

**ĐẠI HỌC THÁI NGUYÊN**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

-----

**Hoàng Xuân Lộc**

**HỆ THỐNG QUAN SÁT ĐA MỤC TIÊU**  
**TỐI ƯU HÀM ĐÁNH GIÁ ĐỐI TƯỢNG**

**Chuyên ngành: Khoa học máy tính**

**Mã số: 60480101**

**TÓM TẮT LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN**

**Thái Nguyên - 2015**

## LỜI CẢM ƠN

Lời đầu tiên, Em xin cảm ơn Thầy TS. Đào Nam Anh đã trực tiếp hướng dẫn, giúp đỡ và tận tình chỉ bảo Em trong suốt thời gian từ khi có ý tưởng cho luận văn đến khi hoàn thành luận văn này.

Em xin chân thành cảm ơn đến các Thầy bộ môn đã trực tiếp dạy dỗ Em như Thầy PGS.TS Ngô Quốc Tạo, PGS. TS. Phạm Việt Bình, PGS.TSKH Nguyễn Xuân Huy, PGS.TS Đỗ Năng Toàn,...đã giúp Em có kiến thức và ý tưởng cho luận văn.

Em xin cảm ơn các Thầy trong hội đồng bảo vệ đề cương đã góp ý, giúp Em có những thay đổi phù hợp hơn trong luận văn này.

Em xin cảm ơn trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên đã tạo rất nhiều điều kiện thuận lợi cho Em.

## **LỜI CAM ĐOAN**

Tôi xin cam đoan luận văn là kết quả nghiên cứu của tôi, không sao chép của ai. Nội dung luận văn có tham khảo và sử dụng các tài liệu liên quan, các thông tin trong tài liệu được đăng tải trên các tạp chí và các trang website theo danh mục tài liệu của luận văn.

**Tác giả luận văn**

**Hoàng Xuân Lộc**

## MỤC LỤC

DANH MỤC HÌNH VẼ.....	V
MỞ ĐẦU.....	1
CHƯƠNG 1. TỔNG QUAN THEO DỐI ĐỐI TƯỢNG.....	3
1.1 Xử lý video .....	3
1.2 Bài toán theo dõi đối tượng .....	7
1.3 Ứng dụng thực tế - giám sát thông minh .....	10
1.4 Thành phần của thuật toán.....	14
1.5 Các thuật toán so sánh .....	25
1.6 Phát hiện các đối tượng chuyển động .....	26
CHƯƠNG 2. THUẬT TOÁN THEO DỐI ĐA MỤC TIÊU TỐI ƯU HÀM ĐÁNH GIÁ ĐỐI TƯỢNG.....	35
2.1 Ký hiệu.....	37
2.2 Phát hiện các đối tượng.....	37
2.3 Hàm đánh giá đối tượng .....	44
2.4 Tối ưu hoá hàm đánh giá đối tượng.....	48
2.5 Theo vết đối tượng.....	51
2.6 Thuật toán .....	56
2.7 Kết luận chương 2.....	57
CHƯƠNG 3. CÀI ĐẶT THỬ NGHIỆM .....	58
3.1 Môi trường cài đặt .....	58
3.2 Trích dẫn một số mã nguồn .....	58
3.3 Kết quả thực nghiệm.....	63
3.4 Kết luận chương 3.....	74
KẾT LUẬN.....	75
TÀI LIỆU THAM KHẢO.....	76

## DANH MỤC HÌNH VẼ

Hình 1-1 Cấu trúc phân cấp của video .....	4
Hình 1-2. Khung hình khác nhau nhưng có cùng biểu đồ màu .....	5
Hình 1-3. Theo dõi khách bộ hành.....	7
Hình 1-4 Hệ thống camera điều khiển xe tự hành SCABOR .....	7
Hình 1-5 Mô hình hệ thống giám sát .....	11
Hình 1-6: Hệ thống camera giám sát giao thông. ....	13
Hình 1-7 Hệ thống camera giám sát an ninh.....	13
Hình 1-8: Công nghệ nhận diện khuôn mặt. ....	14
Hình 3-9: Bộ cơ sở dữ liệu .....	66
Hình 3-10 IMG1: Đối tượng ảnh số 1 được theo dõi qua các khuôn hình từ 3-16 ..	67
Hình 3-11 IMG2: Đối tượng ảnh số 14 được theo dõi qua các khuôn hình 85-88, bị khuất trong khuôn hình 89, xuất hiện lại trong hình 90 với mã số mới 16.....	68
Hình 3-12 IMG2: Đối tượng ảnh số 16 được theo dõi tiếp qua các khuôn hình 140-154.....	69
Hình 3-13 IMG2: Đối tượng ảnh số 16 được theo dõi tiếp trong mã số 29 qua các khuôn hình 195-210. Có một số đối tượng mới xuất hiện, tuy nhiên còn nhỏ nên không hiện mã số trên màn hình .....	70
Hình 3-14 IMG3: Có 2 đối tượng ảnh trong khuôn hình 666-680. Do kích thước của đối tượng nhỏ hơn ngưỡng đặt trước nên mã số không hiển thị trên khuôn hình.....	71
Hình 3-15 IMG3: Đối tượng ảnh số 89 được theo dõi trong các khuôn hình 780-800. Một đối tượng ảnh đi cùng không được hiển thị mã số do kích thước của đối tượng nhỏ hơn ngưỡng đặt trước.....	72

## MỞ ĐẦU

Theo dõi đa đối tượng là một nhiệm vụ đã có từ lâu với ngành an ninh, nay là nhiệm vụ quan trọng trong lĩnh vực thị giác máy tính. Sự gia tăng của máy tính cấu hình mạnh, sự sẵn có của các máy quay video chất lượng cao và không tốn kém, và nhu cầu ngày càng tăng về phân tích video tự động đã tạo ra rất nhiều sự quan tâm về các thuật toán theo dõi đa đối tượng.

Các thuật toán này đã có ứng dụng rộng rãi là hệ thống giám sát thông minh tự động - trợ giúp đắc lực cho con người thực hiện theo dõi, giám sát. Ví dụ giám sát giao thông một hệ thống giám sát thông minh có thể cho ta biết được số lượng phương tiện lưu thông qua đoạn đường, đưa ra thông tin về tốc độ chuyển động, đường đi của đối tượng được theo dõi.

Khó khăn chính của bài toán theo dõi đa đối tượng là sự phức tạp của không gian quan sát với số lượng lớn quỹ đạo đối tượng liên tục, và một số quỹ đạo rời rạc. Để giải quyết theo dõi tương tác giữa các đối tượng cần phải có các mô hình thuật toán tối ưu, đảm bảo phân tích được trong thời gian thực. Ý thức được những lợi ích mà hệ thống giám sát thông minh mang lại: ” Hệ thống quan sát đa đối tượng tối ưu hàm đánh giá đối tượng” để từ đó nắm được các thuật toán giải quyết bài toán theo dõi giám sát đa đối tượng chuyển động trong video với sự tối ưu hóa hàm đánh giá đối tượng mà kết quả là tối ưu nguồn lực máy tính.

Đối tượng của luận văn này là tìm hiểu kết quả nghiên cứu trong lĩnh vực giám sát đa đối tượng (multi object tracking) di chuyển trong video bằng sự tối ưu hàm đánh giá đối tượng. Luận văn sẽ phân tích và thực nghiệm thuật toán cho bài toán trên.

Ngoài phần mở đầu và kết luận, luận văn được chia làm 3 chương, luận văn có các chương như sau:

Chương 1: Tổng quan về theo dõi đối tượng trong xử lý ảnh

Trình bày các vấn đề đặt ra cần giải quyết trong hệ thống theo dõi thông minh, giới thiệu bài toán phát hiện, phân loại, theo dõi đối tượng chuyển động trong hệ thống theo dõi đối tượng.

Chương 2: Phương pháp phát hiện và theo dõi đa đối tượng

Phân tích giải pháp bài toán theo dõi đa đối tượng thời gian thực bằng cách tối ưu hàm đánh giá đối tượng. Trình bày chi tiết về mô hình theo dõi đối tượng và xây dựng hàm số hàm đánh giá đối tượng. Tối ưu hàm đánh giá đối tượng cục bộ hàm đánh giá đối tượng để đạt giá trị cực tiểu của hàm đánh giá đối tượng, từ đó tính toán theo vết đường đi của đối tượng trong thời gian thực.

Chương 3: Thực nghiệm và đánh giá

Trình bày về việc cài đặt chương trình, xây dựng dữ liệu thực nghiệm, các quá trình thực nghiệm, kết quả thực nghiệm và các kết quả đánh giá, nhận xét các xử lý từ thực nghiệm.

## CHƯƠNG 1. TỔNG QUAN THEO ĐỐI ĐỐI TƯỢNG

Giám sát bằng video là quá trình phân tích khung hình tuần tự trong video.. Đó là một ứng dụng của thị giác máy tính. Có ba loại hoạt động giám sát bằng video: giám sát bằng video có thể là bằng tay, bán tự động hoặc hoàn toàn tự động.

Giám sát video bằng tay là việc phân tích các nội dung video được con người thực hiện trực tiếp. Hệ thống dạng này đang được sử dụng rộng rãi.

Video giám sát bán tự động liên quan đến một số hình thức xử lý video với sự can thiệp đáng kể của con người. Ví dụ điển hình là các hệ thống phát hiện chuyển động đơn giản. Chỉ khi có sự chuyển động đáng kể thì video mới được ghi lại và gửi cho chuyên gia là con người phân tích.

Với một hệ thống hoàn toàn tự động, không có sự tham gia của con người vào quá trình phân tích. Trong một hệ thống như vậy hệ thống làm cả hai nhiệm vụ cấp thấp, là phát hiện chuyển động và theo dõi, và nhiệm vụ ra quyết định cấp cao, như nhận diện sự kiện bất thường và xác định loại chuyển động. Hệ thống giám sát video hỗ trợ động phân loại tự các đối tượng và theo dõi đối tượng. Điều hành hệ thống giám sát video trong thời gian dài trực tiếp bằng con người là không thực tế và khó khả thi.

### 1.1 Xử lý video

#### 1.1.1. Một số khái niệm

Video là tập hợp các khung hình, mỗi khung hình là một ảnh. Video hay còn gọi là chuỗi ảnh (image sequence) tượng trưng cho thông tin hình ảnh. Đó là một chuỗi các hình ảnh liên tục theo thời gian.



$$S=f(x,y,t) \quad (1-1)$$

Trong đó:

$x, y \in R$ : là tọa độ của điểm ảnh (thông tin về không gian).

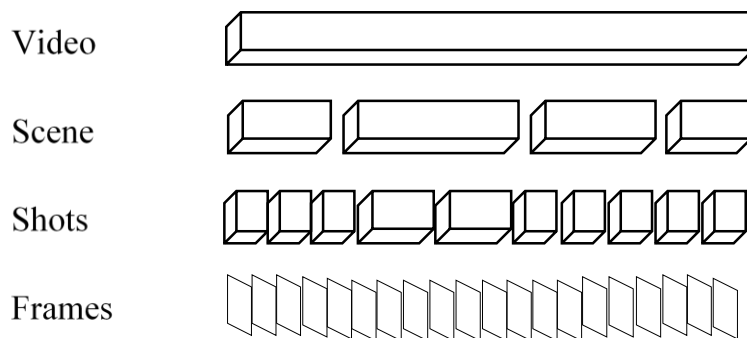
$t \in R$ : thông tin về thời gian.

Dựa trên công thức 1.1 ta thấy ảnh tĩnh là một trường hợp đặc biệt của video. Khi đó nó là một chuỗi các hình ảnh không thay đổi theo thời gian.

$$f(x, y, t_1) = f(x, y, t_2); \quad \forall x, y \in R \quad (1-2)$$

**Shot (liạ):** là một đơn vị cơ sở của video. Một lia là một đơn vị vật lý của dòng video, gồm chuỗi các khung hình liên tiếp, không thể chia nhỏ hơn.

**Scene (cảnh):** là các đơn vị logic của dòng video. Cấu trúc phân cấp của video được mô tả trong hình vẽ.



Hình 1-1 Cấu trúc phân cấp của video

Khi phim được chiếu, các khung hình lần lượt được hiển thị ở tốc độ nhất định. Tốc độ thường thấy ở các định dạng video khác nhau là 30 và 24 hình/s. Như vậy một giờ video sẽ có số lượng khung hình tương ứng là 108000 hoặc 86400. Dù là video ở định dạng nào thì nó cũng có dung lượng rất lớn và nếu xử lý với tất cả các khung hình thì thật không hiệu quả.

Phân đoạn là quá trình phân tích và chia nội dung hình ảnh video thành các đơn vị cơ sở gọi là các lia (shots). Việc lấy mẫu chính là chọn gán đúng một khung hình video đại diện cho mỗi lia (hoặc nhiều hơn tùy theo mức độ phức tạp của nội dung hình ảnh của lia) và được gọi là các khung – khóa [3].

Khung – khóa là khung hình đại diện mô tả nội dung chính của shot.

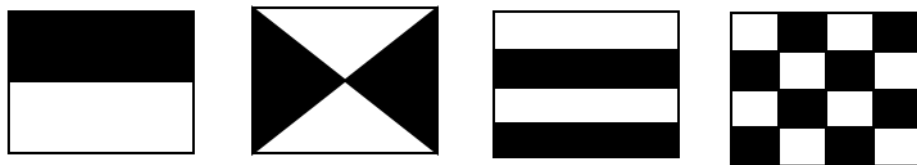
Quá trình phân đoạn dữ liệu video tiến hành phân tích, phát hiện sự chuyển đổi từ lia này sang lia khác hay chính là sự phát hiện ranh giới giữa các lia đó (đó chính là đo sự khác nhau giữa các khung hình liền kề).

### **1.1.2. Một số thuộc tính đặc trưng**

Video có bốn đặc tính chung là: màu (color), kết cấu (texture), hình dáng (shape), và chuyển động (motion).

#### **Màu (Color):**

Màu là một thuộc tính đặc trưng của ảnh. Biểu đồ ảnh, biểu diễn sự phân bố màu, là một đặc trưng màu phổ biến nhất hiện nay. Biểu đồ màu không phụ thuộc vào sự quay, dịch chuyển ảnh cũng như chiều nhìn ảnh. Tuy nhiên với biểu đồ màu không biểu diễn thông tin về không gian phân bố các điểm ảnh, do đó hai điểm ảnh có cùng biểu đồ màu có thể lại có nội dung khác nhau.



Hình 1-2. Khung hình khác nhau nhưng có cùng biểu đồ màu

#### **Kết cấu (Texture):**