

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

CHƯƠNG I

ĐẶT VẤN ĐỀ

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. 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.

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*”. Luận văn này được cấu trúc các phần như sau. Chương 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 cụm fuzzy c-partition và giải thuật di truyền. Chương ba 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 bốn. Và cuối cùng các kết luận được đưa ra ở chương năm

CHƯƠNG II

CƠ SỞ LÝ THUYẾT

2.1. Fuzzy c-partition

Fuzzy C-Means (FCM) là một phương pháp của phân nhóm cho phép một phần dữ liệu thuộc về hai hoặc nhiều cụm

FCM được thực hiện lần lượt theo các bước:

- Bước 1: Khởi tạo ma trận $U=[u_{ij}]$, $U(0)$
- Bước 2: Tại lần lặp thứ k: tính toán véc-tơ trung tâm $C(k)=[c_j]$ với $U(k)$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m}$$

- Bước 3: Cập nhật $U(k)$ và $U(k+1)$

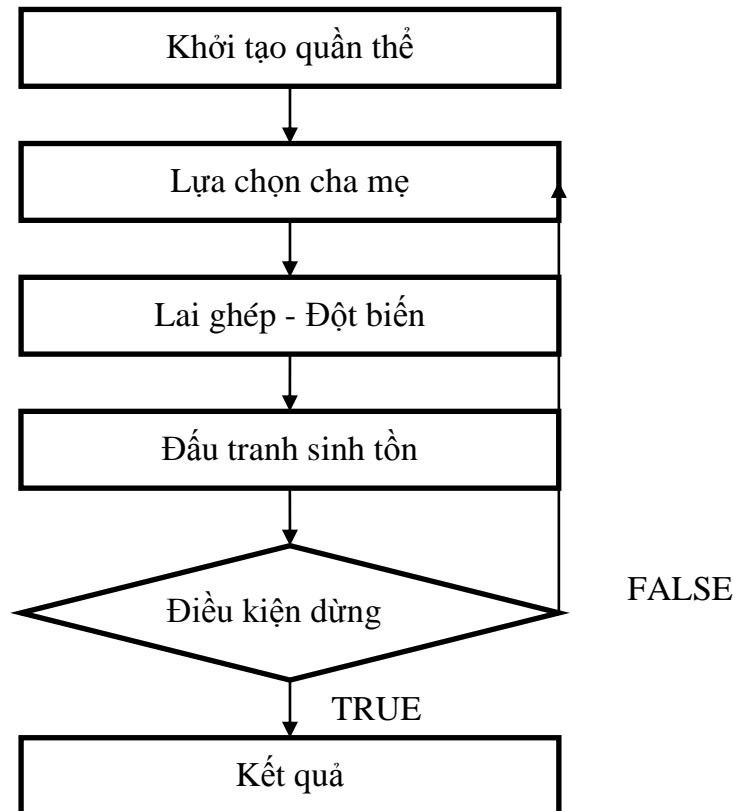
$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

- Bước 4: Kiểm tra

$$\|U^{k+1} - U^k\| < \varepsilon \left(\max_{ij} \|u_{ij}^{k+1} - u_{ij}^k\| \right)$$

Nếu kết vẫn chưa thỏa, ta quay lại bước 2, nếu đã thỏa mãn, ta kết thúc tính toán.

2.2. Giải thuật di truyền



Hình 1. Sơ đồ chung giải thuật di truyền

1. Khởi tạo một quần thể ban đầu (tập lời giải ban đầu của bài toán).
2. 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.
3. 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.

4. 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.

CHƯƠNG III

ÁP DỤNG GIẢI THUẬT DI TRUYỀN TRONG BÀI TOÁN THEO DÕI ĐỐI TƯỢNG

Bài toán “theo dõi đối tượng” được chia thành hai phần.

- Rút trích đối tượng. Trong phần này ta phải tách các đối tượng chuyển động ra khỏi khung cảnh.
- Theo vết đối tượng. Với từng đối tượng ta phải lưu lại được quỹ đạo của chúng khi chúng vào vùng quan sát của ta về phân loại đối tượng.

Trong luận văn, tôi sử dụng colour histogram là đặc trưng ảnh để thực hiện rút trích đối tượng.

3.1. Colour histogram

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 (Pratt, 1991). Thông thường, một colour histogram biểu diễn phân bố thống kê của các màu trong một ảnh màu trên một không gian cho trước.

3.2. Áp dụng giải thuật di truyền tối ưu trích xuất đối tượng

Dựa trên định nghĩa về fuzzy c-partition, tôi 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 vector

trọng tâm khởi tạo U_0 và chỉ ra khoảng các vector 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 fuzzy c-partition tốt nhất. Giải thuật di truyền được sử dụng để tìm một ma trận fuzzy c-partition tốt nhất.
- iii. Thủ tục giải mờ. Bước này tiến hành chuyển đổi ma trận fuzz c-partition thành ma trận c-partition rõ ràng.

Để tìm kiếm cách phân đoạn tối ưu của fuzzy c-partition, một giải thuật di truyền được sử dụng và thiết kế như sau.

Biểu diễn di truyền

Với một tập vector m-chiều cho trước, cá thể trong quần thể được biểu diễn bởi một chromosome – một chuỗi chứa c vector m-chiều số thực, nó mã hoá các vector trọng tâm tương ứng với c cụm trong c-partition. **Khởi tạo quần thể**

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

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 sau

$$\frac{1}{c} \sum_{i=1}^c \sum_{j=1}^n p_{ij}$$

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. Sau đó lai ghép một điểm cắt được áp dụng với hai chuỗi được chọn để sinh ra hai 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 bin trong tập chuỗi ban đầu. Thao tác đột biến được tiến hành với xác suất cố định là p_m .

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.

Giải mờ

Để thu được ảnh phân vùng, ta cần phải biến đổi ma trận fuzzy c-partition thành ma trận c-partition rõ ràng. 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 fuzzy c-partition, p_{ij}

thể hiện khả năng pixel j thuộc về cụm i . Ma trận phân trăm P_p được tính

như sau

$$p_{pij} = \frac{p_{ij}}{\sum_{j=1}^n p_{ij}}$$

Ma trận partition crisp P_c được định nghĩa là

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

Rõ ràng việc biến đổi thành ma trận partition crisp, mỗi pixel 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. **Theo dõi đối tượng**

Sau khi các đối tượng đã phát hiện, ta tiến hành theo vết đối tượng. Ở đây tôi cũng sử dụng bộ lọc Kalman nhưng có một số sửa đổi. Mỗi đối tượng sẽ được đặc trưng bởi một số tham số sau:

1. p = Vị trí trong tọa độ ảnh. Tọa độ của trọng tâm của đối tượng sẽ chính là tọa độ của đối tượng.
2. 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.
3. \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.
4. $\delta \vec{v}$ = độ không chắc chắn vận tốc.
5. \mathcal{R} = hình chữ nhật bao quanh đối tượng.
6. 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

Gọi At là thời gian giữa 2 frame, theo mô hình chuyển động tuyến tính thì ta có

$$p_{n+1} = p_n + v * At$$

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.

$$\delta p_{n+1} = \delta p_n + \delta v_n * At$$

Những vị trí này sẽ được chọn để ước lượng vị trí mới của đối tượng trong frame tiếp theo. Mỗi đối tượng trong frame hiện tại sẽ được dự đoán vị trí mới bằng cách rộng hình chữ nhật bao quanh đối tượng một khoảng là

$v_n * At$ 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 frame hiện tại, chúng ta tìm vùng khớp với vùng này ở frame tiếp theo bằng phương pháp so khớp tương quan ảnh. Hàm tương quan giữa vùng M ở frame hiện tại với vùng ứng viên ở frame tiếp theo và cách vùng này một khoảng d là:

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

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

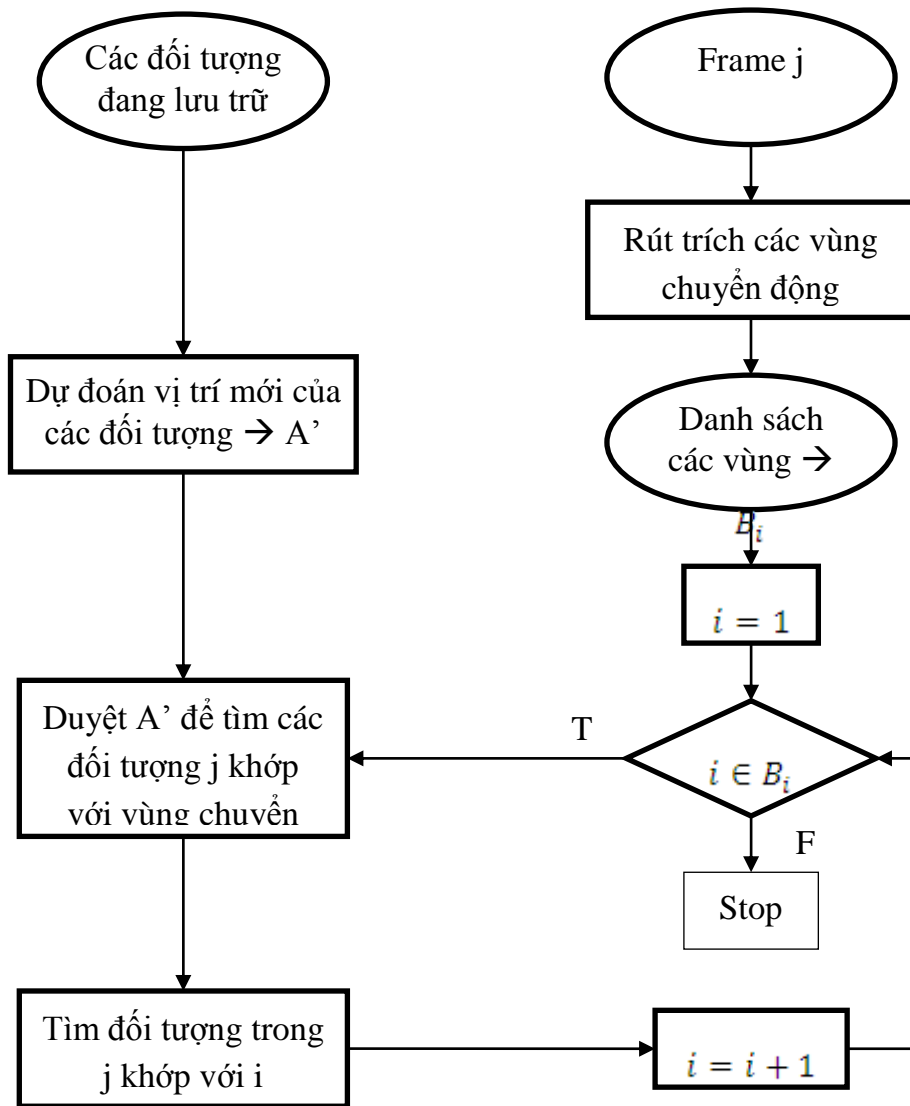
$$\|W\| = \sum_{x \in \mathcal{R}} W(x)$$

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

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 $O(m^2 n^2)$. 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 2 lần theo chiều x và 1 lần theo chiều y . Do đó trong quá trình so khớp, theo chiều x có

4 điểm thì sẽ chọn một điểm để so khớp và cứ 2 điểm theo chiều y thì sẽ chọn 1 điểm.

Theo vết đối tượng



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

1. 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)

2. 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.
3. 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.
4. Trường hợp nhiều đối tượng cùng khớp với một vùng chuyển động. Trường hợp này xảy ra do các đối tượng giao nhau.
5. 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 2 frame I_n và $I_{(n+1)}$, ta có

$$d = \frac{p_{n+1} - p_n}{\Delta t}$$

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

Khi đó ta có vận tốc mới của đối tượng được cập nhật như sau:

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

Và độ không chắc chắn của vận tốc được cập nhật như sau:

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

CHƯƠNG IV

CÀI ĐẶT THỬ NGHIỆM

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 sau.

Cấu hình	Chi tiết
Phần cứng	Dell Optiplex 7020, CPU Intel Core i7-4790 @ 8x3.60Ghz, RAM DDR3 16GB
Hệ điều hành	Window 7 SP1 Professional 64-bit
Ngôn ngữ & môi trường	Matlab 2013a 64-bit

Bảng 1. Cấu hình và môi trường thử nghiệm

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 10 phút 07 giây.
- Tốc độ frame là 30fps.

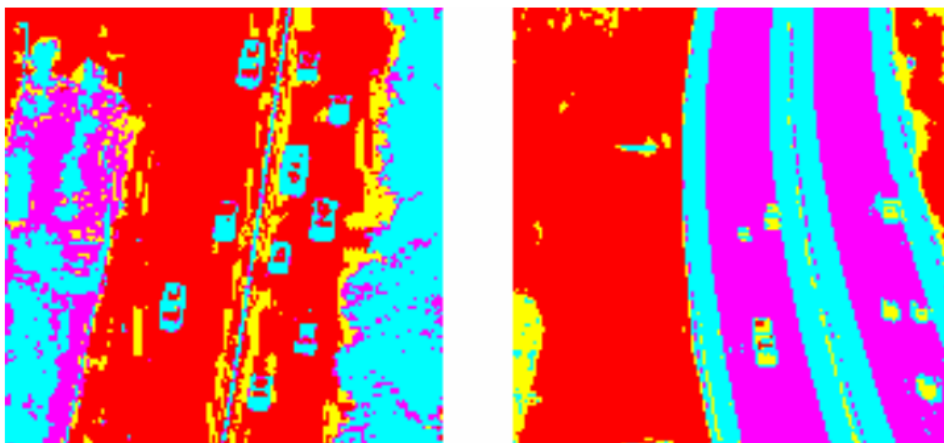
- Kích thước khung hình là 320×320 pixel

4.2. Kết quả thử nghiệm



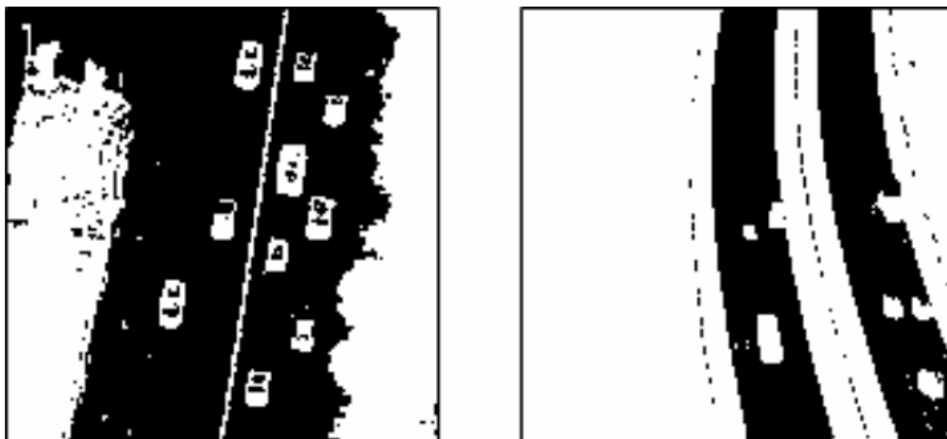
Hình 2. Frame ảnh thử nghiệm

Trong bước phân vùng ảnh, tham số k_1 và k_2 trong công thức ? 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 3 là ảnh sau khi được phân vùng với giải thuật phân vùng cùng các tham số như trên.



Hình 3. 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 là ảnh nhị phân chứa đối tượng đường.



Hình 4. 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 liên mạch (theo kinh nghiệm thành phần cấu trúc được đặt là 5). Kết quả của thao tác hình thái học như hình 5.



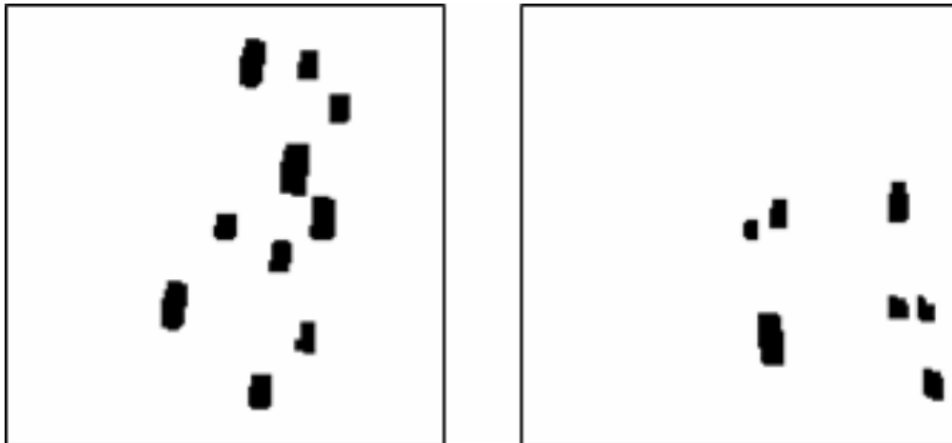
Hình 5. Ảnh nhị phân đường sau khi lọc

Để trích xuất được phương tiện giao thông từ ảnh hình 4, ta tiến hành trừ hình 4 cho hình 5. Hình 6 là kết quả trích xuất phương tiện giao thông bởi phép trừ.



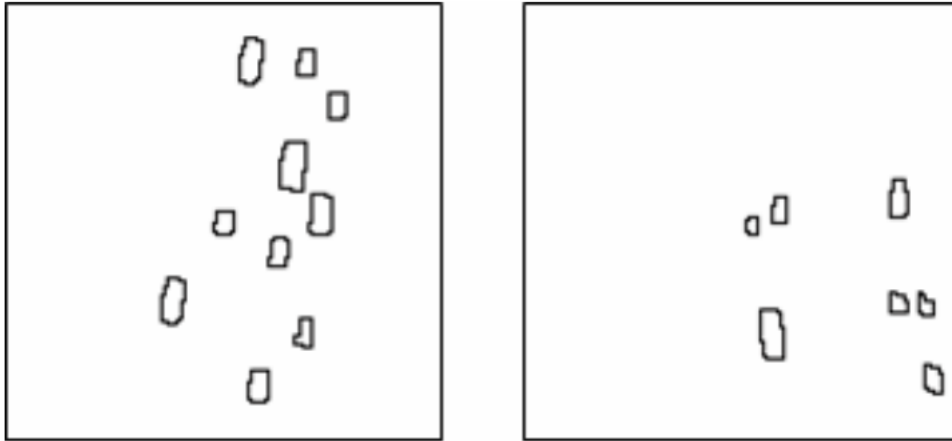
Hình 6. Ảnh nhị phân phương tiện giao thông

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 7.



Hình 7. 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 8.



Hình 8. 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 9. Trong hình được ghép các đường màu đỏ là bao của đối tượng. Có thể quan sát từ Hình 9 rằng phần lớn đường bao khớp chính xác với các phương tiện giao thông.



Hình 9. 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 đo Multiple Object Tracking Accuracy (MOTA)

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

trong đó m_t , fp_t và mme_t lần lượt là số lượt misses, false positives và mismatches tương ứng với frame t .

STT	Đánh giá	Kết quả
1	Miss rate	5.73%
2	False positive rate	7.78%
3	Mismatches	236
4	MOTA	85.96%

Bảng 2. Kết quả theo dõi đối tượng theo độ đo MOTA

Kết quả theo dõi đối tượng tính theo MOTA được ghi nhận ở bảng 2, cho thấy chương trình thử nghiệm cho kết quả khá tốt, nhận dạng và theo dõi đối tượng là phương tiện giao thông với độ chính xác cao.

CHƯƠNG V

KẾT LUẬN

Về mặt lý thuyết, luận văn đã trình bày được các nội dung sau:

- Các khái niệm cơ bản về fuzzy c-partition và giải thuật di truyền ứng dụng cho các bài toán tối ưu.
- Trình bày giải thuật fuzzy segmentation trong bài toán trích xuất đối tượng và áp dụng giải thuật di truyền tối ưu cho fuzzy segmentation.
- Áp dụng bộ lọc Kalman vào bài toán theo dõi đối tượng sau khi đã trích rút được các đối tượng trong khung hình.

Về mặt thực nghiệm, luận văn đã thu được một số kết quả:

- Đã cài đặt thành công mô hình giải thuật như đã trình bày với đối tượng cụ thể là các phương tiện giao thông quay từ thiết bị UAV.
- Kết quả thử nghiệm với độ đo MOTA rất tốt (85.96%)

Như vậy luận văn đã giải quyết được bài toán “*theo dõi đối tượng*” đã đề ra với hiệu quả tốt.

Hạn chế

Như vậy, kết quả thực nghiệm là khá là phù hợp với những gì được trình bày trong lý thuyết. Tuy vậy, do còn nhiều hạn chế về kiến thức, kinh nghiệm của

bản thân, cũng như thời gian thực hiện, luận văn còn các hạn chế như: việc sử dụng giải thuật di truyền còn khá đơn giản, lựa chọn các tham số chủ yếu dựa trên thử nghiệm; đặc trưng sử dụng thì tương đối đơn giản.

TÀI LIỆU THAM KHẢO

- Angel, A., Hickman, M., Mirchandani, P., & Chandnani, D. (2003). Methods of analyzing traffic imagery collected from aerial platforms. *IEEE Transactions on Intelligent Transportation Systems*, 4, 99-107.
- Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International journal of computer vision*, 12, 43-77.
- Collins, R. T., Lipton, A. J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., . . . others. (2000). *A system for video surveillance and monitoring*. Tech. rep., Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University.
- Fleet, D., & Weiss, Y. (2006). Optical flow estimation. Trong *Handbook of mathematical models in computer vision* (trang 237-257). Springer.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press.
- Klir, G. J., & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Lucas, B. D., Kanade, T., & others. (1981). An iterative image registration technique with an application to stereo vision. *IJCAI*, 81, trang 674-679.

- Mittal, A., & Paragios, N. (2004). Motion-based background subtraction using adaptive kernel density estimation. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2, trang II--302.
- Pratt, W. K. (1991). Image segmentation. *Digital Image Processing: PIKS Inside, Third Edition*, 551-587.
- Stauffer, C., & Grimson, W. E. (1999). Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, 2.
- Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *Acm computing surveys (CSUR)*, 38, 13.