

# ĐỀ XUẤT MÔ HÌNH HỆ THỐNG QUẢNG CÁO TỰ ĐỘNG TÙY BIẾN THEO KHÁCH HÀNG TRONG NGỮ CẢNH THỜI GIAN THỰC

Lê Nguyên Bảo<sup>1</sup>, Lê Đắc Như<sup>2</sup>, Lê Văn Chung<sup>1</sup>, Nguyễn Gia Như<sup>1</sup>, Đỗ Năng Toàn<sup>3</sup>

<sup>1</sup> Đại học Duy Tân, Đà Nẵng

<sup>2</sup> Khoa Công nghệ thông tin, Đại học Hải Phòng

<sup>3</sup> Viện Công nghệ thông tin, Đại học Quốc Gia Hà Nội

{lenguyenbao, levanchung, nguyengianhu}@duytan.edu.vn, nhuongld@hus.edu.vn, dntoan@vnu.edu.vn

**TÓM TẮT**— Tùy biến quảng cáo là thay đổi các nội dung quảng cáo theo ngữ cảnh cụ thể. Điều này làm cho nội dung quảng cáo phù hợp và thực sự hữu ích đối với các khách hàng trong ngữ cảnh thời gian thực. Bài báo này trình bày đề xuất mô hình hệ thống quảng cáo tự động tùy biến theo đối tượng khách hàng. Từ những hình ảnh thực thu nhận được từ camera, hệ thống sẽ phân tích các đối tượng dựa trên tập đặc trưng cho trước để xác định lớp khách hàng phù hợp. Dựa trên lớp khách hàng được xác định, hệ thống sẽ truy xuất đến cơ sở dữ liệu quảng cáo đa phương tiện và tự động lựa chọn, phát những nội dung phù hợp. Với mỗi giai đoạn, chúng tôi sẽ tập trung phân tích và đánh giá các kỹ thuật được sử dụng để nâng cao tính khả thi và hiệu quả của hệ thống.

**Từ khóa**— Quảng cáo tùy biến người dùng, nhận dạng đối tượng, phân lớp đối tượng.

## I. GIỚI THIỆU

Trong những năm gần đây, các loại hình quảng cáo đã có sự phát triển mạnh mẽ trên cả chiều rộng mà cả chiều sâu. Nói đến chiều rộng của nó, chúng ta thấy quảng cáo có mặt ở khắp mọi nơi, từ những quốc gia tư bản phát triển đến các nước đang phát triển theo chủ nghĩa xã hội. Quảng cáo đã trở thành một hoạt động kinh tế quan trọng trong nền kinh tế thị trường. Đây là xu thế tất yếu, là một trong những phương tiện hữu hiệu thể hiện sự cạnh tranh thương mại, là một động lực kích thích phát triển nền kinh tế. Về bề sâu, quảng cáo không những đã làm biến dạng những cách thức sinh hoạt của người dùng mà còn thay đổi tư duy, ảnh hưởng sâu sắc đến văn hóa của nhiều tầng lớp người trong xã hội. Trong bài báo này, chúng tôi muốn nhìn nhận theo một chiều thứ ba nữa, chiều phát triển kỹ thuật cho các mô hình quảng cáo. Ban đầu, các kỹ thuật quảng cáo từ những phương tiện thô sơ như lời đồn, tin tức truyền miệng giữa bạn bè, lời dẫn giải của người bán hàng đến các hình thức mới khác như: tặng quà, trao giải,... đã tận dụng sức mạnh của các phương tiện truyền đại chúng như: nhật báo, tạp chí, truyền thanh, truyền hình, điện ảnh.

Ngày nay, quảng cáo đã có sự hội tụ và bước phát triển mới khi tiến lên mô hình truyền thông đa phương tiện với với khả năng tương tác cao thông qua mạng Internet. Đây cũng là một thách thức của hình thức quảng cáo mới, năng động, tương phản với hình thức quảng cáo đơn phương một chiều khi con người thụ động trước các hệ thống quảng cáo. Sự tùy biến quảng cáo với khả năng cập nhật theo thời gian thực hướng đến thay đổi quảng cáo văn bản theo ngữ cảnh đầy đủ của một tìm kiếm hoặc trang web ai đó đang xem như: Google AdWords, Google AdSense. Đây là công nghệ quảng cáo hiển thị theo ngữ cảnh (Contextual Advertising hoặc Targeting Advertising) và rất hiệu quả [1]. Với các hình thức quảng cáo phổ biến như logo, banner, pop-up... theo thời gian, người dùng khi truy cập vào trang web đều có thể nhìn thấy quảng cáo. Tuy nhiên, nhược điểm của hình thức này là phụ thuộc vào lượng truy cập không đều, khó kiểm soát, tính định hướng thấp, nội dung quảng cáo khó nhắm trúng đối tượng khách hàng mục tiêu do quảng cáo chỉ đặt ở một website cố định và hiển thị ở cả những bài viết, chuyên mục không liên quan.

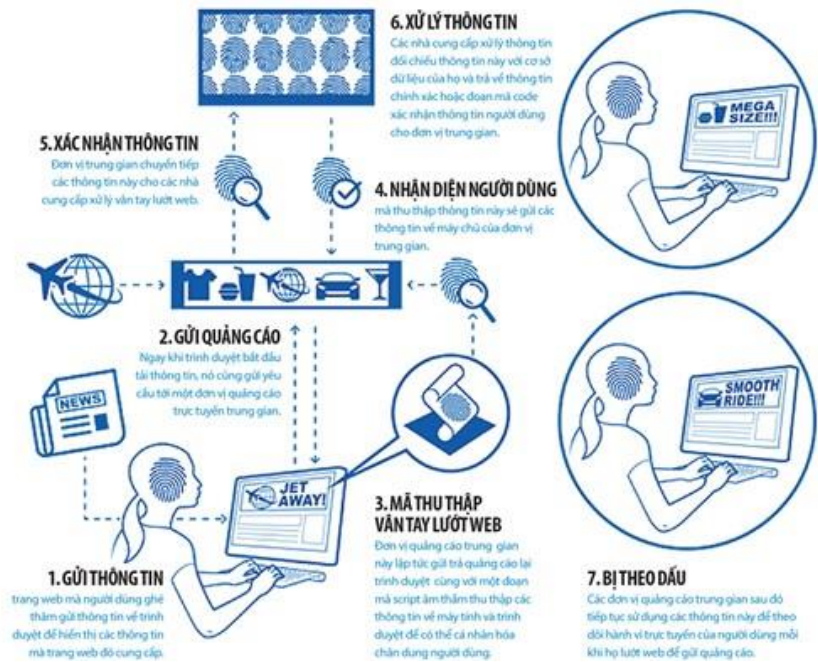
Công nghệ tiếp theo là hiển thị nội dung quảng cáo theo ngữ cảnh trên các website hoặc các phương tiện truyền thông khác như điện thoại di động hướng đến người dùng dựa vào ngữ cảnh của bài viết, vị trí địa lý, thời gian của người dùng tiếp cận với quảng cáo và thói quen truy cập của khách hàng tiềm năng. Công nghệ hiển thị quảng cáo theo ngữ cảnh xác định nội dung trên trang website, xác định hoàn cảnh sử dụng của người dùng để đưa ra quảng cáo phù hợp nhất với nhu cầu của người dùng. Các tính năng chính của công nghệ tùy biến hiển thị quảng cáo theo ngữ cảnh là tự động phân tích người dùng, chọn đối tượng hiển thị quảng cáo, vị trí địa lý, quản lý ngân sách và hệ thống đặt giá trực tuyến. Đồng thời hệ thống tính phí linh hoạt: thu phí theo số lượt hiển thị (Cost Per Mile), theo giá trị của mỗi lần click (Cost Per Click) hoặc theo hoạt động tương tác của người dùng trên quảng cáo (Cost Per Action) đã giúp các doanh nghiệp tối ưu hóa chi phí quảng cáo trong khi đảm bảo tính hiệu quả của nó [2].

Trong bài báo này, chúng tôi sẽ tập trung phân tích và đề xuất mô hình hệ thống quảng cáo tự động tùy biến theo đối tượng khách hàng trong ngữ cảnh thời gian thực được áp dụng trong tùy biến các nội dung video quảng cáo với mục tiêu hướng đến là làm cho nội dung quảng cáo phù hợp và thực sự hữu ích đối với các khách hàng trong từng ngữ cảnh cụ thể.

## II. MỘT SỐ NGHIÊN CỨU LIÊN QUAN

Đối với các hệ thống quảng cáo trực tuyến, điều quan trọng nhất để đưa ra được các quảng cáo phù hợp người dùng là âm thầm theo dõi thói quen lướt web của người dùng ngay cả khi không được phép. Khi đó, các nhà quảng cáo trực tuyến đã yêu cầu các trang web để thu thập dữ liệu lướt web của người dùng, thậm chí cả hồ sơ cá nhân chi tiết với sở thích và hành vi của họ. Các hồ sơ dữ liệu này có thể được phân tích và cho phép các doanh nghiệp quảng cáo trực tuyến nhắm tới những cộng đồng mục tiêu cụ thể. Có thể người dùng không để tâm song thói quen lướt web chứa rất

nhiều thông tin mà các doanh nghiệp quan tâm. Vì qua đó, họ có thể xác định được các người dùng tiềm năng được dựng nên qua những thông tin về nhân thân được tiết lộ trên mạng xã hội, lịch sử mua hàng hay các địa điểm đã ghé qua bị lưu vết trên mạng bởi chính điện thoại người dùng. Có thể nói không cần đến cơ quan chuyên trách của chính phủ, các công ty quan tâm đến dữ liệu lướt web của người dùng cũng rút ra được nhiều điều về người dùng qua những thông tin này. Sau đây chúng ta sẽ xem xét quá trình theo dõi lịch sử duyệt web gọi là lưu vân tay (*fingerprinting*) để do thám thông tin người dùng. Công cụ thu thập dữ liệu web sớm nhất là Cookie-một dạng bản ghi được tạo ra và lưu lại trên trình duyệt khi người dùng truy cập một website để định danh bạn. Khi truy cập vào một trang web, website này sẽ đặt một cookie trên máy đó, thay cho việc liên tục hỏi các thông tin như nhau, chương trình trên website có thể sao lưu thông tin vào một cookie mà khi cần thông tin nó sẽ đọc cookie chứ không cần yêu cầu người dùng nhập lại thông tin. Tiếp đến phải kể đến kỹ thuật theo dõi người dùng mới phổ biến là sử dụng Flash cookie với plug-in Adobe Flash. Trước đây, các trang web thường giấu thông tin vào trong Flash cookie để trốn tránh việc phát hiện và gỡ bỏ của người dùng. Thông tin tại đây sau khi quét sẽ tự động sản sinh ra các cookie thông thường cài vào trình duyệt [1].



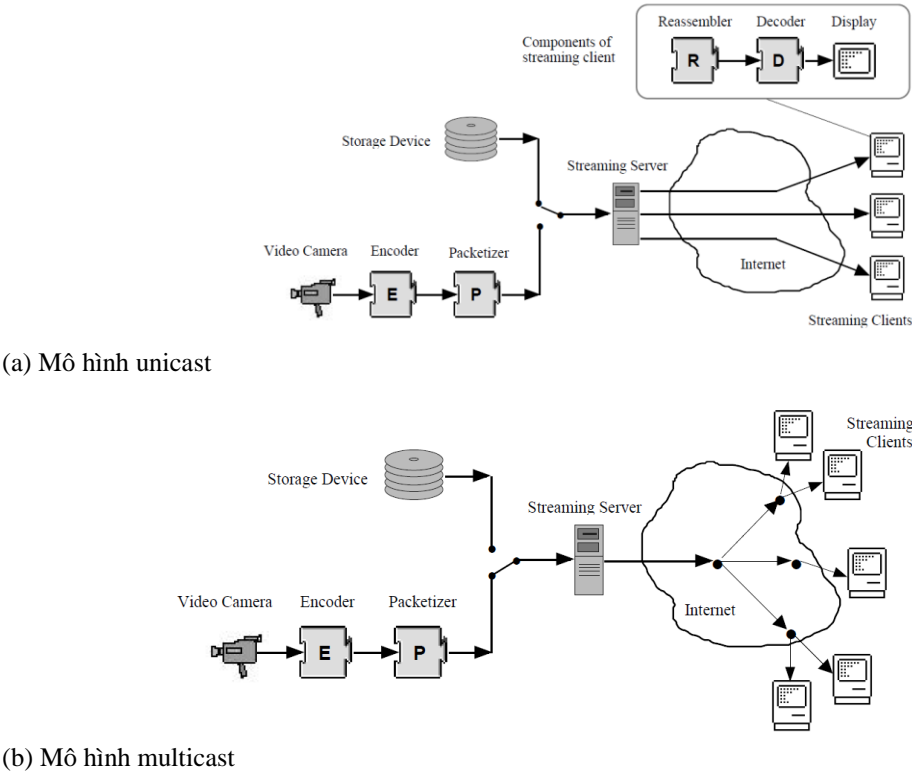
Hình 1. Quá trình theo dõi người dùng và tùy biến quảng cáo trên các trang Web

Về nội dung các video quảng cáo trên mạng đều sử dụng kỹ thuật Streaming video [2]. Đây là kỹ thuật sử dụng cách thức phát lại các đoạn video được lưu trữ trên các máy tính trên mạng tới người dùng đầu cuối muốn xem đoạn video mà không cần tải đoạn video đó về trên máy tính. Về bản chất, streaming video là quá trình chia nhỏ file video thành các frame, rồi lần lượt gửi từng frame tới một bộ đệm trên máy tính của người xem và hiển thị nội dung frame đó. Với các định dạng tập tin Video truyền thống, dữ liệu chỉ có thể hiển thị khi đã được tải về toàn bộ, vì vậy đối với các tập tin Video chất lượng cao có dung lượng lớn thì công việc này sẽ tiêu tốn rất nhiều thời gian. Streaming Video tiết kiệm thời gian cho người dùng bằng cách sử dụng các công nghệ giải nén kết hợp với player hiển thị dữ liệu đồng thời trong lúc vẫn tiếp tục download gọi là buffering. Streaming Video được thể hiện dưới hai dạng:

- Video theo yêu cầu (*on demand*): là các dữ liệu Video được lưu trữ trên multimedia server và được truyền đến người dùng khi có yêu cầu, người dùng có toàn quyền để hiển thị cũng như thực hiện các thao tác (tua, dừng, qua ...) với các đoạn dữ liệu này.
- Video thời gian thực (*live event*): là các dữ liệu Video được convert trực tiếp từ các nguồn cung cấp dữ liệu theo thời gian thực (máy camera, microphone, thiết bị phát dữ liệu Video...).

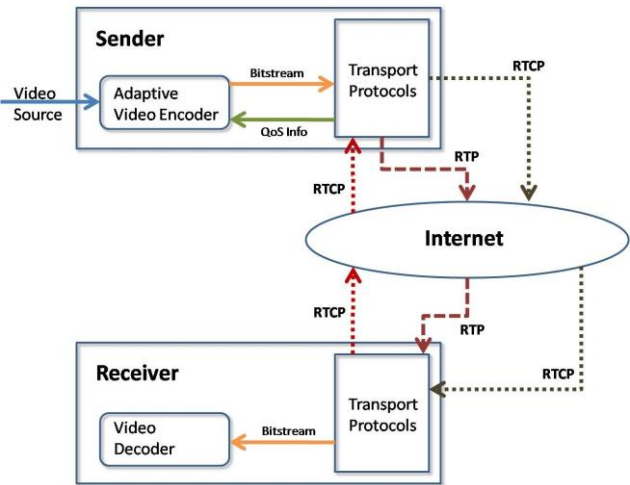
Và quá trình này tuân thủ chặt chẽ về ràng buộc theo thời gian, nói khác là tuân thủ chặt chẽ theo giao thức RTSP (*Real Time Streaming Protocol*), RTP (*Real-time Transport Protocol*) và RTCP (*Real Time Control Protocol*) [16]. Kiến trúc hệ thống streaming được mô tả như Hình 2. Trong đó, Streaming Server có nhiệm vụ phân phối Stream cho mỗi thiết bị người dùng. Streaming Server hoạt động sẽ lấy nội dung (những tập tin Video hoặc những tập tin khác trong hệ thống) và tạo Stream cho mỗi yêu cầu người dùng gửi đến. Những Stream này có thể hiện thực theo hai phương thức (mỗi phương thức được điều khiển bởi cơ chế khác nhau) là unicast và multicast. Phổ biến nhất hiện nay là cơ chế unicast. Chức năng thường được tích hợp trong Streaming Server là lưu trữ và phục hồi nội dung. Công việc xử lý quan trọng của Streaming Server là tạo một gói tin cho mỗi Stream theo thời gian thực. Ngoài ra, Streaming Server còn phải giải quyết các yêu cầu về mã hóa gói dữ liệu để đảm bảo tính bảo mật, nén dữ liệu để gửi đến những kênh Stream có tùy chọn kết nối với tốc độ khác nhau. Với những hệ thống máy chủ cho phép phát Video trực tuyến hoặc cung cấp cho việc tải về theo yêu cầu người dùng đang được sử dụng hiện nay, thì với mỗi yêu cầu của người dùng đến - sẽ được máy chủ xử lý thông tin đảm bảo yêu cầu đó là hợp lệ để thiết lập kết nối; tiếp sau đó, nó sẽ truy xuất đến ổ cứng (hoặc thiết bị

lưu trữ cục bộ) để truy xuất tập tin yêu cầu đó và gửi về cho người dùng. Như chúng ta đều biết tốc độ truy xuất trên ổ cứng thường là khá chậm. Bên cạnh đó, có những tập tin được nhiều người dùng cùng quan tâm trong một thời điểm (sai khác nhỏ về thời gian), việc truy xuất ổ cứng liên tục để truy xuất cùng một tập tin như cách thức làm việc hiện tại là rất lãng phí và tốn kém cho máy chủ. Một vấn đề quan trọng khác trong hệ thống là phân tích dữ liệu và đưa ra giải thuật giải quyết vấn đề quản lý Cache một cách hiệu quả trong Streaming Server.



Hình 2. Kiến trúc hệ thống Streaming Video

Kiến trúc MCN Streaming (*Multimedia Communications and Networking*) [3] mức tổng quát với nền tảng tương thích được biểu diễn trong Hình 3. Trong đó, MCN streaming bao gồm hai thành phần: người gửi và người nhận. Người gửi có trách nhiệm mã hóa video và lớp vận chuyển gói tin mã hóa, trong khi người nhận có trách nhiệm giải mã gói tin tầng vận chuyển và giải mã video. Các gói dữ liệu được mã hóa trước khi chuyển lên trên đường truyền Internet. Bên nhận cũng gửi thông tin phản hồi để nhận thông qua các gói RTCP. Các phản hồi về chất lượng dịch vụ được sử dụng bởi bộ mã hóa video để điều chỉnh các thông số mã hóa của nó on-the-fly. Bằng cách này, bộ mã hóa video thích nghi với các điều kiện kênh tức thời và chất lượng video tốt hơn dự kiến ở phía người nhận [4].



Hình 3. Kiến trúc hệ thống MCN Streaming

Razib Iqbal và Shervin Shirmohammadi đã phân tích và đánh giá ưu nhược điểm các hướng tiếp cận để mô hình hóa Video Streaming tương thích trong [5]. Tiếp đó, Razib Iqbal đề xuất kiến trúc để xử lý video liên và streaming trực tuyến trong mạng ngang hàng sử dụng mô tả cú pháp luồng bit biểu diễn theo chuẩn MPEG-21 [6]. Trong [7], Giuseppe

Cofano và cộng sự đề xuất kiến trúc điều khiển cung cấp các luồng dữ liệu lớn (MAVSD- *Massive Adaptive Video Streaming Delivery*).

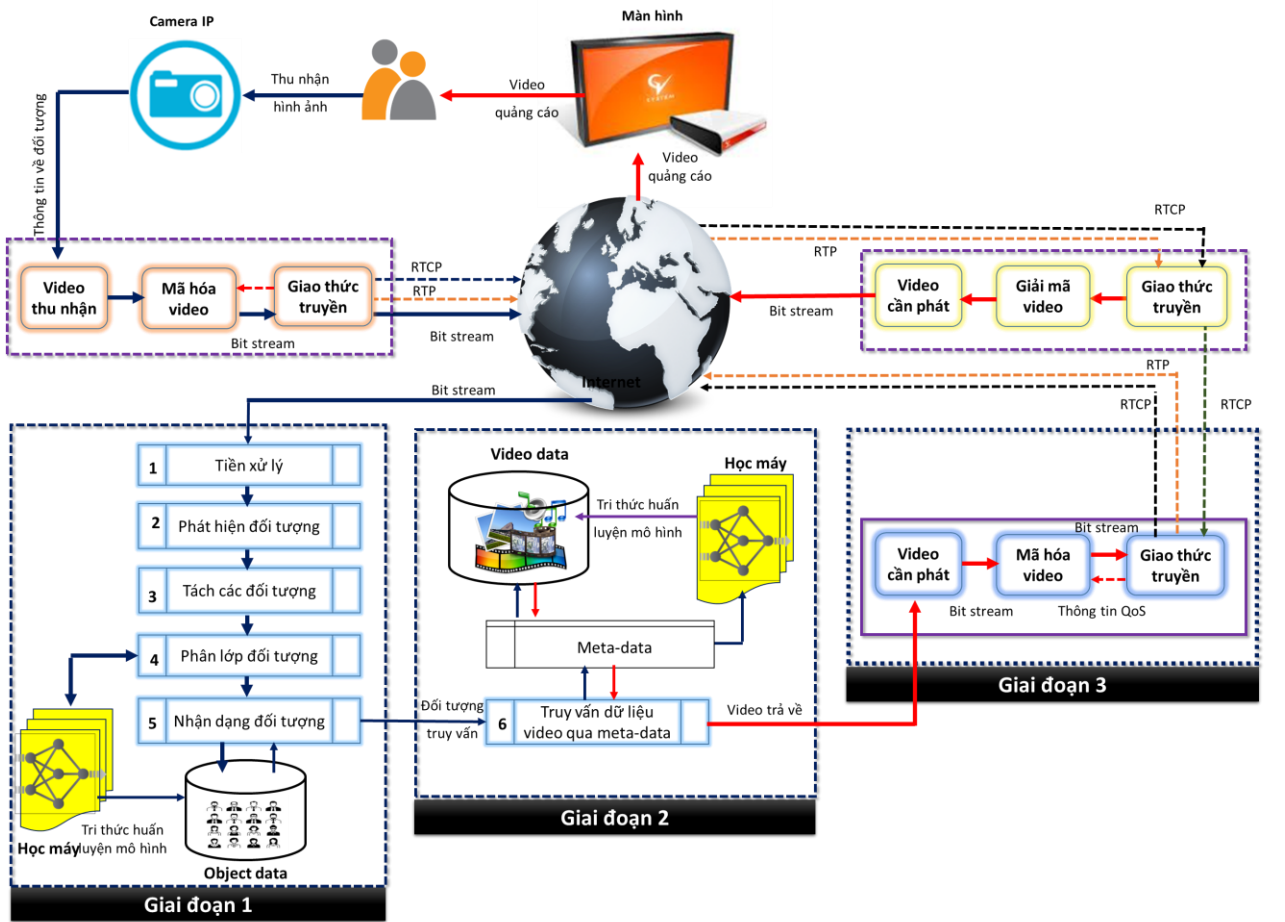
III. ĐỀ XUẤT MÔ HÌNH HỆ THỐNG QUẢNG CÁO TỰ ĐỘNG TÙY BIẾN THEO KHÁCH HÀNG TRONG NGỮ CẢNH THỜI GIAN THỰC

A. MÔ HÌNH HỆ THỐNG QUẢNG CÁO TỰ ĐỘNG TÙY BIẾN THEO KHÁCH HÀNG

Trong phần này, chúng tôi sẽ tập trung phân tích và đề xuất mô hình hệ thống quảng cáo tự động tùy biến theo đối tượng khách hàng trong ngữ cảnh thời gian thực được áp dụng trong tùy biến các nội dung video quảng cáo với mục tiêu hướng đến là làm cho nội dung quảng cáo phù hợp và thực sự hữu ích đối với các khách hàng trong từng ngữ cảnh cụ thể thông qua nhận dạng phân tích đối tượng thu nhận được từ camera. Hệ thống là sự kết hợp giữa xử lý ảnh và nhận dạng với kiến trúc MCN Streaming sử dụng cấu trúc metadata để lưu trữ dữ liệu video cộng với mô hình học máy nhằm mang lại hiệu quả tối đa cho hệ thống.

Mô hình hệ thống quảng cáo tự động tùy biến theo khách hàng trong ngữ cảnh thời gian thực được biểu diễn như Hình 4. Hệ thống gồm 3 giai đoạn chính là:

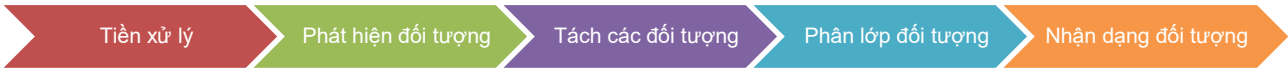
- Giai đoạn 1: Phát hiện và phân lớp đối tượng dựa trên hình ảnh thu nhập được từ Camera
- Giai đoạn 2: Truy xuất cơ sở dữ liệu video theo đối tượng phân lớp
- Giai đoạn 3: Truyền nội dung video



Hình 4. Kiến trúc hệ thống quảng cáo tự động tùy chỉnh theo khách hàng trong ngữ cảnh thời gian thực

B. GIAI ĐOẠN 1. PHÁT HIỆN VÀ PHÂN LỚP ĐỐI TƯỢNG DỰA TRÊN HÌNH ẢNH TỪ CAMERA

Giai đoạn đầu tiên của hệ thống là thu nhận hình ảnh đối tượng trực tiếp từ Camera, sau đó tiến hành phân tích để phát hiện và tách các đối tượng ra khỏi nền để trích chọn các đặc trưng về đối tượng được lựa chọn. Các đặc trưng về đối tượng sẽ được phân lớp rồi nhận dạng đối tượng.



Hình 5. Các bước trong giai đoạn phát hiện và phân lớp đối tượng dựa trên hình ảnh từ Camera



1. TIỀN XỬ LÝ

Trong quá trình thực hiện thu nhận ảnh từ video 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 ...). Video là do một chuỗi ảnh liên tiếp chuyển động tạo thành. Để có thể truyền tải video, các thông tin và dữ liệu này được mã hóa. Do đó, để có thể xử lý được các tín hiệu video bên cạnh việc xử lý các vấn đề khách quan như điều kiện ánh sáng, vị trí đặt camera...ta cũng cần quan tâm tới các vấn đề về tiêu chuẩn nén của video. Do đó các công việc như khử nhiễu, cân chỉnh mức xám thường được xác định thông qua các ngưỡng (*Threshhold*) trong chương trình và do người sử dụng quyết định (tinh chỉnh) và chưa có khả năng tự động cân bằng. Giai đoạn tiền xử lý này có chức năng chuẩn hóa ảnh cần tìm giúp cho việc tìm kiếm được hiệu quả hơn. Các công việc trong bước tiền xử lý có thể là: Chuẩn hóa kích cỡ giữa ảnh trong cơ sở dữ liệu và ảnh cần tìm, hiệu chỉnh độ sáng, tối của ảnh, lọc nhiễu, chuẩn hóa về vị trí, tư thế ảnh mặt...

Sau khi dữ liệu video được tiền xử lý loại bỏ những thành phần không mong muốn sẽ được chuyển sang giai đoạn phát hiện đối tượng. Mục tiêu hướng đến của chúng tôi là phát hiện đối tượng dựa vào khuôn mặt. Phần nhận dạng mặt người bao gồm hai bước: phát hiện khuôn mặt và nhận dạng tự động đối tượng. Công việc chính dựa vào các kỹ thuật rút trích đặc trưng từ ảnh đối tượng và thực hiện đối sánh để định danh tự động. Sau đây, chúng tôi sẽ phân tích hiệu quả của các phương pháp sử dụng trong hệ thống nhận dạng đối tượng.

2. PHÁT HIỆN ĐỐI TƯỢNG

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. Hệ thống sẽ phát hiện khoảng vùng khuôn mặt trong khung hình video có nhiều đối tượng xung quanh hay giữa đám đông nhiều người. Có rất nhiều hướng tiếp cận để giải quyết vấn đề trên, các tác giả trong [8] đã phân loại các hướng tiếp cận này như Bảng 1.

**Bảng 1.** Phân loại các kỹ thuật phát hiện đối tượng

Phân loại	Các kỹ thuật và 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 5. Optical flow
Supervised Classifier	1. Support Vector Machines 2. Neural Networks 3. Adaptive Boosting

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. Hướng giải quyết là xây dựng mô hình nền (*Background Modeling*), sau đó sử dụng mô hình này cùng với frame hiện tại để rút ra được các foreground chuyển động. Có nhiều phương pháp xây dựng mô hình background như: mô hình sử dụng adaptive kernel density estimation của Anurag Mittal 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. Giải thuật W4 của Haritaoglu, sử dụng Mixture of Gaussian của Stauffer 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? 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 [8]. 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.

3. TÁCH CÁC ĐỐI TƯỢNG

Bước này làm nhiệm vụ xác định vị trí, kích cỡ của một hoặc nhiều khuôn mặt trên ảnh chụp từ đó tách ra phần mặt. Về cơ bản hệ thống xử lý video gần giống với hệ thống xử lý ảnh. Tuy nhiên, do đặc trưng chuyển động của hình ảnh nên ta cần quan tâm tới các vấn đề:

- a) Xác định đối tượng, số lượng đối tượng có trong ảnh
- b) Xác định đối tượng chuyển động trong ảnh: các đối tượng dời đi hay chuyển tới, chuyển động đơn hay chuyển động trong đám đông
- c) Tách đối tượng từ một nhóm đối tượng chuyển động
- d) Tách đối tượng với bóng của bản thân

Phần ảnh mặt được tách ra thường nhỏ hơn nhiều so với ảnh chụp ban đầu, nó sẽ là các khuôn mặt cần tìm và chức năng trích chọn đặc trưng sẽ sử dụng các ảnh được tách ra này [9].

4. PHÂN LỚP ĐỐI TƯỢNG DỰA VÀO CÁC ĐẶC TRƯNG

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.

Các nghiên cứu liên quan đến trích chọn đặc trưng [10, 11, 12, 13]:

**Bảng 2.** Phân loại các kỹ thuật trích chọn đặc trưng

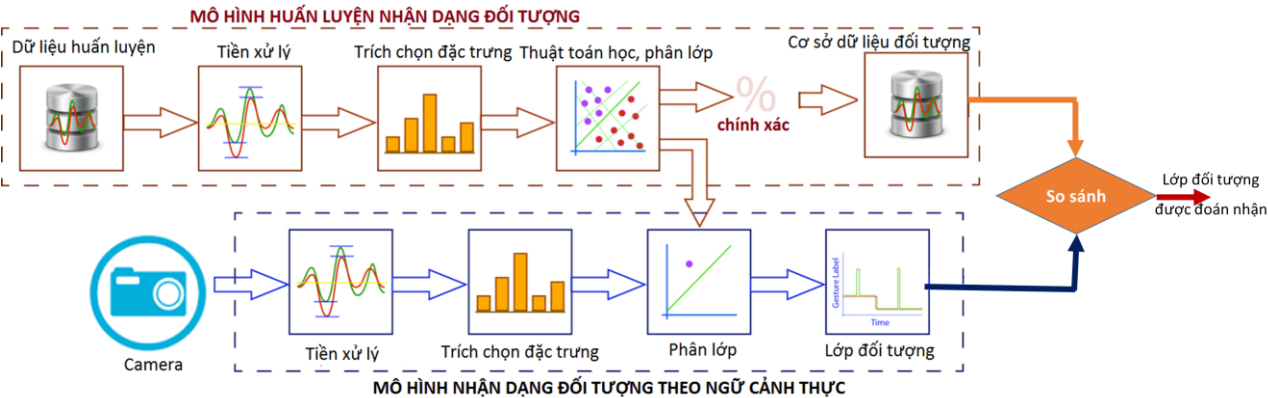
Tác giả	Các kỹ thuật và nghiên cứu liên quan
Bledsoe (1960) Goldstein (1970)	Sử dụng tiếp cận dựa trên các đặc trưng như mắt, tai, màu tóc, độ dày môi để tự động nhận dạng
Kirby và Sirovich (1988) Turk và Pentland (1991)	Sử dụng phương pháp phân tích thành phần chính PCA ( <i>Principal Components Analysis</i> ) và thuật toán Eigenfaces để nhận dạng khuôn mặt
Trần Phước Long, Nguyễn Văn Lượng (2003)	Sử dụng mạng nơron để dò tìm khuôn mặt trong ảnh, kết hợp với phương pháp PCA và biến đổi cosine rời rạc (DCT) để rút ra các đặc trưng là đầu vào cho bộ nhận dạng máy học SVM và mô hình Markov ẩn HMM
Lu Boun Vinh, Hoàng Phương Anh (2004)	Sử dụng thuật toán AdaBoost để dò tìm khuôn mặt kết hợp với thuật toán FSVM để tiến hành nhận dạng mặt người.
Zuo và cộng sự (2006)	Kết hợp phương pháp PCA hai chiều với phân tích độc lập tuyến tính (LDA)
Chen và Yaou Zhao (2006)	Kết hợp DCT và HRBF ( <i>Hierarchical Radial Basis Function Network</i> )
Sharif, Sajjad Mohsin, Muhammad Younas Javed, và Muhammad Atif Ali (2012)	Kết hợp thuật toán LoG ( <i>Laplacian of Gaussian</i> ) và DCT
Suhas, Kurhe Ajay B., và Dr. Prakash Khanale B (2012)	Áp dụng PCA kết hợp với DCT theo hướng tiếp cận toàn diện cho khuôn mặt
Sunil (2013)	Kết hợp PCA với DCT
Bouzalmat, Jamal Kharroubi, và Arsalane Zarghili (2013)	Áp dụng thuật toán máy học SVM dựa trên LDA, để nhận dạng khuôn mặt người trên ảnh
Lowe (2004)	Tiếp cận dựa trên các đặc trưng cục bộ không đổi SIFT. Đặc trưng cục bộ SIFT không bị thay đổi trước những biến đổi tỷ lệ ảnh, tịnh tiến, phép quay, không bị thay đổi một phần đối với phép biến đổi hình học affine (thay đổi góc nhìn) và mạnh với những thay đổi về độ sáng, sự nhiễu và che khuất.
Aly (2006)	Sử dụng các đặc trưng SIFT để nhận dạng mặt người
Chennamma, Lalitha Rangarajan, Veerabhadrappe (2011)	Kết hợp phương pháp phân tích thành phần chính và đặc trưng SIFT để nhận dạng khuôn mặt người
Kumar và Padmavati (2012)	Đề xuất các cách tính khoảng cách khác nhau khi so khớp các đặc trưng SIFT để nhận dạng mặt người, góp phần cải thiện tốc độ so khớp các đặc trưng SIFT.

5. NHẬN DẠNG ĐỐI TƯỢNG

Đặc tính sinh trắc của người là đặc tính đo được các nét hành vi riêng của con người. Nhận diện người được thực hiện trong quá trình kiểm tra tính đồng nhất của đặc tính sinh trắc giữa người cần kiểm tra với đặc tính tương tự của người đã được lưu sẵn trong cơ sở dữ liệu đối tượng. Chúng ta có thể nhận diện người dựa trên các đặc điểm sinh trắc tính trên người như khuôn mặt, mắt, vân tay, bàn tay, gen...hay các đặc trưng hành vi như dáng đi, chữ viết, giọng nói...Tuy nhiên, trong phạm vi nghiên cứu của hệ thống chúng tôi hướng đến việc nhận dạng đối tượng dựa trên khuôn mặt để phân lớp đối tượng đang tương tác.

Nhiệm vụ của giai đoạn nhận dạng đối tượng mặt người là xử lý tự động thông tin đặc trưng được trích chọn từ các ảnh để tìm ra độ tương tự giữa các khuôn mặt và đưa ra quyết định về tính đồng nhất giữa chúng. Quá trình này được thực hiện bằng cách so sánh giữa đối tượng thu nhận được với các đối tượng được lưu trữ trong cơ sở dữ liệu. Quá trình nhận dạng được tích hợp thêm khả năng tự học để huấn luyện hệ thống [14, 15]. Các bước thực hiện bao gồm:

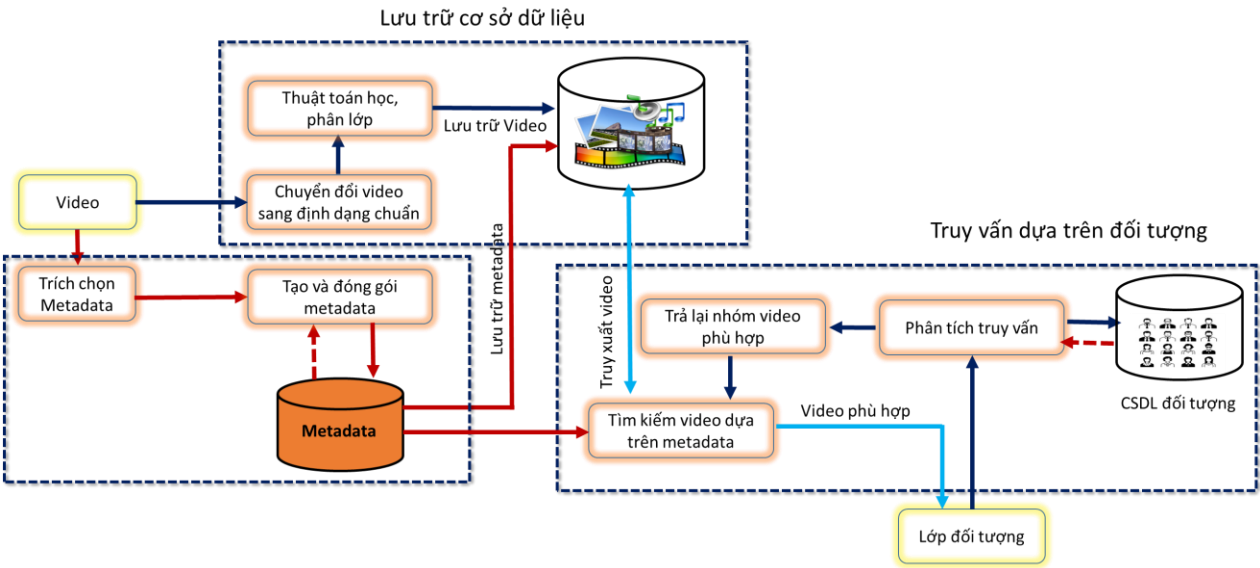
1. Chuẩn bị tập ảnh huấn luyện chứa các khuôn mặt người. Với ảnh mỗi người đặt vào trong thư mục tên của người đó, như vậy tiện cho việc viết code và phân lớp.
  2. Tiến hành tìm khuôn mặt người trong ảnh, nếu có thì tách ra file ảnh.
  3. Huấn luyện: vì nhãn là kiểu số nguyên nên chúng ta phải quy ước là tên người ứng với 1 số nguyên nào đó. Khi dự báo chương trình sẽ trả về nhãn và chuyển sang tên người.
  4. Dự đoán: với 1 ảnh cần phải tách khuôn mặt, sau đó dự đoán khuôn mặt đó.
- Có 3 thuật toán hiệu quả và chính xác thường được sử dụng phổ biến đó là: Eigen, Fisher và LBPH.



Hình 6. Mô hình nhận dạng đối tượng dựa trên hình ảnh từ Camera

**C. GIAI ĐOẠN 2. TRUY XUẤT CƠ SỞ DỮ LIỆU VIDEO THEO ĐỐI TƯỢNG PHÂN LỚP**

Để tăng cường hiệu quả trong truy xuất dữ liệu video dựa trên đối tượng nhận được từ giai đoạn 1, chúng tôi tổ chức cơ sở dữ liệu video theo dạng metadata. Các file video sẽ được trích chọn các đặc trưng metadata để thuận lợi trong quá trình truy xuất tìm kiếm dữ liệu tương tự định dạng chuẩn XML. Quá trình tạo và đóng gói các metadata từ file video được mô tả như sau:



Hình 7. Mô hình truy xuất dữ liệu video dựa trên metadata

**D. GIAI ĐOẠN 3. TRUYỀN NỘI DUNG VIDEO STREAMING QUA MẠNG**

Live streaming, truyền trực tiếp qua Internet, bao gồm một camera ghi lại nội dung cần phát, bộ mã hóa để số hóa nội dung, một server streaming để truyền dữ liệu, và một mạng lưới phân phối nội dung để phân phối và cung cấp nội dung. Các Client nhận được dữ liệu phải có khả năng tập hợp dữ liệu và gửi nó như một dòng ổn định cho các ứng dụng, sau đó là xử lý các dữ liệu và chuyển đổi nó thành âm thanh hoặc hình ảnh [17, 20]. Điều này có nghĩa rằng nếu các Streaming Client nhận được các dữ liệu nhanh hơn so với yêu cầu, nó cần phải lưu dữ liệu dư thừa trong một bộ đệm. Nếu luồng dữ liệu không đủ nhanh tải về, dữ liệu sẽ không được mịn màng khi hiển thị.

**1. ĐỊNH DẠNG DỮ LIỆU**

Âm thanh stream được nén bằng cách sử dụng định dạng âm thanh như MP3, Vorbis hoặc AAC. Hình ảnh video stream được nén bằng cách sử dụng codec video như H.264 hoặc VP8. Mã hóa âm thanh và video stream được những

trong một gói bitstream như FLV, WebM, ASF hoặc ISMA. Bitstream được phân phối từ một streaming server tới một streaming client bằng cách sử dụng một giao thức truyền tải, ví dụ như là MMS hoặc RTP. Các streaming client có thể tương tác với streaming server bằng cách sử dụng một giao thức kiểm soát, chẳng hạn như MMS hoặc RTSP [16, 19].

## 2. CÁC GIAO THỨC TRUYỀN TẢI

Để thiết kế một giao thức mạng hỗ trợ streaming media đặt ra nhiều vấn đề. Một số giao thức hiện nay đang được sử dụng để hỗ trợ streaming media gồm:

Giao thức Datagram: chẳng hạn như giao thức UDP (*User Datagram Protocol*) gửi media stream như là một loạt các gói tin nhỏ. Điều này dường như là đơn giản và hiệu quả, tuy nhiên, không có cơ chế nào trong giao thức để đảm bảo việc giao tập tin media đi. Nếu dữ liệu bị mất, việc stream có thể khiến cho chất lượng bị sụt giảm. Giao thức TCP (*Transmission Control Protocol*) sẽ đảm bảo chuyển đúng mỗi bit trong media stream. Khi xảy ra hiện tượng mất dữ liệu trên mạng, các gói media stream, Trong khi giao thức xử lý, phát hiện những dữ liệu bị mất và truyền lại, thì Clients có thể giảm thiểu hiệu ứng này bằng cách sử dụng dữ liệu đệm để hiển thị. Trong khi việc trễ dữ liệu này là chấp nhận được trong đoạn video về các kịch bản yêu cầu, người sử dụng các ứng dụng tương tác như hội nghị truyền hình sẽ cảm nhận được nếu sự chậm trễ vượt quá 200ms. Giao thức Unicast gửi một bản sao riêng media stream từ máy chủ đến mỗi người nhận. Unicast là tiêu chuẩn cho hầu hết các kết nối Internet, nhưng quy mô không lớn khi mà nhiều người muốn xem cùng một chương trình truyền hình tại một thời điểm. Giao thức Multicasting cũng tương tự như vậy khi mà copy cùng một bản sao trên toàn bộ mạng cho một nhóm khách hàng. Giao thức Multicast được phát triển để giảm tải máy chủ / network từ những luồng data trùng lặp xảy ra khi nhiều người nhận được nhiều luồng độc lập. Các giao thức này gửi một luồng duy nhất từ nguồn đến một nhóm người nhận. Tùy thuộc vào loại hình và cơ sở hạ tầng mạng, Multicast có thể hoặc không thể có thực thi. Thêm một sự bất lợi của multicasting đó là sự mất mát dữ liệu khi sử dụng tính năng Video theo yêu cầu. Tính streaming liên tục của luồng video – âm thanh của Radio/TV thường ngăn cản khả năng điều khiển việc phát lại của người nhận. Tuy nhiên, vấn đề này có thể được giảm nhẹ bởi các yếu tố chẳng hạn như bộ nhớ đệm của máy chủ, hộp set-top kỹ thuật số, và bộ đệm của thiết bị phát.

Giao thức Multicast IP cung cấp một phương tiện để gửi một luồng media duy nhất cho một nhóm người nhận trên một mạng máy tính. Một giao thức multicast, thường là Internet Group Management Protocol, được sử dụng để quản lý việc cung cấp các dòng multicast cho các nhóm người nhận trên một mạng LAN. Một trong những thách thức trong việc triển khai IP multicast là các bộ định tuyến và tường lửa giữa mạng LAN phải cho phép thông qua để các gói tin đến nhóm multicast. Giao thức P2P (*Peer-to-peer*) sắp xếp để mỗi máy tính đều là máy chủ, có thể streaming luồng dữ liệu. Hoạt động của giao thức chủ yếu dựa vào khả năng tính toán và băng thông của các máy tham gia chứ không tập trung vào một số nhỏ các máy chủ trung tâm như các mạng thông thường, tất cả các máy tham gia đều đóng vai trò đồng thời là máy khách và máy chủ đối với các máy khác trong mạng. Các máy tham gia đều đóng góp tài nguyên, bao gồm băng thông, lưu trữ, và khả năng tính toán.. Tuy nhiên, Một nhược điểm khác của hệ thống này là do không có định hướng, một yêu cầu tìm kiếm thường được chuyển cho một số lượng lớn máy trong mạng làm tiêu tốn một lượng lớn băng thông của mạng, nó làm tăng các vấn đề kỹ thuật, hiệu suất, chất lượng.

Trong mô hình đề xuất, chúng tôi sử dụng các giao thức RTSP (*Real-time Streaming Protocol*), RTP (*Real-time Transport Protocol*) và RTCP (*Real-time Transport Control Protocol*) được thiết kế đặc biệt để stream media qua mạng [16, 17]. RTSP là giao thức điều khiển trên mạng được thiết kế để sử dụng giao tiếp giữa máy client và máy streaming server. Giao thức này được sử dụng để thiết lập và điều khiển phiên giao dịch giữa các máy tính. RTSP định nghĩa một bộ các tín hiệu điều khiển tuần tự, phục vụ cho việc điều khiển quá trình playback. Trong khi giao thức HTTP là giao thức không có trạng thái thì RTSP là giao thức có xác định trạng thái. Một định danh được sử dụng khi cần thiết để theo dõi các phiên giao dịch hiện tại của quá trình streaming video gọi là số hiệu session. Cũng giống như HTTP, RTSP sử dụng TCP là giao thức để duy trì một kết nối đầu cuối tới đầu cuối và các thông điệp điều khiển của RTSP được gửi bởi máy client tới máy server. Nó cũng thực hiện điều khiển lại các đáp trả từ máy server tới máy client trên cổng mặc định 554. Để thực hiện kỹ thuật streaming video theo giao thức RTSP máy client phải gửi lên máy server (streaming server) những request sau và phải theo một trình tự nhất định: gửi yêu cầu OPTIONS, DESCRIBE, SETUP, PLAY. Trong quá trình streaming video, nếu như người dùng rời khỏi phạm vi quan sát của camera hệ thống sẽ tạm dừng quá trình streaming thì sẽ gửi yêu cầu PAUSE tới máy server, yêu cầu này sẽ làm tạm dừng một hay nhiều luồng dữ liệu đang truyền các frame về máy client. RTP định dạng một gói tin RTP được dùng để truyền trên luồng dữ liệu video hay audio dựa trên địa chỉ IP. RTP được sử dụng kết hợp với giao thức RTCP. Trong đó, RTP được sử dụng để đóng gói các frame dữ liệu ( audio và video) để truyền trên luồng dữ liệu thì RTCP được sử dụng để giám sát chất lượng của dịch vụ (QoS) hoặc để thống kê theo các tiêu chí trong quá trình truyền tải. Thường thì giao thức RTP sử dụng cổng có số hiệu chẵn còn giao thức RTCP sử dụng cổng có số hiệu lẻ. Giao thức truyền RTP quy định cách thức truyền dữ liệu theo thời gian thực. Thông tin được cung cấp bởi giao thức này bao gồm thời gian đồng bộ (timestamps), số thứ tự gói tin (phục vụ cho việc tìm gói tin bị lạc ) và chi phí cho việc mã hóa định dạng dữ liệu. Giao thức điều khiển RTCP để kiểm tra QoS luồng dữ liệu và thực hiện đồng bộ giữa các luồng dữ liệu. So với RTP, thì băng thông của RTCP sẽ nhỏ hơn [18, 19]. Một giao thức cho phép miêu tả dữ liệu đa phương tiện nhưng không bắt buộc phải kèm theo là giao thức miêu tả phiên (Session Description Protocol – SDP).



#### IV. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, chúng tôi đã trình bày đề xuất mô hình hệ thống quảng cáo tự động tùy biến theo đối tượng khách hàng. Từ những hình ảnh thực thu nhận được từ camera, hệ thống sẽ phân tích các đối tượng dựa trên tập đặc trưng cho trước để xác định lớp khách hàng phù hợp. Dựa trên lớp khách hàng được xác định, hệ thống sẽ truy xuất đến cơ sở dữ liệu quảng cáo đa phương tiện và tự động lựa chọn, phát những nội dung phù hợp. Với mỗi giai đoạn, chúng tôi đã tập trung phân tích và đánh giá các kỹ thuật được sử dụng để nâng cao tính khả thi và hiệu quả của hệ thống.

Trong thời gian tới, chúng tôi sẽ triển khai và tiến hành phân tích đánh giá hiệu quả của hệ thống thực với nhiều kỹ thuật khác nhau trong mỗi giai đoạn. Các nghiên cứu sẽ hướng đến xây dựng một hệ thống hoàn chỉnh hỗ trợ phân tích chính xác đối tượng và tốc độ truyền video qua mạng nhanh chóng.

#### V. TÀI LIỆU THAM KHẢO

- [1] Yun Ji Na, Choon Seong Leem, and Il Seok Ko. 2006. ACASH: an adaptive web caching method based on the heterogeneity of web object and reference characteristics. *Inf. Sci.* 176, 12, 1695-1711.
- [2] Dapeng Wu, Yiwei Thomas Hou et. al. Streaming Video over the Internet: Approaches and Directions.
- [3] Dapeng Oliver Wu, Qin Chen, MCN Streaming: An Adaptive Video Streaming Platform, Multimedia Communications and Networking (MCN) Lab Dept. of Electrical & Computer Engineering, University of Florida,
- [4] Hayder Radha, Mihaela van der Schaar, and Shirish Karande. 2004. Scalable video transcaling for the wireless internet. *EURASIP J. Appl. Signal Process.* 2004, 265-279.
- [5] Razib Iqbal and Shervin Shirmohammadi. 2010. An analytical approach to model adaptive video streaming and delivery. In *Proceedings of the 2010 ACM workshop on Advanced video streaming techniques for peer-to-peer networks and social networking (AVSTP2P '10)*. ACM, New York, NY, USA, 55-58.
- [6] Giuseppe Cofano, Luca De Cicco, and Saverio Mascolo. 2014. A Control Architecture for Massive Adaptive Video Streaming Delivery. In *Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming (VideoNext '14)*. ACM, New York, NY, USA, 7-12.
- [7] 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.
- [8] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, 'Face Recognition Using Laplacianfaces,' IEEE Trans. Pattern Analysis and Machine Intelligence, no. 3, pp. 328-340, 2005.
- [9] 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
- [10] E. Rentzeperis, A. Stergiou, A. Pnevmatikakis and L. Polymenakos, 'Impact of Face Registration Errors on Recognition,' in I. Maglogiannis, K. Karpouzis and M. Bramer (eds.), Artificial Intelligence Applications and Innovations (AIAI06), Springer, Berlin Heidelberg, pp. 187-194, June 2006.
- [11] A. Pnevmatikakis and L. Polymenakos, 'Subclass Linear Discriminant Analysis for Video-Based Face Recognition,' J. of Visual Communication and Image Representation, vol. 20, no. 8, pp. 543-551, 2009.
- [12] 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.
- [13] Anurag Mittal and Mikos Paragios, "Motion Based Background Subtraction using Adaptive Kernel Density Estimation" pp. 302-309, 2004
- [14] C. Stauffer and W. Grimson, "Adaptive Background mixture models for Real-time tracking" pp. 750-755, 2009.
- [15] Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object Tracking: A Survey" pp. 7-15.
- [16] Rodriguez, D., Rosa, R.; Costa, E.; Abrahao, J.; Bressan, G., Video quality assessment in video streaming services considering user preference for video content, IEEE Transactions on Consumer Electronics, Vol.60(3), 436 - 444
- [17] Björn Richerzhagen, Stefan Wilk, Julius Rückert, Denny Stohr, and Wolfgang Effelsberg. 2014. Transitions in Live Video Streaming Services. In *Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming (VideoNext '14)*. ACM, New York, NY, USA, 37-38.
- [18] Carlo Giuliotti, Dan Schonfeld, and Rashid Ansari. 2010. A novel cache optimization algorithm and protocol for video streaming in pure peer-to-peer networks. In *Proceedings of the 2010 ACM workshop on Advanced video streaming techniques for peer-to-peer networks and social networking (AVSTP2P '10)*. ACM, New York, NY, USA, 19-24.
- [19] Xiangyang Zhang and Hossam Hassanein. 2012. A survey of peer-to-peer live video streaming schemes - An algorithmic perspective. *Comput. Netw.* 56, 15 (October 2012), 3548-3579.
- [20] Luca De Cicco and Saverio Mascolo. 2014. An adaptive video streaming control system: modeling, validation, and performance evaluation. *IEEE/ACM Trans. Netw.* 22, 2 (April 2014), 526-539.