

DOI:10.22144/ctu.jvn.2017.140

## NGHIÊN CỨU ỨNG DỤNG KỸ THUẬT THEO DÕI ĐỐI TƯỢNG XÂY DỰNG HỆ THỐNG CAMERA GIÁM SÁT THÔNG MINH

Phạm Nguyên Khang, Đỗ Thanh Nghị và Phạm Thế Phi

Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

### Thông tin chung:

Ngày nhận bài: 31/05/2017

Ngày nhận bài sửa: 07/09/2017

Ngày duyệt đăng: 29/11/2017

### Title:

Using object tracking techniques for intelligent surveillance systems

### Từ khóa:

Camera giám sát, đặc trưng cục bộ, luồng quang học, so khớp đặc trưng, theo dõi đối tượng

### Keywords:

Feature matching, local feature, object tracking, optical flow, surveillance camera

### ABSTRACT

This paper presents some results of building intelligent surveillance camera systems using object tracking. Main steps of the object tracker include (i) keypoint tracking using optical flow, (ii) keypoint matching, and (iii) consensus-base voting. A novel algorithm to accelerate processing using pipeline technique on multicores systems has also been proposed. The algorithm divides the whole processing frame into 4 stages which are executed on 4 different threads. Synchronization of threads is realized producer – consumer model. The proposed method achieved a 3.3 times increased computational time compared to the original one. The surveillance system continuously tracks target object and gives a warning sound if the object disappears in a predefined interval. Experimental results show that the proposed method achieves very promising results.

### TÓM TẮT

Bài báo trình bày các kết quả nghiên cứu về việc xây dựng hệ thống camera giám sát thông minh sử dụng kỹ thuật theo dõi đối tượng. Phần cốt lõi của hệ thống là bộ theo dõi đối tượng, hoạt động dựa trên trên việc kết hợp giữa (i) truy vết đối tượng bằng luồng quang học, (ii) so khớp các đặc trưng cục bộ và (iii) tìm sự đồng thuận lớn nhất của các đặc trưng cục bộ. Cũng trong bài báo này, một thuật giải thuật mới đã được đề xuất nhằm tăng tốc độ xử lý các khung ảnh bằng kỹ thuật ống dẫn (pipeline) trên các hệ thống máy tính đa nhân. Giải thuật này chia quá trình xử lý thành 4 giai đoạn liên tiếp, phụ thuộc nhau và giao cho 4 tiến trình xử lý chúng một cách độc lập. Việc đồng bộ giữa các tiến trình được thực hiện bằng mô hình sản xuất – tiêu thụ (producer – consumer). Điều này giúp tăng tốc độ xử lý lên đến 3,3 lần trên hệ thống máy tính đa nhân. Hệ thống camera giám sát thông minh sẽ theo dõi đối tượng liên tục và phát tín hiệu cảnh báo khi đối tượng cần theo dõi biến mất trong một khoảng thời gian được định trước. Kết quả thực nghiệm cho thấy rằng các giải pháp đề xuất là hoàn toàn phù hợp.

Trích dẫn: Phạm Nguyên Khang, Đỗ Thanh Nghị và Phạm Thế Phi, 2017. Nghiên cứu ứng dụng kỹ thuật theo dõi đối tượng xây dựng hệ thống camera giám sát thông minh. Tạp chí Khoa học Trường Đại học Cần Thơ. 53a: 44-52.

### 1 GIỚI THIỆU

Theo dõi đối tượng (object tracking) là một những lĩnh vực nghiên cứu đầy tiềm năng của thị giác máy tính (computer vision). Việc theo dõi chủ

yếu dựa trên hình ảnh của các đối tượng thu được từ các camera giám sát. Trong các thập niên gần đây, lĩnh vực này đã thu hút đáng kể các nhà nghiên cứu về thị giác máy tính nhờ vào sự đa

dạng của các ứng dụng tiềm năng làm cho lĩnh vực này trở thành mục tiêu nghiên cứu hấp dẫn. Có thể chia bài toán theo dõi đối tượng thành hai nhóm tiêu biểu: biết trước mô hình của đối tượng và không biết trước mô hình. Nhóm bài toán thứ nhất tương đối đơn giản. Dựa trên thông tin đã biết trước về đối tượng như màu sắc hoặc hình dáng của đối tượng, các giải thuật theo dõi sử dụng thông tin này để định vị đối tượng trong các khung hình (frame) của video. Sử dụng màu sắc để mô hình hoá đối tượng là phương pháp đơn giản nhất và phù hợp đối với các đối tượng cần theo dõi có màu (gần) đồng nhất. Phương pháp này hiệu quả khi màu của đối tượng tương đối khác đối với màu nền. Các giải thuật liên quan đến phương pháp này có thể kể đến Mean Shift (Cheng, Y., 1995), CAMShift (Bradski, 1998). Một số phương pháp khác sử dụng thông tin về hình dáng hoặc kết cấu hình học của đối tượng (Lowe, 1992; Jurie and Dhome, 2006). Đối với nhóm bài toán thứ hai, do không cần phải cung cấp trước mô hình của đối tượng cần theo dõi là gì nên ứng dụng của nó cũng rộng rãi hơn, cho phép theo dõi các đối tượng phức tạp. Do phương pháp này không đòi hỏi phải huấn luyện hay bất cứ thông tin đặc biệt nào về đối tượng cần theo dõi nên còn gọi là theo dõi phi mô hình (model-free tracking). Một trong các đặc tính quan trọng của các thuật toán theo dõi là khả năng xử lý hình dáng (bề ngoài) phức tạp của đối tượng trong một thời gian bất kỳ. Mặc dù đã có nhiều tiến triển trong các phương pháp theo dõi đối tượng phi mô hình làm cho nó trở nên mạnh hơn, chịu đựng nhiễu tốt hơn, nhưng bản thân của bài toán cũng phải đối mặt với các khó khăn và trở nên khó giải hơn vì các lý do: hình dáng của đối tượng phức tạp, bị che khuất một phần, ảnh hưởng của ánh sáng, màu sắc, góc chụp/quay của camera (Maggio and Cavallaro, 2011). Ngoài ra, do đối tượng không được biết trước khi theo dõi nên không thể huấn luyện máy học để nhận dạng đối tượng. Một số tác giả sử dụng chiến lược học trực tuyến (online) trong quá trình theo dõi (Safari *et al.*, 2009). Tuy nhiên, việc cập nhật mô hình trong khi học thường cũng gây ra lỗi vì thiếu dữ liệu để huấn luyện.

Hướng tiếp cận được xem là nhiều hứa hẹn nhất hiện nay đối với bài toán theo dõi đối tượng phi mô hình là dựa trên các đặc trưng cục bộ bất biến về hình dáng của đối tượng. Các phương pháp có thể kể đến là chia đối tượng thành các vùng nhỏ (patches) và biểu diễn đối tượng như một tập hợp các vùng này. Việc chia đối tượng thành các vùng nhỏ có thể dựa trên lưới hoặc tìm các vùng bất biến (Adam *et al.*, 2006; Hua and Wu, 2006; Nejjum *et al.*, 2008). Một phương pháp khác của tiếp cận này là sử dụng các điểm đặc biệt (interest points, key points) trên đối tượng và ước lượng vị trí của nó

trong các khung ảnh kế tiếp dựa trên luồng quang học (optical flow). Việc xác định vị trí đối tượng trong khung ảnh sau đó dựa trên việc so khớp các điểm đặc biệt ở khung ảnh trước và khung ảnh sau. Việc truy vết đối tượng dựa trên sự đồng thuận của các đặc trưng truy vết được (Nebehay, 2015). Phương pháp này có ưu điểm là có thể theo dõi được ngay cả khi đối tượng bị quay hay camera bị thay đổi vị trí.

Phần tiếp theo của bài báo được tổ chức như sau: kỹ thuật theo dõi đối tượng dựa trên sự đồng thuận được mô tả trong phần 2; kết quả thực nghiệm được trình bày trong phần 3 và sau cùng là kết luận và hướng phát triển.

## 2 THEO DÕI ĐỐI TƯỢNG DỰA TRÊN SỰ ĐỒNG THUẬN

### 2.1 Phương pháp CMT

So khớp và theo dõi dựa trên sự đồng thuận (Consensus-based matching and tracking hay CMT) được (Nebehay *et al.*, 2014) đề xuất, là một phương pháp dựa trên keypoint để theo dõi đối tượng theo kỹ thuật phi mô hình kết hợp giữa so khớp (matching) và theo dõi (tracking). Để định vị đối tượng trong mỗi khung hình (frame), mỗi keypoint sẽ bình chọn tâm của đối tượng. Trong quá trình so khớp các keypoint giữa hai khung liên tiếp, so khớp sai (false matching: hai keypoint được giải thuật cho khớp với nhau nhưng thực chất chúng không khớp nhau) là điều khó tránh khỏi. Nebehay *et al.* đã đề xuất một mô hình dựa trên sự đồng thuận (consensus-based) để phát hiện các keypoint bị so khớp sai. Tiếp cận mới này phân cụm các bình chọn (votes) trực tiếp trong không gian ảnh.

Bằng cách áp dụng các biến đổi hình học (phép quay, thay đổi tỉ lệ) lên bình chọn, phương pháp này cho phép theo dõi được đối tượng bị biến đổi hình dạng trong quá trình theo dõi. Ngoài ra, phương pháp này không sử dụng các thông tin hình ảnh của các keypoint và chỉ đơn thuần dựa vào vị trí của chúng, và do đó tránh được các sai sót nghiêm trọng trong trường hợp hai vùng keypoint có bề ngoài giống nhau nhưng không có liên quan nhau về mặt hình học. Để tăng tốc độ tính toán, các keypoint được mô tả bằng một vector nhị phân. Phương pháp này đã được thực nghiệm trong môi trường thời gian thực (real-time) trên một tập lớn dữ liệu lớn với kết quả cao.

Ý tưởng chính của phương pháp CMT được mô tả như sau: cho một chuỗi  $n$  khung ảnh liên tục  $I_1, \dots, I_n$ , và một vùng theo dõi  $b_1$  (chứa đối tượng cần được theo dõi) trong khung ảnh  $I_1$ , với mỗi khung ảnh  $I_t$ , cần xác định tư thế (pose) của đối tượng

đang theo dõi hoặc nói rằng đối tượng không có trong khung ảnh.

Việc ước lượng tư thế của đối tượng dựa vào tâm  $\mu$ , tỉ lệ  $s$  và góc quay  $\alpha$  của nó, trong đó  $s$  và  $\alpha$  được ước lượng dựa trên hình dạng ban đầu của đối tượng trong vùng  $b_1$ . Để đơn giản, giả sử rằng vùng theo dõi  $b_1$  có dạng hình chữ nhật có các cạnh song song với các trục tọa độ. Giải thuật CMT được cho trong Bảng 1.

a. So khớp và truy vết các điểm đặc trưng

Mô hình đối tượng được mô tả dựa trên tập các keypoint:

$$O = \{(r_i, f_i)\}_{i=1}^{N^O} \quad (1)$$

Trong đó, mỗi keypoint bao gồm 2 phần tử là vị trí (tọa độ)  $r \in \mathbb{R}^2$  và bộ mô tả  $f$ . Để đơn giản trong tính toán,  $f$  được mô tả theo kiểu nhị phân  $f \in \{0, 1\}^d$ .

Tập các keypoint  $O$  được khởi tạo bằng kỹ thuật trích đặc trưng BRISK hoặc SIFT từ khung hình đầu tiên  $I_1$  bên trong vùng khởi tạo  $b_1$ , sau đó quy tâm (mean-normalisation) vị trí của các keypoint. Để bảo tồn hình dáng của đối tượng, mỗi khung hình  $I_t$  với  $t \geq 2$ , cần tìm ra tập các keypoint tương ứng:

$$K_t = \{(a_i, m_i)\}_{i=1}^{N^{K_t}} \quad (2)$$

Trong đó,  $a$  là vị trí các keypoint trong ảnh,  $m$  là chỉ số (index) của keypoint tương ứng trong  $O$ .

**Bảng 1: Giải thuật CMT**

**Giải thuật CMT**

**Input:**  $I_1, \dots, I_n, b_1$

**Output:**  $b_2, \dots, b_n$

- 1:  $O \leftarrow \text{detect}(I_1, b_1)$
- 2:  $K_1 \leftarrow O$
- 3: for  $t \leftarrow 2, \dots, n$  do
- 4:      $P \leftarrow \text{detect}(I_t)$
- 5:      $M \leftarrow \text{match}(P, O)$
- 6:      $T \leftarrow \text{track}(K_{t-1}, I_{t-1}, I_t)$
- 7:      $K' \leftarrow T \cup M$
- 8:      $s \leftarrow \text{estimate\_scale}(K', O)$
- 9:      $\alpha \leftarrow \text{estimate\_rotation}(K', O)$
- 10:      $V \leftarrow \text{vote}(K', O, s, \alpha)$
- 11:      $V^C \leftarrow \text{consensus}(V)$
- 12:      $K_t \leftarrow \text{vote}^{-1}(V^C)$
- 13:     if  $|V^C| \geq \vartheta \cdot N^O$  then
- 14:          $\mu \leftarrow \frac{1}{n} \sum_{i=1}^n V_i^C$
- 15:          $b_t \leftarrow \text{bounding\_box}(b_1, \mu, s, \alpha)$
- 16:     else
- 17:          $b_t \leftarrow \emptyset$
- 18:     end if
- 19: end for

Để xác định  $K_t$ , ta thực hiện các bước tính toán như sau:

Xác định các keypoint trong khung ảnh thứ  $I_t$ , mỗi keypoint cũng có vị trí  $a$  và bộ mô tả  $f$ . Lưu trữ trong tập  $P$ .

$$P = \{(a_i, f_i)\}_{i=1}^{N^P} \quad (3)$$

Mỗi keypoint trong tập P, cần phải tính khoảng cách Hamming theo công thức (4) giữa bộ mô tả của nó với bộ mô tả của từng keypoint được tìm thấy trong I<sub>1</sub>, bao gồm cả các keypoint nền.

$$d(f^1, f^2) = \sum_{i=1}^d XOR(f_i^1, f_i^2) \quad (4)$$

Để so khớp các keypoint trong P với các keypoint trong I<sub>1</sub>, khoảng cách đến láng giềng gần nhất phải gần hơn các láng giềng khác theo một tỷ lệ cố định ρ nào đó (Lowe, 2014). Tập các keypoint M là một tập con chứa các keypoint đã so khớp giữa keypoint trong P và các keypoint trong O, sau đó bổ sung thêm các chỉ số (index) của keypoint tương ứng (là các định danh m trong công thức (2))

Để theo vết, chúng ta tính toán sự dịch chuyển của K<sub>t-1</sub> từ I<sub>t-1</sub> đến I<sub>t</sub> bằng cách dùng phương pháp ước lượng luồng quang học (Lucas & Kanate, 1981). Tại thời điểm t=2, K<sub>1</sub> bằng với tập các keypoint của O. Tập các keypoint T được tìm ra bằng cách cập nhật vị trí của keypoint trong K<sub>t-1</sub> nhưng vẫn giữ lại các định danh của keypoint.

Sau đó thực hiện kết hợp hai tập T và M và đưa vào tập K' với kích thước NK', khi kết hợp hai tập T, M thì các keypoint giống nhau ở hai tập này sẽ được loại bỏ. Đến bước này trong tập K' vẫn còn chứa các keypoint ngoại lai hoặc các điểm mơ hồ không chính xác.

*b. Bình chọn*

Để xác định vị trí của đối tượng, mỗi keypoint (a, m) trong K' sẽ bình chọn một giá trị h(a,m) → ℝ<sup>2</sup> vào tâm của đối tượng, tập các bình chọn đó được thể hiện như sau:

$$V = \{h(a_i, m_i)\}_{i=1}^{N^{K'}} \quad (5)$$

Ta có thể xem mỗi bình chọn (vote) là một phép biến đổi hình học affine gồm: tịnh tiến, tỉ lệ và quay lên các keypoint trong không gian 2 chiều.

Trước hết, với phép tịnh tiến, h<sup>T</sup>(a, m) được tính như sau:

$$h^T(a, m) = a - r_m \quad (6)$$

Trong đó, r<sub>m</sub> là vị trí tương đối của keypoint tương ứng với chỉ số (index) của m trong O. Khi đối tượng bị thay đổi tỉ lệ (ví dụ thu nhỏ), các vector bình chọn này tiến gần đến trung tâm đối tượng (các vector bị thu ngắn lại xem Hình 5). Khi đó tỷ lệ thu nhỏ của các vector bình chọn được tính lại như sau:

$$h^S(a, m) = a - s \cdot r_m \quad (7)$$

Để tính s, chúng ta tính khoảng các Euclide giữa các cặp a<sub>i</sub> và a<sub>j</sub> trong K', và so sánh khoảng cách vừa tính được với khoảng cách của các keypoint tương ứng r<sub>m<sub>i</sub></sub> và r<sub>m<sub>j</sub></sub> trong O.

Ta có: a<sub>i,j</sub> = a<sub>i</sub> - a<sub>j</sub> và r<sub>i,j</sub> = r<sub>m<sub>i</sub></sub> - r<sub>m<sub>j</sub></sub>

$$D_s = \left\{ \frac{\|a^{i,j}\|}{\|r^{i,j}\|}, i \neq j \right\} \quad (8)$$

Từ đó rút ra được s = median (D<sub>s</sub>)

Khi đối tượng bị quay, các vector bình chọn bị quay theo, và tâm quay là tâm của đối tượng:

$$h^R(a, m) = a - s \cdot R r_m \quad (9)$$

Với R là ma trận quay trong không gian 2D

$$R = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad (10)$$

Để tính được góc quay α, cần phải tính được các góc α<sub>i,j</sub> (theo công thức 11) là góc tạo bởi sự chênh lệch góc quay của hai đường thẳng so với trục ngang:

- Đường thẳng thứ nhất nối hai điểm thứ i và thứ j trong khung hình đầu tiên

- Đường thẳng thứ hai nối hai điểm thứ i và thứ j trong khung hình thứ t.

$$\alpha_{i,j} = \text{atan2}(a_y^{i,j}, a_x^{i,j}) - \text{atan2}(r_y^{i,j}, r_x^{i,j}) \quad (11)$$

Duyệt qua tất cả các cặp keypoint ta được tập D<sub>α</sub> như sau:

$$D_\alpha = \{\alpha_{i,j}, i \neq j\} \quad (12)$$

Từ đó rút ra được α = median (D<sub>α</sub>)

*c. Đồng thuận (Consensus)*

Khi có một keypoint trong K' không thuộc về đối tượng (key point ngoại lai) thì vector bình chọn của nó sẽ không hướng đến tâm của đối tượng mà nó chỉ đến một vùng hình ảnh bất kỳ nào đó. Trước khi tìm tâm của đối tượng μ, chúng ta cần loại bỏ các keypoint ngoại lai này và các vector bình chọn của nó bằng cách tìm kiếm sự đồng thuận của các vector bình chọn.

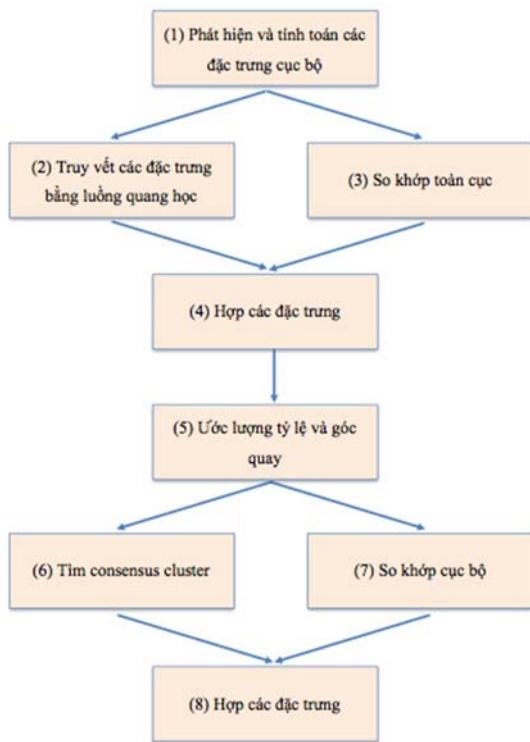
Để thực hiện việc này, chúng ta phân V thành các cụm nhỏ hơn V<sub>1</sub>, V<sub>2</sub>, ..., V<sub>m</sub> bằng cách áp dụng giải thuật phân cụm phân cấp theo hướng từ dưới lên (hierarchical agglomerative clustering) trên các tập điểm trong V sử dụng khoảng cách Euclide. Trong cách phân cụm này, dữ liệu được tổ chức thành các cấu trúc phân cấp theo ma trận khoảng cách, kết quả là một cây phân cấp (dendrogram), sau đó cây phân cấp này sẽ bị cắt theo một ngưỡng δ xác định nào đó (dùng phân cấp khi khoảng cách Euclide giữa hai cụm lớn hơn hoặc bằng ngưỡng δ).

Gọi tập con  $V_c$  là tập có số thành phần lớn nhất trong số các cụm thu được ở bước phân cụm, hay có thể gọi  $V_c$  bằng một cái tên khác là cụm đồng thuận (consensus cluster), và gọi  $K_t$  là tập các keypoint nằm trong  $K'$  sao cho các keypoint này có vector bình chọn nằm trong tập  $V_c$ .

Nếu số lượng thành phần của  $V_c$  nhỏ hơn  $\theta \cdot |O|$  thì xem như đối tượng không xuất hiện. Ngược lại chúng ta sẽ tính tâm của đối tượng dựa trên các thành phần của cụm đồng nhất này.

$$\mu = \frac{1}{n} \sum_{i=1}^n V_i^c \text{ với } n = |V^c|.$$

Với tâm của đối tượng  $\mu$ , tỷ lệ  $s$ , và góc quay  $\alpha$  chúng ta xác định được tư thế của đối tượng đang quan tâm.



**Hình 1: Các bước trong quá trình theo dõi đối tượng bằng giải thuật CMT**

Kết quả thu được của bước này là một cluster lớn nhất tập trung các đặc trưng truy vết được. Mặc dù phương pháp này có khả năng loại bỏ nhiều

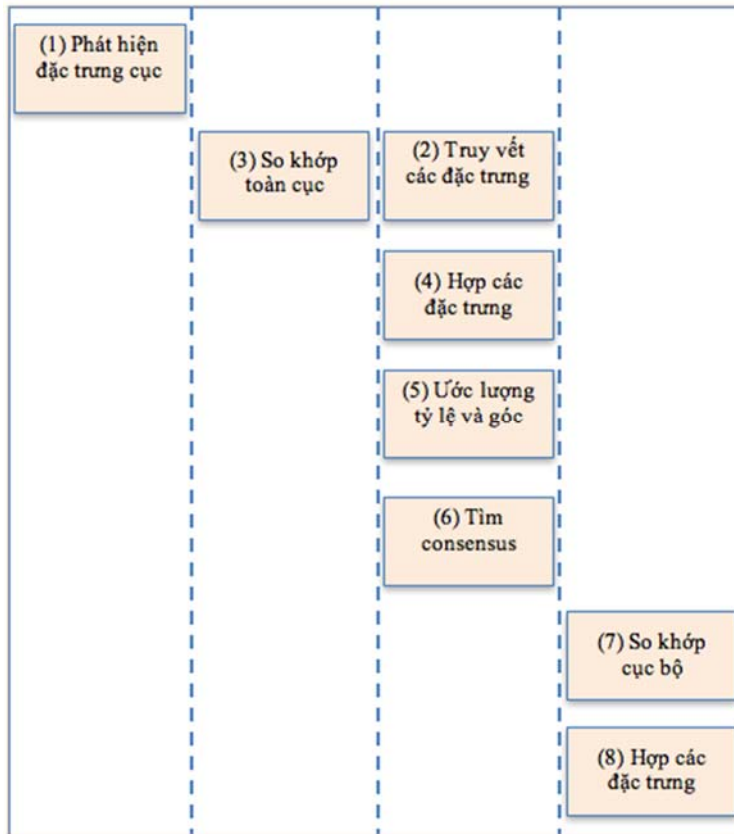
*false match* tuy nhiên nó lại làm giảm số lượng các đặc trưng được truy vết. (Nebhey *et al.*, 2015) đã cải tiến điều này bằng cách bổ sung thêm một bước so khớp cục bộ. Các bước chính của toàn bộ quá trình xử lý được trình bày trong Hình 2.

**2.2 Tăng tốc độ xử lý bằng phương pháp ống dẫn**

Giải thuật CMT gốc được trình bày trong phần trên xử lý tuần tự các khung ảnh vì để xử lý khung ảnh hiện tại ta cần có kết quả của khung ảnh trước đó. Hơn nữa, trong bản thân quá trình xử lý, như mô tả trong Hình 1, các bước này đều tuần tự và phụ thuộc nhau (ngoại trừ bước 2, 3 và 6, 7 có thể thực hiện song song). Vì thế, một giải thuật mới đã được đề xuất để cải tiến tốc độ xử lý việc theo dõi bằng cách sử dụng kỹ thuật ống dẫn (pipeline) trên các hệ thống đa nhân (multicores).

Các bước từ 1 đến 8 trong Hình 1 được chia thành 4 giai đoạn chính (stages): (1, 2), (3), (4, 5, 6) và (7, 8). Mỗi giai đoạn được thực thi trong một luồng (thread) khác nhau. Việc đồng bộ hoá quy trình xử lý được thực hiện dựa trên các hàng đợi kết quả theo mô hình sản xuất – tiêu thụ (producer – consumer problem). Giải thuật hoạt động như sau: đầu tiên, thread 1 tính toán các đặc trưng cục bộ và lưu kết quả vào hàng đợi. Thread 2 chờ cho đến khi thread 1 đưa các đặc trưng cục bộ vào hàng đợi là nó có thể bắt đầu thực hiện việc so khớp toàn cục. Kết quả của việc so khớp toàn cục sẽ được đặt vào hàng đợi tương ứng. Trong khi thread 2 làm công việc này, thread 1 lại có thể xử lý tiếp frame ảnh thứ 2. Thread 3 thực hiện truy vết các đặc trưng của bước trước đó bằng phương pháp luồng quang học và lưu kết quả vào trong hàng đợi. Sau đó, nó chờ cho đến khi thread 2 so khớp toàn cục xong nó sẽ hợp các điểm đặc trưng lại và tính toán ước lượng tỷ lệ, góc quay và tìm cluster lớn nhất theo phương pháp consensus. Thread 4 chờ cho thread 3 thực hiện xong nó sẽ thực hiện so khớp cục bộ hợp các đặc trưng và hợp nhất các đặc trưng để cho ra kết quả sau cùng. Trong 8 bước của giải thuật, các bước (1), (3) và (7) là các bước chiếm nhiều thời gian nhất nên nó đã được chia ra thực hiện trong các thread khác nhau nhằm tăng tốc độ xử lý (Hình 2).



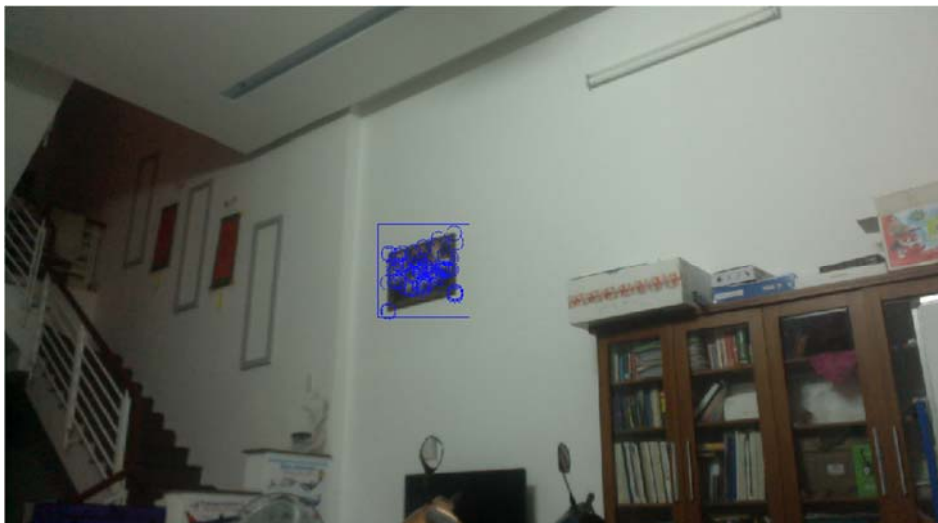


**Hình 2: Giải thuật CMT cải tiến sử dụng kỹ thuật ống dẫn**

### 3 KẾT QUẢ THỰC NGHIỆM

Giải thuật đề xuất được cài đặt bằng ngôn ngữ C++ trên sử dụng thư viện OpenCV. Phiên bản song song sử dụng kỹ thuật ống dẫn được cải tiến từ mã nguồn của giải thuật CMT do tác giả cung cấp. Dữ liệu thực nghiệm thực nghiệm được thu thập từ các camera quan sát ở các siêu thị ở tỉnh

Sóc Trăng với tổng cộng hơn 5 giờ video. Các video được lưu lại với định dạng avi. So sánh thời gian thực thi giữa phiên bản CMT gốc và phiên bản CMT dùng kỹ thuật ống dẫn, kết quả cho thấy rằng giải thuật CMT cải tiến xử lý nhanh hơn 1,7 lần trên máy tính 2 cores và 3,3 lần trên máy tính 4 cores. Kết quả này phù hợp với giả thiết ban đầu.



**Hình 3: Khởi tạo đối tượng cần theo dõi**

Giải thuật mới đã được sử dụng để xây dựng một hệ thống camera giám sát thông minh. Hệ thống này có khả năng theo dõi đối tượng do người dùng chỉ định và phát tín hiệu cảnh báo khi đối tượng biến mất khỏi khung quan sát trong một khoảng thời gian định trước.

Để minh họa hiệu quả của hệ thống camera theo dõi an ninh, các thí nghiệm đã được thiết lập để theo dõi trong ngôi nhà với camera theo dõi IP hiệu SmartZ. Camera được điều chỉnh hướng vào đối tượng là một bức tranh trên tường (Hình 3).

Người dùng sử dụng chuột để xác định đối tượng cần theo dõi (bức tranh). Hệ thống theo dõi khởi động và tính toán các điểm đặc trưng của bức tranh. Nếu số lượng đặc trưng quá ít (nhỏ hơn), hệ thống sẽ thông báo rằng không thể theo dõi đối tượng.

Cho camera quay đi một góc 30° theo phương ngang. Đối tượng lúc này bị lệch đi so với vị trí ban đầu nhưng vẫn còn nằm trong tầm nhìn của hệ thống. Hệ thống vẫn phát hiện được (Hình 4).



**Hình 4: Đối tượng vẫn tiếp tục được theo dõi mặc dù camera bị quay một góc**



**Hình 5: Đối tượng bị lấy ra khỏi vùng theo dõi**



**Hình 6: Đối tượng được trả về chỗ cũ**



**Hình 7: Đối tượng vẫn còn được theo dõi nếu phần nhìn thấy còn > 30%**

Bức tranh được lấy ra khỏi vị trí (Hình 5), lúc này hệ thống còn thấy được đối tượng nên khởi động bộ đếm. Nếu thời gian (số khung ảnh) không nhìn thấy đối tượng vượt ngưỡng, hệ thống sẽ báo động liên tục cho đến khi ta trả bức tranh về chỗ cũ (Hình 6) hoặc hủy bỏ việc theo dõi.

Ngoài ra, do sử dụng các đặc trưng cục bộ, hệ thống có thể theo dõi được đối tượng ngay cả khi đối tượng bị che khuất hoặc ngoài tầm nhìn một phần (phần nhìn thấy được khoảng 30%). Khi phần nhìn thấy được còn ít hơn giá trị ngưỡng, hệ thống

sẽ xem như không còn nhìn thấy được đối tượng nữa (Hình 7).

Để minh họa cho khả năng theo dõi đối tượng động của hệ thống, một thí nghiệm khác đã thiết lập dựa trên video thu được bằng camera giám sát ở một siêu thị. Đối tượng được theo dõi là một nhân viên phụ trách quầy bán sữa. Hình 8 mô tả tình hình siêu thị trước khi theo dõi. Cô nhân viên ở bên trái (ảnh trái). Sau khi chọn đối tượng cần theo dõi, các đặc trưng của đối tượng được hiển thị (ảnh phải).



**Hình 8: Chọn đối tượng cần theo dõi**



**Hình 9: Hệ thống tự động bám theo đối tượng**





**Hình 10: Đối tượng được theo dõi cho đến khi rời khỏi vùng quan sát**

Khi đối tượng di chuyển, hệ thống sẽ tự động theo dõi và ghi nhận lại vị trí của đối tượng (Hình 9) cho đến khi đối tượng ra khỏi vùng quan sát (Hình 10). Thí nghiệm này cho thấy rằng hệ thống hoạt động tốt trong điều kiện thực.

#### 4 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã trình bày một giải pháp về việc xây dựng một hệ thống camera giám sát thông minh sử dụng kỹ thuật theo dõi đối tượng với phương pháp CMT và đề xuất cải tiến tốc độ xử lý của giải thuật CMT bằng kỹ thuật ống dẫn trên các hệ thống máy tính đa nhân. Hệ thống camera giám sát thông minh đã được cài đặt và kiểm thử. Với kết quả rất khả quan, việc nghiên cứu, thực nghiệm sẽ tiếp tục với nhiều tình huống hơn và điều kiện ánh sáng phức tạp hơn. Mặc dù kỹ thuật ống dẫn đã có thể góp phần tăng tốc độ xử lý các khung cảnh nhưng vẫn còn có thể tiếp tục tăng tốc xử lý bằng cách song song hoá hai bước chiếm nhiều thời gian nhất là bước (3) so khớp toàn cục và (7) so khớp cục bộ. Điều này sẽ được thực hiện trong các nghiên cứu tiếp theo.

#### TÀI LIỆU THAM KHẢO

Adam, A., Rivlin, E., and Shimshoni, I., 2006. Robust fragments-based tracking using the integral histogram. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 7-22 June 2006. New York, USA, 1: 798-805.

Bradski, G. R., 1998. Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, 2nd Quarter.

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence. 17:790-799.

Jurie, F., and Dhome, M., 2002. Hyperplane approximation for template matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24: 996-100.

Hua, G., and Wu, Y., 2006. Measurement integration under in consistency for robust tracking. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1: 650-657.

Lowe, D.G., 1992. Robust model-based motion tracking through the integration of search and estimation. International Journal of Computer Vision, 8(2): 113-122.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In the proceedings of the seventh IEEE international conference on Computer vision. 2:1150-1157.

Lucas, B.D., Kanade, B. D., 1981. An iterative image registration technique with an application to stereo vision. In IJCAI, 81:674-679.

Maggio, E. and Cavallaro, A., 2011. Video Tracking: Theory and Practice. John Wiley & Sons, 292 pages.

Nebehay, G. and Pflugfelder, R., 2014. Consensus-based matching and tracking of keypoints for object tracking. In IEEE Winter Conference on Applications of Computer Vision, 862-869.

Nejhum, S. M. S., Ho, J., and Yang, M.-H., 2008. Visual tracking with histograms and articulating blocks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008, 1-8.

Saffari, A., Leistner, C., Santner, J., Godec, M., and Bischof, H., 2009. On-line random forests. In Proceedings of the IEEE 12th International Conference on Computer Vision Workshops, 1- 9.