật theo dõi bao gồm ba bước cơ bản: trích xuất đặc trưng, phát hiện đối tượng và theo dấu đối tượng. Ở bước đầu tiên, histogram màu, một đặc trưng đơn giản, phổ biến và hiệu quả được trích rút, đưa ra các thông tin về màu sắc thể hiện trong ảnh. Tiếp theo, giải thuật tiến hành phát hiện các đối tượng chứa trong khung hình dựa trên giải thuật phân đoạn ảnh mờ, một phát triển của giải thuật phân hoạch mờ. Để chọn được cách phân đoạn tốt nhất, hay cách phân hoạch tối ưu, giải thuật di truyền và tối ưu bầy đàn được áp dụng. Các cách phân hoạch được mã hoá thành các cá thể, phát triển nhiều thế hệ và qua chọn lọc tự nhiên trả về lời giải tối ưu. Sau khi đã được phát hiện trong khung hình, các đối tượng được theo dõi qua các khung hình dựa trên bộ lọc Kalman theo các giả thiết đối tượng thay đổi về vị trí, hướng và tốc độ nhỏ qua các khung hình liên tiếp nhau.

Về mặt thực nghiệm, tôi đã cài đặt thành công công cụ GAObjectTracking cho bài toán theo dõi đối tượng áp dụng giải thuật di truyền. Công cụ được thử nghiệm với dữ liệu quay từ thiết bị bay UAV, theo dõi các đối tượng là phương tiện giao thông trên đường. Các đối tượng trong video, qua công cụ GAObjectTracking cho cảm nhận trực quan, đã được phát hiện và theo dõi tương đối chính xác. Ngoài ra, công cụ còn được tiến hành đánh giá dựa trên độ đo MOTA, cho độ chính xác 85.96% rất ấn tượng. Như vậy luận văn đã giải quyết được bài toán ***“theo dõi đối tượng”*** đã đề ra với hiệu quả tốt.

Hướng phát triển

Luận văn đã trình bày các kiến thức cơ bản về nhận dạng và theo dõi đối tượng, đề xuất hai phương pháp tối ưu là giải thuật di truyền và tối ưu bày đàn cho bài toán. Với giải thuật di truyền đã được áp dụng để thu được phân hoạch tối ưu, tuy nhiên các phép lai ghép, chọn lọc áp dụng là phổ thông, các tham số thử nghiệm chủ yếu dựa trên kinh nghiệm. Với phương pháp tối ưu bày đàn, do thời gian có hạn, nên giải pháp này chưa được cài đặt vào công cụ thử nghiệm.

Hướng nghiên cứu tiếp theo của tôi là tiếp tục nghiên cứu các giải pháp giải thuật di truyền và tối ưu bày đàn: mở rộng phương pháp lai ghép và chọn lọc cho giải thuật di truyền; cài đặt và tinh chỉnh mô hình giải pháp tối ưu bày đàn cho bài toán theo dõi đối tượng.

TÀI LIỆU THAM KHẢO

- [1] P. Simon (2013), “Too Big to Ignore: The Business Case for Big Data”, *vol. 72, John Wiley & Sons.*
- [2] D. A. Forsyth and J. Ponce (2002), “Computer vision: a modern approach”, *Prentice Hall Professional Technical Reference.*
- [3] A. Yilmaz, O. Javed and M. Shah (2006), “Object tracking: A survey”, *Acm computing surveys (CSUR)*, vol. 38, pp. 13.
- [4] K. G. Derpanis (2004), “The harris corner detector”, *York University.*
- [5] D. G. Lowe (1999), “Object recognition from local scale-invariant features”, in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on.*,
- [6] Y. Cheng (1995), “Mean shift, mode seeking, and clustering”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, pp. 790-799.
- [7] D. M. Greig, B. T. Porteous and A. H. Seheult (1989), “Exact maximum a posteriori estimation for binary images”, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 271-279.
- [8] M. Kass, A. Witkin and D. Terzopoulos (1988), “Snakes: Active contour models”, *International journal of computer vision*, vol. 1, pp. 321-331.
- [9] H. Permuter, J. Francos and I. H. Jermyn (2003), “Gaussian mixture models of texture and colour for image database retrieval”, in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on.*
- [10] R. Li, Y. Chen and X. Zhang (2006), “Fast robust eigen-background updating for foreground detection”, in *2006 International Conference on*

Image Processing.

- [11] K. Toyama, J. Krumm, B. Brumitt and B. Meyers (1999), “Wallflower: Principles and practice of background maintenance”, in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*.
- [12] A. Mittal and N. Paragios (2004), “Motion-based background subtraction using adaptive kernel density estimation”, in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*.
- [13] L. Zhao and C. E. Thorpe (2000), “Stereo-and neural network-based pedestrian detection”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, pp. 148-154.
- [14] C. Stauffer and W. E. L. Grimson (1999), “Adaptive background mixture models for real-time tracking”, in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*.
- [15] N. Bouaynaya, W. Qu and D. Schonfeld (2005), “An Online Motion-Based Particle Filter for Head Tracking Applications”, in *ICASSP (2)*.
- [16] L. A. Zadeh (1965), “Fuzzy sets”, *Information and control*, vol. 8, pp. 338-353.
- [17] G. J. Klir and B. Yuan (1995), “Fuzzy Sets and Fuzzy Logic: Theory and Applications”, *Upper Saddle River, NJ, USA: Prentice-Hall, Inc*.
- [18] J. H. Holland (1975), “Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence”, *U Michigan Press*.
- [19] J. D. Bagley (1967), “The behavior of adaptive systems which employ genetic and correlation algorithms”.

- [20] D. E. Goldberg (2006), “Genetic algorithms”, *Pearson Education India*.
- [21] J. M. Fitzpatrick and J. J. Grefenstette (1988), “Genetic algorithms in noisy environments”, *Machine learning*, vol. 3, pp. 101-120.
- [22] J. Kennedy, J. F. Kennedy, R. C. Eberhart and Y. Shi (2001), “Swarm intelligence”, *Morgan Kaufmann*.
- [23] J. Kennedy (2011), “Particle swarm optimization”, in *Encyclopedia of machine learning*, Springer, pp. 760-766.
- [24] M. Clerc (2010), “Particle swarm optimization”, vol. 93, *John Wiley & Sons*.
- [25] M. Zambrano-Bigiarini, M. Clerc and R. Rojas (2013), “Standard particle swarm optimisation 2011 at cec-2013: A baseline for future pso improvements”, in *2013 IEEE Congress on Evolutionary Computation*.
- [26] W. K. Pratt (1991), “Image segmentation”, *Digital Image Processing: PIKS Inside, Third Edition*, pp. 551-587.
- [27] K. Bernardin, A. Elbs and R. Stiefelhagen (2006), “Multiple object tracking performance metrics and evaluation in a smart room environment”, in *Sixth IEEE International Workshop on Visual Surveillance, in conjunction with ECCV*.