MỘT KỸ THUẬT PHÁT HIỆN, BÁM SÁT ĐỐI TƯỢNG VÀ ỨNG DỤNG

Trần Thanh Việt[1], Trần Công Chiến[1], Huỳnh Cao Tuấn[1], Nguyễn Hữu Nam[1], Đỗ Năng Toàn[2], Trần Hành[1]

(1)Information Resource Center, Lac Hong University

Email: {thanhviet, chientran, caotuan, huunam}@lhu.edu.vn

(2)Viện Công nghệ Thông tin, Viện Khoa học và Công nghệ Việt Nam

Email: dntoan@ioit.ac.vn

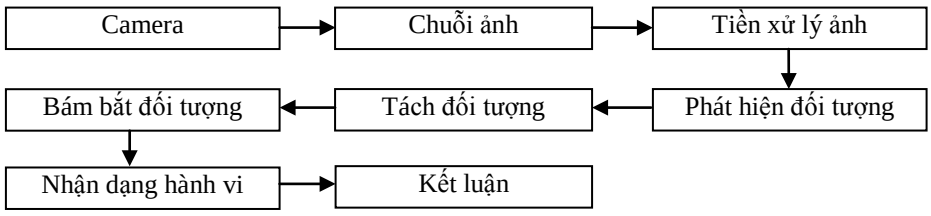
Tóm tắt: Việc phát hiện các đối tượng chuyển động trong camera nhờ các kỹ thuật xử lý ảnh, để khoanh vùng và đoán nhận một số hành vi của đối tượng là một việc làm có ý nghĩa khoa học và thực tiễn. Ở Việt Nam chưa có nhiều nghiên cứu và ứng dụng theo hướng này. Trong bài báo đã trình bày nghiên cứu kỹ thuật ứng dụng thử nghiệm theo vết đối tượng trong camera và dựa trên các hành vi của đối tượng để điều khiển thiết bị máy tính như chuột, lướt web, ra các sự kiện click, double click, right click, zoom out, zoom in. Các kết quả nghiên cứu bằng mô hình thực tế và kết quả đạt được như mong muốn.

Keywords: object tracking, optical flow, meanshift, camshift,computer vision.

1. Đặt vấn đề

Giám sát tự động là một hướng mới được nghiên cứu và phát triển trong lĩnh vực nhận dạng và xử lý ảnh và tạo cách tiếp cận cho phần mềm thiết kế chuyên dụng cho các thiết bị giám sát tự động. Việc phát hiện ra các đối tượng chuyển động trong camera nhờ các kỹ thuật xử lý ảnh đã đoán nhận một số hành vi của đối tượng là một việc làm có ý nghĩa khoa học và thực tiễn.

Chúng ta biết kết quả thu nhận từ các camera giám sát hoặc webcam là các frame ảnh, kết quả nghiên cứu chính của bài báo ở đây là việc phát hiện đối tượng chuyển động trong các frame ảnh đó. Frame ảnh thu nhận được từ các camera hoặc webcam sẽ được xử lý qua các công đoạn sau: Phát hiện đối tượng chuyển động, đánh dấu các đối tượng vừa phát hiện, phân loại chúng được tiến hành xử lý và được kết quả là đối tượng đang cần theo vết ở vị trí nào, để tiến hành đánh dấu (tô màu, kẻ khung) và từ đó liên tục bám sát đối tượng theo một ngưỡng nhất định.



Hình 1 - Sơ đồ mô tả các tiến trình xử lý của hệ thống

Bài báo đã giải quyết bài toán chọn đối tượng muốn theo vết, xác định vị trí đối tượng và điều khiển thiết bị chuột tới vị trí mong muốn (vị trí của đối tượng đang theo vết), đồng thời quyết định ra sự kiện gì (Click, Double click, Drap & Drop, Zoom out, Zoom in…), để đạt được mục đích cuối cùng là có được một ứng dụng mà người sử dụng có thể duyệt web, click chọn liên kết, phóng to, thu nhỏ ảnh … mà không cần sử dụng chuột.

2. Hiện trạng

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 frame 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 [7]:

Loại	Những nghiên cứu liên quan
Point detectors	1. Moravec's detector 2. Harris detector 3. Scale Invariant Feature Transform Affine 4. Invariant Point Detector
Segmentation	1. Mean-shift 2. Graph-cut 3. Active contours
Background Modeling	1. Mixture of Gaussians 2. Eigenbackground 3. Wall flower 4. Dynamic texture background
Supervised Classifier	1. Support Vector Machines 2. Neural Networks 3. Adaptive Boosting

Bảng 1 - Bảng phân loại các thuật toán phát hiện đối tượng

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 frame hiện tại để rút ra được các foreground chuyển động. Để có thể tiếp cận cần phải xây dựng được mô hình background. Có nhiều phương pháp xây dựng mô hình background bởi các tác giả: Anurag Mittal dùng adaptive kernel density estimation được tính bằng[5]. Kết quả tốt tuy nhiên khó khăn về không gian lưu trữ, tính toán phức tạp, tốc độ không đáp ứng thời gian thực. Haritaoglu dùng giải thuật W4, Stauffer sử dụng Mixture of Gaussian [6] 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 and Dan Schonfeld 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 máy học. Các phương pháp này có thể kể đến như: mạng neural, adaptive boosting, cây quyết định, support vector machines. 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 máy học. Một số đặc trưng thường được sử dụng như: đặc trưng về màu sắc, đặc trưng về góc cạnh, đặc trưng histogram… 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 máy học 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 máy học 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.

3. Phương pháp nghiên cứu

3.1 Phương pháp trừ nền

Thuật toán trừ nền xác định mức xám của ảnh Video từ một camera tĩnh [2]. Phương pháp trừ nền này khởi tạo một nền tham khảo với một số frame đầu tiên của Video đầu vào. Sau đó, nó trừ giá trị cường độ của mỗi điểm ảnh trong ảnh hiện thời cho giá trị tương ứng trong ảnh nền tham khảo.

Gọi $I_n(x)$ là biểu diễn của giá trị cường độ mức xám ở điểm ảnh có vị trí (x) và ở trường hợp thứ n của dãy Video I thuộc trong đoạn [0, 255]. Gọi $B_n(x)$ là giá trị cường độ nền tương ứng cho điểm ảnh ở vị trí (x) ước lượng theo thời gian từ ảnh Video I_0 đến I_{n-1}. Một điểm ảnh ở vị trí (x) trong ảnh hiện thời thuộc thành phần nổi trội nếu nó thỏa mãn:

$$|I_n(x) - B_n(x)| > T_n(x) \qquad (1)$$

Trong đó $T_n(x)$ là giá trị ngưỡng có khả năng thích hợp được khởi tạo cùng với ảnh Video đầu tiên I_0, $B_0 = I_0$, và ngưỡng được khởi tạo bởi giá trị đã được xác định trước.

Nền cơ sở và các ảnh ngưỡng phải được cập nhật liên tục từ các ảnh đầu vào. Sự phối hợp cập nhật này là khác nhau đối với các vị trí điểm, chẳng hạn như một điểm $x \in FG$ thì sẽ khác với $x \in BG$:

$$B_{n+1}(x) = \begin{cases} \alpha B_n(x) + (1-\alpha) I_n(x), x \in BG \\ \beta B_n(x) + (1-\beta) I_n(x), x \in FG \end{cases} \qquad (2)$$

$$T_{n+1}(x) = \begin{cases} \alpha T_n(x) + (1-\alpha)(\gamma \times |I_n(x) - B_n(x)|), & x \in BG \\ T_n(x), & x \in FG \end{cases} \qquad (3)$$

Trong đó $\alpha, \beta (\in [0.0,1.0])$ là các hằng số chỉ ra rằng có bao nhiêu thông tin từ các ảnh vào được đẩy vào nền và các ảnh ngưỡng. Nói cách khác, nếu mỗi điểm ảnh nền được coi như là chuỗi các lần, các ảnh nền là một giá trị trung bình của trọng số vùng theo thời gian của chuỗi các ảnh đầu vào và ảnh ngưỡng là giá trị trung bình của trọng số vùng của γ lần khác nhau của các ảnh đầu vào và nền đó, ví dụ :

(a) (b) (c)

Hình 2 - Ảnh (a) là ước lượng nền cơ sở, ảnh (b) thu được ở bước tiếp theo. ảnh (c) thể hiện bản đồ điểm ảnh nổi trội phát hiện được bằng cách sử dụng phép trừ nền.

Thuật toán trừ nền:

Input: ảnh nền B, ảnh hiện tại I và ma trận ngưỡng T

Output: ảnh M là mặt nạ chuyển động

```
m:=getHeight(M);
n:=getWidth(M);
for x:=1 to m do
        for y:=1 to n do
                if  | B[x,y]-I[x,y] | > T[x,y] then
                        M[x,y]:=255;
                else
                        M[x,y]:=0;
```

Thuật toán cập nhật nền:

Input: nền B, ảnh hiện tại I và mặt nạ chuyển động M

Output: nền B được cập nhật lại

```
m:=getHeight(B);
n:=getWidth(B);
for x:=1 to m do
        for y:=1 to n do
                if M[x,y]=0 then
                        B[x,y]:=α*B[x,y]+(1-α)*B[x,y];
```

3.2 Phương pháp Optical flow

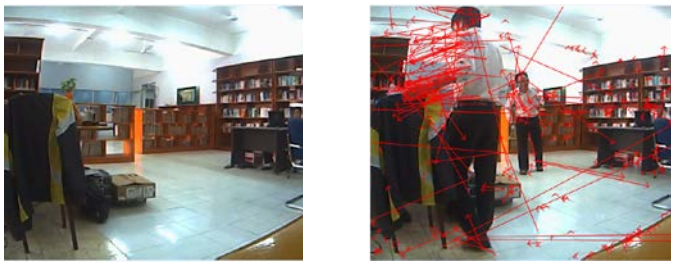
Phương pháp Optical flow[3] thực hiện bằng cách sử dụng các vector có hướng của các đối tượng chuyển động theo thời gian để phát hiện các vùng chuyển động trong một ảnh[1].

Ý tưởng quan trọng của phương pháp tính optical flow dựa trên giả định sau:

Bề ngoài của đối tượng không có nhiều thay đổi (về cường độ sáng) khi xét từ frame thứ n sang frame n+1.

Nghĩa là: $I(\vec{x},t) = I(\vec{x}+\vec{u},t+1)$ (4)

Trong đó $I(\vec{x},t)$ là hàm trả về cường độ sáng[4] của điểm ảnh $\vec{x}$ tại thời điểm t (frame thứ t). $\vec{x} = (x, y)^T$ là tọa độ của điểm ảnh trên bề mặt (2D), $\vec{u} = (u_1, u_2)$ là vector vận tốc, thể hiện sự thay đổi vị trí của điểm ảnh từ frame thứ t sang frame t+1.



Hình 3 - Frame ảnh tại thời điểm t trước và sau khi vẽ các vector có hướng.

3.3 Đề xuất giải pháp

Có nhiều kỹ thuật tiếp cận để phát hiện chuyển động trong hình ảnh Video liên tục. Có thể so sánh khung hình hiện tại với hình nền chúng ta chụp từ ban đầu khi bật camera hoặc từ khung hình trước. Đối với kỹ thuật thứ nhất thì đơn giản và giảm được việc xử lý. Tuy nhiên, cách tiếp cận có một bất lợi lớn, ví dụ nếu có một đối tượng đang di chuyển ở frame đầu tiên nhưng sau đó nó đã biến mất. Kỹ thuật thứ hai thì xử lý phức tạp hơn, xử lý nhiều hơn nhưng lại thích nghi với mọi môi trường, kể cả môi trường ít thay đổi hoặc thay đổi nhiều. Nhược điểm là nếu đối tượng di chuyển một cách rất chậm thì hệ thống không phát hiện ra. Nhưng có thể giải quyết bằng cách tăng số khung hình trên giây. Giải pháp mà bài báo muốn đề xuất là kết hợp phương pháp Optical Flow với phương pháp trích chọn mẫu.

Giải thuật đề xuất:

Gọi $x_t = \{x_{m,t}\ ;\ m = 1,\ldots,M\}$ là tập các đối tượng tại thời điểm t. Trong đó, M là số đối tượng có trong hệ thống, M có thể thay đổi theo thời gian. Gọi $x_t^* = \{x_{d,t}^*; d = 1,...,D\}$ là tập biểu diễn kết quả phát hiện đối tượng của hệ tại thời điểm t tương ứng. Ta có

$$x_t^* = \{x_{d,t}^*; d = 1,...,D\} = Detect(z_t)$$

với D là số đối tượng phát hiện được.

Gọi $x_{old,t}^* = \{x_{d,t}^* \in x_t^*; \min \| x_{d,t}^* - x_{m,t-1} \| \leq d_{thresh}\}$, với ngưỡng cho trước d_{thresh}, là tập các kết quả phát hiện "cũ", được hiểu theo nghĩa, nếu một phát hiện trong thời điểm t quá gần với một trạng

thái đã có tại thời điểm t-1 thì nó sẽ được xem là trùng với đối tượng đó. Một cách gần đúng, ta giả định những phát hiện này xuất phát từ đối tượng đã có từ thời điểm t-1 trước đó.

Tương tự, ta định nghĩa $x^*_{new,t} = x^*_d / x^*_{old,t}$ là tập những phát hiện "mới", được hiểu là giữa tập các điểm mới và tập các điểm cũ cách nhau một khoảng là *d*.

Thuật toán thực nghiệm:

Khởi tạo (t=0)

Với i=1,…,N

Gán $w_0^{(i)} = 1/N$ và sinh ra $w_0^{(i)} \sim p(x_0)$

Với t=1,2,…

- Bước lấy mẫu đối tượng lần đầu tiên

$$\hat{x}_{opticalflow} = Opticalflow(\hat{x}_{t-1})$$

Với i=1,…,N

Sinh ra $w_t^i \sim q(x_t \mid x_{t-1}^i, z_t) = \alpha N(\hat{x}_{opticalflow}, P) + (1-\alpha)\, p(x_t \mid x_{t-1})$

- Các đối tượng giống nhau lân cận $p(z \mid x_t^i) = N(D_z, 0, \sigma^2)$, với D_z là hệ số khoảng cách giữa 2 đối tượng giống nhau.

- Tính các trọng số quan trọng (chưa được chuẩn hóa)

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p(z_t \mid x_t^{(i)})\, p(x_t^{(i)} \mid x_{t-1}^{(i)})}{q(x_t^{(i)} \mid x_{t-1}^{(i)}, z_t)}$$

Với i=1,…,N

- Chuẩn hóa các trọng số quan trọng

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$$

-Bước lấy lại mẫu

Ước lượng kích thước mẫu

$$\hat{N}_{eff} = \frac{1}{\sum_{j=1}^{N} (w_k^{(j)})^2}$$

Nếu $\hat{N}_{eff} < N_{th}$

-Đạt được các mẫu mới $\{x_k^{(j_i)}\}_{i=1}^N$ bởi việc lấy mẫu N lần và thay thế $\{x_k^{(j_i)}\}_{i=1}^N$ sao cho $\Pr\{x_k^{(j_i)} = x_k^{(j)}\} = w_k^{(j)}$

-Khởi tạo $w_k^i = 1/N$

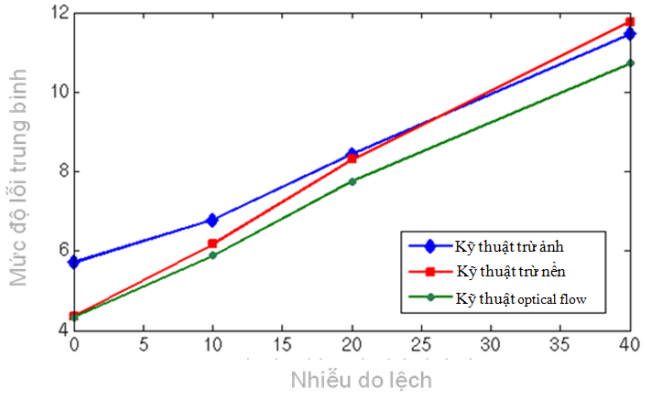
-Bước xuất kết quả

$$\hat{x}_{t|t} = \sum_{i=1}^{N} w_t^{(i)} x_t^{(i)}$$

-Kết thúc

4. Kết quả

Sau khi tiến hành thử nghiệm và so sánh với các kỹ thuật trừ ảnh và trừ nền về mức độ lỗi trung bình, độ nhiễu và tỷ lệ chính xác khi gặp phải nguồn ảnh hoặc nguồn video chất lượng thấp thì phương pháp Optical flow kết hợp tái chọn mẫu đạt được độ ổn định hơn qua bảng đánh giá sau (được khảo sát thử nghiệm trên 1000 frame ảnh):

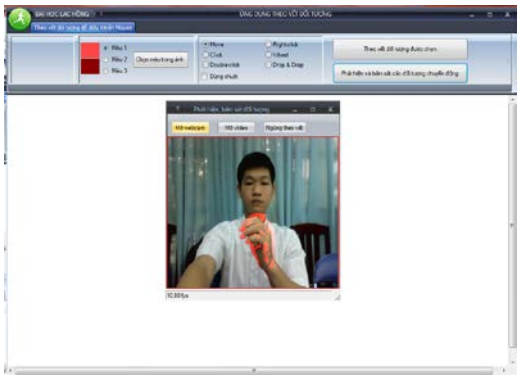


Hình 4 - Biểu đồ so sánh mức độ lỗi và nhiễu giữa các kỹ thuật trừ ảnh, trừ nền, Optical flow

Kết quả đạt được là một ứng dụng cho phép phát hiện và bám sát đối tượng từ webcam gắn trực tiếp với máy tính hoặc từ một file Video (định dạng AVI). Sau đây là các chức năng chính của ứng dụng:

4.1 Phát hiện và bám sát tất cả các đối tượng đang chuyển động

Chức năng này cho phép mở các Video từ một file AVI hoặc từ camera(webcam) gắn trực tiếp với máy tính.

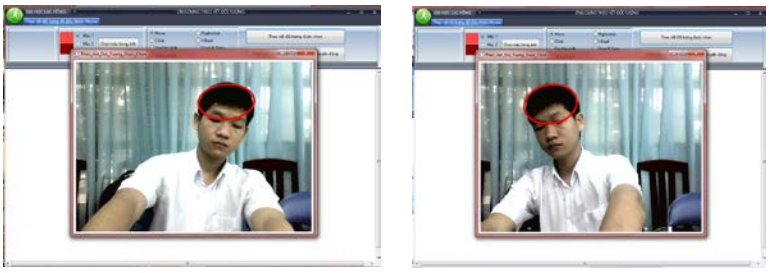


Hình 5 - Phát hiện đối tượng chuyển động

4.2 Chức năng theo vết đối tượng được lựa chọn để theo vết

Chức năng này giám sát đối tượng mà mình muốn theo vết

Quét chọn đối tượng cần theo vết và đối tượng đó sẽ bị theo vết



Hình 6 - Theo vết đối tượng được chọn đang di chuyển

4.3 Chức năng dùng đối tượng đang theo vết để điều khiển "Mouse"

Chức năng này theo màu của một đối tượng và xác định vị trí của đối tượng để đưa con trỏ tới đúng vị trí đối tượng đang đứng, mục đích là muốn thông qua đối tượng bên ngoài như bàn tay để qua camera có thể duyệt web, sử dụng các thao tác căn bản như Move, Click, Right click, Double click, Zoom out, Zoom in…

Hình 7 - Duyệt web và ra lệnh phóng lớn(Zoom Out)

5. Kết luận

Bài báo đã nghiên cứu một số kỹ thuật phát hiện và bám sát đối tượng, đồng thời tiến hành xử lý cho ra kết quả là đối tượng đang cần theo vết đang ở vị trí nào để đánh dấu (tô màu, kẻ khung. Sau khi xác định vị trí đối tượng, ứng dụng sẽ tiếp tục điều khiển thiết bị chuột tới vị trí mong muốn

(vị trí của đối tượng đang theo vết), đồng thời quyết định ra sự kiện gì (Click, Double click, Drap & Drop, Zoom out, Zoom in…)

Hệ thống đầu vào trong bản demo này là lấy hình ảnh trực tiếp từ webcam gắn vào máy tính hoặc lấy một file Video có phần đặc trưng là AVI từ nguồn có sẵn hoặc các Video lấy từ nguồn trực tuyến.

Trong quá trình thực hiện thu nhận ảnh (từ webcam) thường bị biến dạng do các thiết bị thu nhận chất lượng thấp dẫn tới việc cân chỉnh lại rất phức tạp vì nó phụ thuộc quá nhiều vào môi trường xung quanh (bị nhiễu, thay đổi ánh sáng, độ tương phản …). Do đó các công việc như khử nhiễu, cân chỉnh mức xám thường được xác định thông qua các ngưỡng (Thresshold) trong chương trình và do người sử dụng quyết định (tinh chỉnh) và chưa có khả năng tự động cân bằng.

Hướng nghiên cứu sâu hơn của đề tài này là phát hiện và phân loại từng phần chuyển động của đối tượng, từ đó xây dựng các ứng dụng hỗ trợ cho con người nhằm xây dựng các ngôi nhà thông minh hoặc tích hợp cho các robot tự hành.

TÀI LIỆU THAM KHẢO

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. International Journal of Computer Vision, 12(1):43–77, 2007.

[2] R. T. Collins et al. A system for Video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2006.

[3] David J. Fleet, Yair Weiss. Optical flow estimation, Mathematical models for Computer Vision: The Handbook. N. Paragios, Y. Chen, and O. Faugeras (eds.), Springer, 2005.

[4] B. D. Lucas and T. Kanade. An iterative image registration technique with an application in stereo vision. In Seventh International Joint Conference on Artificial Intelligence, pages 674–679, Vancouver, 2007.

[5] Anurag Mittal and Mikos Paragios, "Motion Based Background Subtraction using Adaptive Kernel Density Estimation" pp. 302-309, 2004

[6] C.Stauffer and W.Grimson, "Adaptive Background mixture models for Real-time tracking" pp. 750-755, 2009.

[7] Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object Tracking: A Survey" pp. 7-15.