

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



NGUYỄN THANH TÙNG

**NGHIÊN CỨU THUẬT TOÁN PHÁT HIỆN ĐIỂM
CẮT, GHÉP TRONG VIDEO**

Chuyên ngành: Hệ thống thông tin

Mã số : 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ

HÀ NỘI - NĂM 2021

Luận văn được hoàn thành tại:

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN
THÔNG**

Người hướng dẫn khoa học: PGS.TS. HÀ HẢI NAM
(Ghi rõ học hàm, học vị)

Phản biện 1: PGS.TS. NGÔ QUỐC TẠO

Phản biện 2: TS. HOÀNG XUÂN DẬU

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn
thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: 10 giờ 00 ngày 28 tháng 8 năm 2021

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn
thông.

MỞ ĐẦU

1. Lý do chọn đề tài

Sự ra đời và ứng dụng các phương tiện ghi âm, ghi hình trong công tác điều tra, phá án cũng như tố tụng ngày càng được triển khai sâu rộng, phổ biến trên thế giới; dữ liệu hình ảnh, video thu được từ các hiện trường vụ án đã trở thành một nguồn chứng cứ quan trọng, giúp cơ quan chức năng củng cố chứng cứ, chứng minh các hoạt động phạm tội. Tuy nhiên, bên cạnh những thuận lợi do sự phát triển của khoa học kỹ thuật hiện đại đem lại đó, nó cũng kéo theo nhiều ảnh hưởng tiêu cực trong đời sống, như: việc các video/hình ảnh giả mạo, chứa thông tin sai sự thật (*Deep-fakes*), các video/hình ảnh hiện trường bị chỉnh sửa, cắt ghép, bị các đối tượng phạm tội tác động làm sai lệch thông tin ngày càng phổ biến; thông tin sai sự thật lan tràn ngày càng nhiều trên Internet.

Tại Việt Nam, công tác giám định hình ảnh cũng được Viện Khoa học hình sự - Bộ Công an nghiên cứu, triển khai đạt được nhiều kết quả tích cực; tuy nhiên, do số lượng vụ án hàng năm ngày càng tăng, dữ liệu video thu được từ hiện trường các vụ án ngày càng lớn đã làm tăng cao nhu cầu phát hiện video giả mạo, bị chỉnh sửa. Đáng chú ý, hiện nay công tác giám định video giả mạo cắt ghép chủ yếu được thực hiện hoàn toàn thủ công dựa trên quan sát trực tiếp video của các chuyên gia. Công việc này tốn rất nhiều thời gian và công sức đặc biệt khi các đoạn video thu từ camera có thời lượng lớn. Do đó, việc tự động hoá phát hiện video bị cắt ghép là nhu cầu cấp bách trong công tác điều tra, phá án. Nếu ứng dụng thành công các công nghệ, kỹ thuật hiện đại, hệ thống phát hiện video bị cắt ghép, giả mạo sẽ giúp giảm công sức của các chuyên gia và tăng hiệu quả xử lý công tác giám định kỹ thuật hình sự.

Với yêu cầu thực tiễn nêu trên, việc nghiên cứu đề tài "*Nghiên cứu thuật toán phát hiện điểm cắt, ghép trong video*" là cần thiết cả về phương diện lý luận và thực tiễn.

2. Tổng quan về vấn đề nghiên cứu

Video đã trở thành một phần không thể thiếu trong xã hội hiện đại; tuy nhiên, việc chỉnh sửa video ngày càng trở nên dễ dàng hơn, rất dễ dàng để một số người dùng tạo video được chỉnh sửa với ý đồ xấu. Trong khi nhiều nghiên cứu được nhắm mục tiêu vào những kỹ thuật tiên tiến như phát hiện *Deepfakes*, các kỹ thuật cũ, đơn giản hơn lại không được kiểm tra, không có phương tiện phát hiện. Các thao tác chỉnh sửa video như cắt xén, nối và điều chỉnh tốc độ vẫn có thể dẫn đến các cuộc tấn công hiệu quả. Trong đề tài này, học viên sẽ nghiên cứu đánh giá một số cách tiếp cận phát hiện video bị chỉnh sửa, như: phát hiện dựa trên đặc trưng điểm ảnh, phát hiện dựa trên đặc trưng luồng video và phát hiện dựa trên đặc trưng audio của luồng đa phương tiện; trong đó, so sánh hiệu năng và độ chính xác của từng cách tiếp cận làm cơ sở cho việc khuyến nghị sử dụng các kỹ thuật khác nhau cho từng trường hợp sử dụng cụ thể.

3. Mục đích nghiên cứu

- Rèn luyện phương pháp và khả năng nghiên cứu.
- Nghiên cứu đặc trưng video cắt, ghép.
- Nghiên cứu một số thuật toán phân tích và xử lý hình ảnh.
- Ứng dụng trong một bài toán cụ thể.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng và phạm vi nghiên cứu của luận văn bao gồm:

- Bài toán phát hiện điểm cắt, ghép trong video.
- Các thuật toán, phương pháp phân tích và xử lý hình ảnh.

5. Phương pháp nghiên cứu

- Phương pháp lý thuyết: Khảo sát, phân tích các tài liệu khoa học liên quan đến các thuật toán và bài toán phát hiện điểm cắt, ghép trong video.

- Phương pháp thực nghiệm: Sử dụng các công cụ, phần mềm để thử nghiệm và đánh giá hiệu quả của các thuật toán đề xuất.

Chương 1 - TỔNG QUAN VỀ BÀI TOÁN PHÁT HIỆN ĐIỂM CẮT, GHÉP TRONG VIDEO

1.1. Đặt vấn đề bài toán

Ngày nay, sự phát triển nhanh chóng của mạng Internet kèm theo khối lượng dữ liệu khổng lồ, đa dạng và tăng trưởng không ngừng. Tính đến tháng 01/2021, thế giới có khoảng 4,66 tỷ người dùng Internet, chiếm 59,5% dân số thế giới. Ước tính cứ mỗi ngày có hơn 2 Exabyte (10^{18} byte) dữ liệu được tạo ra trên Internet, mỗi phút có 4,2 triệu câu lệnh tìm kiếm Google; trên Facebook, có thêm 400 người dùng mới, hơn 200.000 bức ảnh được tải lên; trên Youtube, **72 tiếng** video được tải lên, **4.7 triệu video** được xem... [23].

Đối với dữ liệu trên mạng internet, chủ yếu là nội dung do người dùng tạo (UGC), trong đó, dữ liệu là video được quay bằng thiết bị cầm tay, thiết bị điều khiển từ xa, như: điện thoại thông minh, camera, flycam... của người dùng ngày càng chiếm khối lượng lớn. Mọi người có thể chỉnh sửa video cho nhiều mục đích khác nhau, kể cả ủng hộ vấn đề chính trị hoặc giải trí, nhưng những video giả mạo như vậy đặt ra một thách thức lớn cho các tổ chức tin tức, vì việc đăng tải các video giả mạo có thể gây tổn hại nghiêm trọng đến danh tiếng, quyền lợi, sức khỏe của các tổ chức, cá nhân và thậm chí là cả xã hội. Điều này tạo ra nhu cầu cấp thiết về các công cụ có thể hỗ trợ các chuyên gia xác định và tránh nội dung video bị giả mạo.

Việc phát hiện thao tác chỉnh sửa, phân biệt dấu vết chỉnh sửa, cắt ghép so với hình ảnh gốc ngày càng trở nên khó khăn khi các phương pháp giả mạo hình ảnh tinh vi mới được xuất hiện và phổ biến. Vì các công cụ giả mạo ngày càng thông minh, nên một hệ thống phát hiện giả mạo kỹ thuật số đáng tin cậy đang ngày càng trở nên quan trọng trong các lĩnh vực an ninh công cộng, cũng như đối với các lĩnh vực khác, như: điều tra tội phạm, pháp y, dịch vụ tình báo, bảo hiểm, báo chí, nghiên cứu khoa học, hình ảnh y tế và giám sát... Tuy nhiên, các hành vi chỉnh sửa hình ảnh không phải lúc nào

cũng độc hại đối với việc giám định video. Bên cạnh những trường hợp có thể xảy ra như chèn hoặc xóa người, đồ vật quan trọng, có thể làm thay đổi nội dung của video và đây là những trường hợp mà giám định video đề tài chủ yếu nhắm đến, còn có rất nhiều kiểu giả mạo khác có thể diễn ra trên video nhưng không ảnh hưởng lớn tới tính chính xác của chứng cứ. Chúng có thể bao gồm các hoạt động như điều chỉnh độ sắc nét hoặc màu sắc vì lý do thẩm mỹ cho toàn bộ video hoặc việc bổ sung các biểu tượng và hình mờ trên video. Do đó có thể dẫn đến các kết luận, đánh giá không chính xác của thuật toán hệ thống hoặc cũng có thể là một trong những yếu tố che đi các dấu vết của bộ chỉnh sửa độc hại khác.

Việc phát hiện các thao tác chỉnh sửa trong video là một nhiệm vụ đầy thách thức vì các thao tác giả mạo để lại dấu vết trên video - thường không thể nhìn thấy bằng mắt thường và liên quan đến một số thuộc tính của nhiều ảnh cơ bản hoặc các mẫu nén của video và dấu vết đó chỉ có thể được phát hiện bằng các thuật toán thích hợp nhưng hiện nay vẫn còn tồn tại nhiều phức tạp trong cách tiếp cận này. Nhìn chung, có nhiều kiểu hành vi chỉnh sửa khác nhau có thể diễn ra, như: xóa đối tượng, sao chép đối tượng từ cùng một cảnh hoặc từ một video khác, chèn nội dung tổng hợp, chèn hoặc xóa khung, chọn khung hoặc thay đổi màu sắc/độ sáng toàn cục... mỗi loại có khả năng để lại các loại dấu vết khác nhau trên video. Hơn nữa, một vấn đề khác của bài toán thực tế là việc nén video bao gồm một số quy trình khác nhau, tất cả đều có thể phá vỡ các dấu vết giả mạo.

Với những thách thức này, các nhà nghiên cứu đã và đang nghiên cứu xây dựng, triển khai nhiều hệ thống theo các hướng khác nhau nhằm hướng hỗ trợ các chuyên gia trong việc xác định các video giả mạo hoặc nâng cao hiện đại hóa lĩnh vực kỹ thuật hình sự. Các nghiên cứu trong giám định hình ảnh là tiền đề hết sức cần thiết cho mở rộng nghiên cứu các thuật toán hay "bộ lọc" nhằm xử lý video và giúp người dùng cụ thể hóa các điểm mâu thuẫn đáng ngờ trong video. Những bộ lọc này hướng tới khả năng đưa ra kết quả

được hiển thị cho người dùng, giúp họ xác minh video một cách trực quan. Đi kèm với đó, việc sử dụng kiến trúc mạng nơ-ron nhân tạo (deep neuron) để phát hiện những điểm không nhất quán trong video và phân loại video là "gốc" hoặc bị giả mạo vào xây dựng hệ thống tự động hóa quá trình phát hiện cũng là một nội dung tất yếu của công nghệ tự động hóa, một bước tiến của Trí tuệ nhân tạo (AI).

1.2. Một số nội dung cơ bản liên quan bài toán

Giám định hình ảnh và video về cơ bản là các lĩnh vực phụ của xử lý hình ảnh và video, do đó một số khái niệm từ các lĩnh vực xử lý hình ảnh/video đặc biệt quan trọng đối với nhiệm vụ của đề tài, như: hình ảnh (hoặc khung hình - frame); giới hạn nhiễu hình ảnh (*image noise*); sự nhạy bén (*acuity*) hoặc sắc nét (*sharpness*) của hình ảnh; vấn đề nén MPEG...

1.3. Nghiên cứu, ứng dụng hiện nay về phát hiện điểm cắt ghép trong video

Ngày nay, giám định đa phương tiện kỹ thuật số đã trở thành một lĩnh vực nghiên cứu mới nổi, nhận được sự chú ý đáng kể nhằm xác định nguồn gốc và tính xác thực của phương tiện kỹ thuật số. Tính xác thực của hình ảnh rất quan trọng trong nhiều lĩnh vực xã hội, chẳng hạn như: *trong lĩnh vực y tế*, các bác sĩ đưa ra các quyết định quan trọng dựa trên hình ảnh kỹ thuật số; *trong các cơ quan thực thi pháp luật và trong tố tụng hình sự*, tính chính xác của các bức ảnh có một vai trò thiết yếu để chúng có thể được sử dụng làm bằng chứng.

Một số nhà nghiên cứu khoa học đã xem xét tính xác thực của phương tiện truyền thông nhưng do khối lượng đa phương tiện khổng lồ và phức tạp cần phân tích khiến việc xây dựng thuật toán phát hiện giả mạo đa phương tiện trở nên khó khăn. Nghiên cứu trong lĩnh vực này chưa đưa ra được các giải pháp mạnh mẽ và phổ biến, đến nay vẫn cần nhiều những nghiên cứu, đóng góp sâu rộng hơn. Trong những năm gần đây, hầu hết các nỗ lực đã được dành cho việc phát hiện *giả mạo tĩnh*, việc phát hiện *giả mạo động* đã không

nhận được nhiều sự chú ý vì sự phức tạp của phân tích cảnh động và chi phí tính toán, vấn đề này trở nên khó khăn hơn với giám định video.

Một số công trình khoa học có liên quan đã được phát triển để phát hiện video giả mạo hoặc có khả năng phát hiện các đối tượng hoặc khung hình đáng ngờ dựa trên các đặc điểm của video kỹ thuật số. Một số phương pháp được triển khai tập trung vào việc xác định giả mạo giữa các khung hoặc nội khung. Các phương pháp dựa trên xem xét nội khung có thể thực hiện trong miền không gian hoặc không gian - thời gian (như sao chép - di chuyển hoặc nội khung). Các phương pháp dựa trên liên khung diễn ra trong miền thời gian (như chèn, loại bỏ và sao chép khung). Một trong những công trình tiên phong trong lĩnh vực này đã xử lý việc phát hiện trùng lặp khung [33], bằng cách tính đến thông tin tương quan giữa các khung liên tiếp, các manh mối hiệu quả cần được khai thác gồm: tốc độ và sự không nhất quán về mặt vật lý [5]; phần dư chuyển động [35]; và các tính năng đường bao thống kê [4].

Nhìn chung, video giả mạo có thể được phát hiện bằng cách xác minh các thay đổi về không gian, chẳng hạn như nén khung hình [14] [22] hoặc các phương thức tạm thời như thêm hoặc xóa khung [2] [12]. Trong số các kỹ thuật của giám định thụ động, nén kép là một trong những manh mối quan trọng để phát hiện giả mạo video. Một số nghiên cứu đã giải quyết vấn đề phát hiện nén kép như dựa trên việc sử dụng các đặc trưng không gian - thời gian được đánh giá trên cơ sở trường vector chuyển động cục bộ [15]. Một số nghiên cứu tập trung vào việc phát hiện sự trùng lặp khung, ví dụ, trong [34], các tác giả khai thác mối tương quan của các đặc điểm phân tách giá trị kỳ dị giữa các khung gốc và khung đáng ngờ, việc giả mạo sao chép khung được phát hiện bằng cách sử dụng phương pháp dựa trên phân tích tương tự. Ngoài ra, các đặc điểm dư chuyển động trong mỗi khung có thể được sử dụng để xác định các khung bị chỉnh sửa, giả mạo. Một kỹ thuật thụ động khác dựa trên việc trích xuất các đặc trưng thống kê và phân loại của các đặc điểm này thành các mẫu

dương tính hoặc mẫu âm tính [26]. Các công trình nghiên cứu khác đã giải quyết đồng thời nhiều loại tấn công khác nhau như xóa khung và chèn khung bằng cách sử dụng biểu đồ của các tính năng gradient có định hướng (HOG).

Việc khai thác các đặc điểm không gian-thời gian hiệu quả vẫn là thách thức chính đối với hầu hết các nhà nghiên cứu để xác định các khung hình sao chép với độ chính xác cao [27]. Ví dụ: phép phân tách giá trị số ít (SVD) được thực hiện cùng với phép đo độ tương tự Euclid trong [34]; độ lệch chuẩn của các khung hình dư được sử dụng để chọn một số khung hình từ chuỗi video và sau đó giá trị entropy của Biến đổi Cosine rời rạc (DCT) được khai thác để phát hiện sự trùng lặp giữa các khung hình [9]. Trong [28], các tác giả đã sử dụng DCT để tạo một tập hợp các tính năng cho mỗi khung và sau đó để phát hiện ra sự hiện diện của giả mạo bằng cách sử dụng hệ số tương quan. Phương pháp này cho kết quả tốt nhưng thời gian tính toán tương đối lớn.

Gần đây, một số kỹ thuật phát hiện giả mạo video tự động đã được triển khai, trong đó, có những cách tiếp cận tận dụng các mô hình thống kê được áp dụng thành công. Việc chuyển đổi tốc độ khung hình bù theo chuyển động cũng được khai thác cho các mục đích phát hiện giả mạo như làm giả tốc độ khung hình. Vấn đề này cũng được xử lý trong, trong đó tín hiệu dư được coi là dấu hiệu để xác định vị trí các khung giả mạo nội suy [6]. Thời điểm trên xung dao động wavelet và cường độ gradient trung bình cũng được ước tính cùng với khái niệm về ranh giới đối tượng có chiều rộng có thể điều chỉnh (AWOB) và phân loại SVM để xác định các mẫu dương tính (video gốc) và mẫu âm tính (video giả mạo) [26].

Có thể thấy, các nghiên cứu hiện nay trong lĩnh vực giám định video đã đạt được nhiều thành tựu lớn, kết quả khả quan. Tuy nhiên, còn tồn tại một số khó khăn như: hiệu quả khử nhiễu thấp, chưa hoạt động hiệu quả trên video chất lượng cao, khó để định vị tất

cả các khung hình nội suy và không thể khôi phục video đã bị chỉnh sửa, cắt ghép trong nhiều trường hợp.

Chương 2 - THUẬT TOÁN VÀ MÔ HÌNH HỆ THỐNG TỰ ĐỘNG PHÁT HIỆN ĐIỂM CẮT, GHÉP TRONG VIDEO

2.1. Các đặc trưng của video bị cắt ghép, giả mạo

Video là một tập hợp của các chuỗi khung hình/hình ảnh kết hợp với các kỹ thuật nén khác nhau, do đó, ở một mức độ nào đó các loại giả mạo video có thể có những thông tin sai lệch tương tự như các loại giả mạo trong hình ảnh, như: có thể gặp phải các thao tác sao chép chuyển động, ghép nối, nội khung hoặc chỉnh sửa toàn bộ video như thay đổi độ sáng hoặc độ nét. Tuy nhiên, một điểm khác biệt quan trọng trong giám định video là các thao tác giả mạo có thể tác động đến phương diện thời gian của video.

Ngoài ra, các thuật toán giám định hình ảnh dựa trên định dạng ảnh JPEG là không đủ để phát hiện hoặc xác định vị trí các điểm giả mạo trong video. Do một video không chỉ là một chuỗi hình ảnh; việc nén MPEG tái tạo lại hầu hết các khung hình bằng cách kết hợp các khối từ các khung hình khác với một hình ảnh dư. Quá trình này về cơ bản phá hủy các dấu vết mà các thuật toán dựa trên hình ảnh nhằm mục đích phát hiện. Hơn nữa, việc yêu cầu và giải nén được thực hiện bởi các nền tảng trực tuyến như YouTube, Facebook và Twitter gây khó khăn hơn nhiều đối với giám định các dấu vết giả mạo nhỏ, khó phát hiện so với các thuật toán giải nén tương ứng cho hình ảnh. Do đó, việc phát hiện giả mạo video đòi hỏi sự phát triển của các thuật toán cụ thể hướng mục tiêu đến đối tượng là các video. Có thể tìm thấy một số lượng lớn các phương pháp giám định tích cực, tuy nhiên, các phương pháp này không áp dụng được trong khá nhiều trường hợp, nơi chúng ta không kiểm soát được quá trình quay video; các phương pháp tiếp cận giám định video được đề xuất có thể được phân theo ba loại: phát hiện lượng tử hóa kép/nhiều, phát hiện giả mạo giữa các khung và phát hiện giả mạo vùng.

- Trong trường hợp đầu tiên, các hệ thống cố gắng phát hiện xem một video hoặc các phần của nó đã được lượng tử hóa nhiều lần

hay chưa [30]. Một video là Nội dung do người dùng tạo (UGC) trên máy ảnh nhưng thể hiện dấu vết của nhiều phép lượng hóa thì video đó có thể đáng ngờ. Tuy nhiên, các cách tiếp cận như vậy không phù hợp vì trong phần lớn các trường hợp, video được lấy từ các nguồn truyền thông xã hội. Do đó, cả video bị giả mạo và chưa được kiểm tra thường trải qua nhiều lần lượng tử hóa và rất khó để xác thực nếu không có quyền truy cập vào bản gốc của máy ảnh.

- Trong loại thứ hai, để phát hiện giả mạo giữa các khung, các thuật toán nhằm mục đích phát hiện các trường hợp khung mới đã được chèn thêm vào video [37]. Có hai loại video mà sự giả mạo như vậy có thể thực sự thành công để đánh lừa người xem: **Một là**, trường hợp video đã có các đoạn cắt, tức là cảnh đã chỉnh sửa. Ở đó, một cảnh quay có thể bị xóa hoặc thêm vào trong số các ảnh hiện có, nếu bản âm thanh có thể được chỉnh sửa tương ứng. **Hai là**, trường hợp của video CCTV hoặc các cảnh video được quay từ một camera tĩnh, ở đó, các khung hình có thể được chèn, xóa hoặc thay thế mà không gây chú ý về mặt trực quan. Tuy nhiên, các nghiên cứu thử nghiệm sơ bộ cho thấy rằng các thuật toán được thiết kế cho hình ảnh không hoạt động tốt trên video [27].

- Cuối cùng, loại thứ ba phát hiện giả mạo vùng liên quan đến các trường hợp các phần của chuỗi video (ví dụ: một đối tượng) đã được chèn vào các khung của một video khác - đây là kịch bản thường gặp nhất cho UGC. Các thuật toán phát hiện giả mạo vùng video chia sẻ nhiều nguyên tắc chung với các thuật toán phát hiện ghép nối hình ảnh. Một số cách tiếp cận dựa trên thông tin về mặt không gian được trích xuất riêng từ các khung; trong số đó, những phương pháp nổi bật nhất là sử dụng gradient có định hướng hoặc biểu đồ hệ số biến đổi Cosine rời rạc (DCT). Chúng hoạt động tốt trên những video chất lượng cao, nhưng có xu hướng không thành công ở độ nén cao hơn vì các dấu vết hầu như đã bị xóa. Các chiến lược phát hiện giả mạo vùng khác dựa trên thành phần chuyển động của mã hóa video, lập mô hình thống kê vector chuyển động hoặc thống kê lỗi bù chuyển động. Các phương pháp này hoạt động tốt

hơn với nền tĩnh và các đối tượng chuyển động chậm, sử dụng chuyển động để xác định hình dạng/đối tượng cần quan tâm trong video. Tuy nhiên, những điều kiện này thường không đáp ứng được UGC.

2.2. Một số thuật toán phát hiện điểm cắt, ghép trong video và đề xuất

2.2.1. Một số thuật toán phát hiện điểm cắt, ghép trong video

2.2.1.1. Phương pháp tiếp cận dựa trên đặc trưng ảnh

Giám định hình ảnh là một lĩnh vực lâu đời hơn giám định video; với khối lượng lớn các thuật toán đã được xây dựng dựa trên khai thác các đặc trưng ảnh kỹ thuật số cùng lượng lớn các bộ dữ liệu thử nghiệm, giám định hình ảnh đang dần đạt đến độ chín khi các thuật toán hoặc các tổ hợp thuật toán đang đạt đến độ chính xác tối đa cho ứng dụng trong thế giới thực. Việc phát hiện giả mạo hình ảnh thường dựa trên việc phát hiện sự không nhất quán cục bộ trong thông tin nén JPEG, hoặc phát hiện sự không nhất quán cục bộ trong các mẫu nhiễu tần số cao do thiết bị chụp để lại (đặc biệt trong những trường hợp hình ảnh chất lượng cao, độ nén thấp). Sự tiến bộ trong giám định hình ảnh có thể đưa ra kết luận rằng các phương pháp tương tự có thể hoạt động để phát hiện video giả mạo. Cụ thể:

Các phương pháp giám định hình ảnh thường được tổ chức theo một trong hai loại sau: (1) *Giám định tích cực*, trong đó hình mờ hoặc phần thông tin tương tự (thường không nhìn thấy) được nhúng vào hình ảnh tại thời điểm chụp, trong đó tính toàn vẹn được đảm bảo rằng hình ảnh không bị chỉnh sửa kể từ khi chụp [13] [24] [25]; và (2) *Giám định thụ động*, khi không tồn tại thông tin trước đó và việc phân tích xem một hình ảnh có bị giả mạo hay không hoàn toàn phụ thuộc vào chính nội dung hình ảnh đó. Mặc dù giám sát thụ động là một nhiệm vụ khó khăn hơn nhiều, nhưng nó có liên quan nhất trong phần lớn các trường hợp sử dụng thực tế, khi chúng ta thường không có quyền truy cập vào quá trình chụp ảnh. Ba nhóm chính của sự chỉnh sửa là: (1) *Sao chép di chuyển (copy-move)*, một phần của

hình ảnh được sao chép và đặt ở vị trí khác trong ảnh; (2) *Ghép nối hoặc giả mạo nội khung*, một phần của hình ảnh này được đặt trong hình ảnh khác; (3) *Thay đổi toàn bộ hình ảnh*, tức là khi một phần của hình ảnh bị xóa và sau đó được tự động vẽ bằng thuật toán in-painting về nguyên tắc tương tự. Các thuật toán phát hiện copy-move cố gắng nắm bắt sự giả mạo bằng cách tìm kiếm các điểm tự tương đồng trong hình ảnh [29] [34]; các thuật toán phát hiện và xác định vị trí ghép nối dựa trên tiền đề rằng, ở một mức độ nào đó (có thể không nhìn thấy) khu vực được ghép sẽ khác với phần còn lại của hình ảnh do lịch sử chụp và nén khác nhau của chúng; dựa trên nhận định phần do máy tính tạo ra sẽ mang một đặc điểm khác với phần còn lại của hình ảnh.

Nhận xét: Mặc dù, với sự ra đời của học sâu (deep learning), các phương pháp tiếp cận mới, cố gắng tận dụng sức mạnh của mạng nơ-ron phức hợp trong xử lý ảnh để xác định và phát hiện vị trí giả mạo trong video, nhưng cho đến nay, phương pháp giám định video dựa trên đặc trưng ảnh mới chỉ có thể hoạt động hiệu quả cao nếu video chỉ đơn giản là chuỗi các khung hình, do việc nén video hiện đại là một quá trình phức tạp hơn nhiều, nó thường loại bỏ tất cả các dấu vết như lỗi máy ảnh và dấu vết nén khung hình đơn. Vì vậy, phương pháp tiếp cận trên là chưa khả thi trong ứng dụng thực tế và không đáp ứng được sự phát triển của thế giới công nghệ video hiện nay.

2.2.1.2. Phương pháp tiếp cận dựa trên đặc trưng luồng đa phương tiện

Để giải quyết bài toán giám định video nói chung và xác định vị trí điểm cắt, ghép trong video nói riêng, các nhà khoa học đã nghiên cứu, xây dựng nhiều phương pháp dựa trên đặc trưng luồng đa phương tiện, trong đó sự ra đời của các bộ lọc đã mang lại nhiều hiệu quả ứng dụng cao, cụ thể:

2.2.1.2.1. Sử dụng các bộ lọc số học

Thuật ngữ các bộ lọc số học đề cập đến các cách tiếp cận đại số cho phép chiếu thông tin vào một không gian đặc trưng nhằm giúp việc phân tích trong công tác giám định video dễ dàng hơn [13].

- Các bộ lọc **Q4** được sử dụng để phân tích sự phân rã của hình ảnh thông qua Biến đổi Cosine rời rạc (DCT). DCT 2 chiều chuyển đổi một ma trận $N \times N$ khối hình ảnh thành một ma trận mới $N \times N$ khối, trong đó hệ số được tính toán dựa trên tần số của chúng. Cụ thể trong mỗi khối, hệ số đầu tiên nằm ở vị trí (0,0) đại diện cho thông tin tần số thấp nhất và giá trị của nó có liên quan đến giá trị trung bình của toàn bộ khối, hệ số (0,1) bên cạnh nó đặc trưng cho sự thay đổi chậm từ tối sang sáng theo hướng ngang, v.v. Ngoài ra, người ta có thể tạo ra hình ảnh màu sai bằng cách chọn ba vị trí khối và sử dụng ba mảng kết quả làm kênh màu đỏ, xanh lục và xanh lam của hình ảnh đầu ra, như được minh họa bằng biểu thức dưới đây:

- Các bộ lọc **Chrome** chuyên dùng để phân tích nhiễu độ chói của hình ảnh, làm nổi bật tính đồng nhất của nhiễu được mong đợi trong một hệ thống quan sát bình thường và được chiếu sáng tự nhiên. Bộ lọc này tương tự như thuật toán nhiễu trung bình (Median Noise) cho giám định hình ảnh, được sử dụng để phát hiện sự không nhất quán trong hình ảnh vì *nó có khả năng cô lập nhiễu tần số cao*. Cách tiếp cận này cung cấp cái nhìn tổng quan về toàn bộ khung hình, các vị trí có dấu vết nhiễu khác nhau có thể được phát hiện và xác định là nổi bật hơn so với phần còn lại của khung hình.

2.2.1.2.2. Bộ lọc quang học

Đối với các video thu được từ hệ thống quang học kết hợp với hệ thống cảm biến, ánh sáng và thông tin quang học được chuyển thành dữ liệu kỹ thuật số dưới dạng một luồng video. Rất nhiều thông tin liên quan trực tiếp đến ánh sáng và thông tin quang học ban đầu được thiết bị thu nhận được ẩn trong cấu trúc của video. Mục đích của bộ lọc quang học là để trích xuất thông tin này cho phép người điều tra tìm kiếm sự bất thường trong các mẫu thông tin quang học.

- Bộ lọc **Fluor** được sử dụng để nghiên cứu màu sắc của hình ảnh bất kể mức độ chói của nó. Bộ lọc tạo ra một hình ảnh bình thường trong đó màu sắc của hình ảnh ban đầu đã được khôi phục độc lập với độ chói; thông qua quá trình chuẩn hóa màu sắc được thực hiện bởi bộ lọc Fluor giúp nó có thể phá vỡ sự tương đồng của màu sắc khi nó được hệ thống thị giác của con người cảm nhận và làm nổi bật các màu có sự khác biệt rõ rệt hơn dựa trên màu sắc thực tế của chúng.

- Bộ lọc **Focus** được sử dụng để xác định các khu vực sắc nét trong một hình ảnh hoặc các vùng có độ nét mạnh hơn. Một hình ảnh có độ sắc nét cao chứa lượng tần số cao nhiều hơn, trong khi ngược lại, các tần số cao không phù hợp khi đối tượng bị mờ hoặc mất nét. Bộ lọc Focus chỉ xem xét độ phân giải của wavelet thông qua bộ lọc phi tuyến dựa trên việc xử lý bộ RGB của mỗi khung. Nó tạo ra bố cục màu sai nơi mà các vùng tần số thấp bị mờ vẫn có màu xám và các đường viền sắc nét xuất hiện màu.

- Bộ lọc **Acutance** đề cập đến thuật ngữ vật lý để chỉ độ sắc nét trong hình ảnh. Thông thường, nó là một phép đo đơn giản về độ dốc của gradient cục bộ nhưng ở đây nó được chuẩn hóa với giá trị cục bộ của các mức xám, khác với bộ lọc Focus. Bộ lọc Acutance được tính bằng tỷ số giữa các đầu ra của một bộ lọc high-pass và một bộ lọc low-pass. Trong thực tế, có thể sử dụng hai bộ lọc Gaussian với các kích thước khác nhau.

2.2.1.2.3. Các bộ lọc thời gian

Những bộ lọc này nhằm mục đích làm nổi bật hoạt động của luồng video theo thời gian. Việc nén video MPEG-4 khai thác khả năng dư thừa theo thời gian để giảm kích thước video nén - đây là lý do tại sao nén một video phức tạp hơn nhiều so với nén một chuỗi các hình ảnh. Hơn nữa, trong nhiều khung hình, MPEG-4 kết hợp các dự đoán theo một hướng hoặc theo chiến lược tiến/lùi, do đó, việc biểu diễn khung phụ thuộc nhiều vào nội dung khung và mức độ lượng tử hóa. Do đó, việc phân tích theo thời gian của các tham số

lượng tử hóa có thể giúp chúng ta phát hiện ra sự mâu thuẫn trong biểu diễn khung.

- Bộ lọc **Cobalt** so sánh video gốc với phiên bản sửa đổi được định dạng lại bằng MPEG-4 với mức chất lượng khác (tốc độ bit tương ứng khác). Nguyên tắc của bộ lọc Cobalt rất đơn giản: một bộ quan sát video về các lỗi giữa video đầu tiên và video được định lượng lại bằng MPEG-4 với mức chất lượng thay đổi hoặc mức tốc độ bit thay đổi, nếu mức lượng tử hóa trùng với mức chất lượng thực sự được sử dụng trên khu vực sửa đổi nhỏ, thì sẽ không có lỗi tại đó.

- Bộ lọc các **Vector chuyển động** biểu diễn video dựa trên màu sắc của các khối chuyển động khi được mã hóa vào luồng video. Loại biểu diễn này sử dụng các mũi tên để hiển thị các sự dịch chuyển của các khối. Các vector chuyển động được mã hóa trong luồng video để tái tạo lại tất cả các khung không phải là khung chính. Sau đó, một đối tượng của cảnh có một tập hợp các vector chuyển động được liên kết với mỗi khối macro bên trong nó. Những chuyển động này được đại diện bởi bộ lọc vector chuyển động phải đồng nhất và nhất quán, nếu không có khả năng cao là một số hoạt động đáng ngờ liên quan việc chỉnh, sửa video đã xảy ra.

- Bộ lọc **Temporal** được sử dụng để áp dụng chuyển đổi thời gian trên video, chẳng hạn như làm mịn hoặc điều chỉnh thời gian. Nó cũng được sử dụng để so sánh khung hình chỉ tập trung vào sự phát triển của độ sáng theo thời gian.

Nhận xét: Phương pháp tiếp cận dựa trên đặc trưng luồng đa phương tiện là một phương pháp tiếp cận hiệu quả trong khi giám định video ở mức khung hình vẫn là chiến lược khả thi duy nhất với dữ liệu có sẵn hạn chế như hiện nay. Trong đó, bộ lọc Q4 (với phép biến đổi Cosine rời rạc) và bộ lọc Cobalt thường được ứng dụng thực tế nhiều hơn trong việc phân loại dữ liệu đầu ra của các hệ thống giám định video bởi tính trực quan và hiệu quả của nó.

2.2.1.3. Phương pháp tiếp cận dựa trên tín hiệu audio

Những sửa đổi trong video thường để lại dấu vết bởi sự khác biệt giữa hình ảnh và âm thanh - một đặc trưng chúng ta có thể khai thác nhằm phát hiện sự chỉnh sửa, cắt/ghép video. *Phát hiện sự không nhất quán về âm thanh - hình ảnh* là một hướng tiếp cận mà nhiều nhà khoa học đang nghiên cứu, phát triển để phát hiện và mô tả đặc điểm của nhiều loại mâu thuẫn liên quan đến các khía cạnh khác nhau tạo nên một video hoàn chỉnh [31]. Sự không nhất quán của *môi trường xung quanh* và *đặc trưng âm thanh/hình ảnh người nói* đại diện cho các loại chỉnh sửa khác nhau và là bằng chứng xác thực, cũng như thách thức cho công tác giám định video. Một hệ thống cơ bản để triển khai phát hiện video đã bị chỉnh sửa theo hướng tiếp cận này thường gồm 3 giai đoạn chính: (1) Phát hiện sự mâu thuẫn âm thanh và hình ảnh trong các cảnh; (2) Phát hiện sự không nhất quán trong những thay đổi của âm thanh nghe được và âm thanh do nhận dạng hình ảnh trong video; và (3) Xây dựng các bộ dữ liệu và triển khai để đánh giá hiệu quả thuật toán.

Một số kỹ thuật trong phương pháp tiếp cận này như:

(1) *Phân tích, phát hiện âm thanh của cảnh*: Hệ thống phát hiện âm thanh cảnh dựa trên một I-vector được mô hình hóa bằng cách sử dụng một nền tảng Gaussian (số lượng của bộ Gaussian dựa trên số lượng các lớp).

(2) *Phát hiện cảnh không nhất quán*: So sánh các track âm thanh và hình ảnh đã được chú thích, chúng ta có thể dễ dàng nhận thấy sự mâu thuẫn giữa âm thanh và hình ảnh trong video.

(3) *Phát hiện sự không nhất quán của người nói*: Một số sự chỉnh sửa liên quan đến việc thay thế lời nói của một người bằng lời nói của người khác, chẳng hạn như lồng tiếng hoặc có thể thay thế khuôn mặt (deepfakes) nhưng không chỉnh sửa âm thanh. Những sự chỉnh sửa này có thể tạo ra sự mâu thuẫn giữa trạng thái hay cảm xúc của khuôn mặt đang nói chuyện. Dựa trên cơ sở dữ liệu về các đối tượng đã biết với các mẫu giọng nói và khuôn mặt của họ, chúng ta

có thể kiểm tra các kết quả nghe-nhìn không nhất quán với cơ sở dữ liệu.

Nhận xét: Mặc dù việc thử nghiệm trên một bộ sưu tập video giả mạo của các nhà khoa học đã cho thấy nhiều kết quả hứa hẹn cho phương pháp này, tuy nhiên, phương pháp tiếp cận này mới dừng ở việc phát hiện các video bị giả mạo, chưa phát hiện được vị trí, điểm cắt, ghép trong video theo yêu cầu Đề tài đặt ra.

2.2.2. Đề xuất thuật toán giải quyết bài toán

Những nghiên cứu, khảo sát gần đây đã cho thấy, tương ứng với mỗi loại giả mạo trong video sẽ có những kỹ thuật phát hiện phù hợp [27], như: sử dụng kỹ thuật Khớp khối (Block matching), Biến đổi Cosine rời rạc (DCT), Phân tích thành phần chính (PCA), Tương quan chuỗi để phát hiện giả mạo *Sao chép-di chuyển (nhân bản)*; kỹ thuật Phân tích quang phổ hai mặt, phân tích kết hợp song song, ước tính biến thiên nhiễu, thống kê bậc cao để phát hiện giả mạo *nối (Splicing)*...

Do đó, dựa trên những nghiên cứu hiện có này, học viên định hướng triển khai thực nghiệm bằng cách nghiên cứu xây dựng một hệ thống phần mềm giám định video theo ***phương pháp tiếp cận dựa trên kỹ thuật biến đổi Cosin rời rạc*** - một phương pháp phổ biến trong phát hiện video giả mạo loại copy-move và được đánh giá có kết quả với độ chính xác cao; hướng đến mục tiêu làm nổi bật các dấu vết nào do giả mạo video để lại, với trọng tâm là xác định sự gián đoạn trong các khía cạnh tạm thời của video làm nền tảng cho các nghiên cứu phát triển tiếp theo trong tương lai; qua đó đánh giá hiệu quả của phương pháp thực nghiệm trên.

Qua quá trình khảo sát về thực trạng thế giới công nghệ hiện nay liên quan đến mức độ phù hợp của hướng giải quyết bài toán trên, công cụ giám định mà học viên phát triển có tiềm năng rất lớn trong việc xử lý cục bộ các tập dữ liệu video giả mạo, cũng như phương pháp học sâu mà học viên nghiên cứu đã góp phần tích cực và ảnh hưởng rất lớn tới hiệu quả tự động phát hiện giả mạo video.

Dựa trên kết quả thử nghiệm đối với dữ liệu điểm chuẩn và thế giới thực, đồng thời phân tích kết quả, học viên nhận thấy rằng phương pháp được đề xuất mang lại kết quả đầy hứa hẹn so với phương pháp hiện đại, đặc biệt là đối với khả năng khái quát hóa của thuật toán đối với dữ liệu chưa biết được lấy từ thế giới thực.

Chương 3 - THỬ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Đề tài phát hiện video bị chỉnh sửa hiện nay vẫn là một trong những vấn đề khó, thách thức các chuyên gia, khi phải đối mặt với ngày càng nhiều kỹ thuật chỉnh sửa video hiện đại, tinh vi. Do đó đề tài phát hiện điểm cắt, ghép trong video lại càng khó hơn nữa, đây có thể là giai đoạn tiếp theo của hệ thống sau khi đã nhận diện được các video bị chỉnh sửa; một số phương pháp hay nghiên cứu hiện nay mới chỉ dừng lại ở việc phát hiện các video bị chỉnh sửa, chưa thể phát hiện được các vị trí hoặc loại chỉnh sửa mà video đã bị tác động. Qua quá trình nghiên cứu các phương pháp sử dụng trong phát hiện điểm cắt ghép trong video, học viên đã tiến hành xây dựng một chương trình dựa trên thuật toán bộ lọc Cosin rời rạc để thực nghiệm và đánh giá về các phương pháp nghiên cứu phát hiện điểm cắt, ghép trong video, làm tiền đề phục vụ cho các nghiên cứu sau này.

3.1. Giới thiệu chương trình

3.1.1. Nền tảng công nghệ

- Chương trình được xây dựng trên nền tảng ngôn ngữ lập trình bậc cao Python version 3, sử dụng công cụ lập trình Pycharm - môi trường phát triển hoàn hảo dành cho ngôn ngữ lập trình Python để thực hiện. Các thư viện hỗ trợ bao gồm thư viện OpenCV, Scipy, Pillow, Numpy...

- Chương trình sử dụng phương pháp biến đổi Cosin rời rạc trong xử lý ảnh kết hợp cùng với phương pháp học máy để tìm kiếm những vùng bị cắt ghép trong mỗi frame ảnh của video đã trích xuất.

3.1.2. Nguồn dữ liệu

Học viên đã lựa chọn bộ dữ liệu Fake Video Corpus của InVID, được phát triển trên nguồn của dự án InVID, bao gồm rất nhiều video đã được chỉnh sửa, cắt ghép nội dung và bên cạnh đó là số lượng lớn video gốc không bị chỉnh sửa. Các video sẽ được xử lý trước khi đưa vào mô hình thử nghiệm theo chuẩn mã hóa H.264/AVC, định dạng file *.mp4, độ dài mỗi video khoảng từ 10-30s.

Để thử nghiệm và đánh giá các phương pháp đã nghiên cứu trong những phần trước, học viên sẽ lựa chọn ngẫu nhiên ra các video trong tập dữ liệu video để thử nghiệm, trong đó sẽ sử dụng các video có chỉnh sửa nhằm áp dụng các thuật toán đã đưa ra để đánh giá hiệu quả của phương pháp áp dụng.

Từ tập dữ liệu đã được chọn, học viên đã chọn lọc lấy ra 05 video có nội dung đã bị chỉnh sửa. Sau đó, sử dụng phần mềm chỉnh sửa video để chuẩn hóa kích thước khung hình của video để làm dữ liệu đầu vào cho việc thử nghiệm chương trình.

3.2. Cấu trúc chương trình

Trong chương trình thử nghiệm, học viên chia ra các module để thực hiện bao gồm: (1) **Module xử lý dữ liệu đầu vào:** Module này thực hiện việc đọc dữ liệu video đầu vào, trích xuất video thành các frame ảnh dưới dạng ảnh xám và lưu lại vào trong từng thư mục riêng biệt. Thư viện của ngôn ngữ lập trình Python là OpenCV hỗ trợ rất mạnh trong xử lý công việc này; (2) **Module xử lý phát hiện điểm cắt ghép:** Đây là module quan trọng nhất, đảm nhận nhiệm vụ chính đó là xử lý hình ảnh. Áp dụng phép biến đổi Cosin rời rạc kết hợp với các thuật toán học máy trong bộ thư viện của Python để tìm ra các điểm bất thường, có khả năng là những điểm bị chỉnh sửa trong mỗi khung ảnh của video đã trích xuất. Sau đó xuất khung ảnh đã qua xử lý ra thư mục đầu ra riêng biệt; (3) **Module chuyển đổi ảnh sang video:** Sau khi đã xử lý toàn bộ các khung ảnh, công việc tiếp theo là phải chuyển đổi các khung ảnh đã xử lý thành video kết

quả hoàn chỉnh. Các video được chuyển đổi theo các chuẩn mà thư viện Python hỗ trợ. Cụ thể:

3.2.1. Xử lý dữ liệu đầu vào

Trong chương trình, ta sử dụng dữ liệu đầu vào định dạng *.MP4. Mỗi video có một kích thước khung hình, độ lớn khác nhau, nó sẽ ảnh hưởng tới hiệu suất xử lý việc phát hiện điểm cắt ghép trong video. Qua thử nghiệm, với mỗi khung hình khác nhau sẽ có thời gian xử lý khác nhau, do đó trong chương trình này, học viên áp dụng thử nghiệm trên video có độ phân giải 360 pixels và chuẩn hóa sang video vuông có kích thước dài, rộng như nhau bằng 360 pixel để thuận lợi cho việc thử nghiệm trên mô hình.

Sau khi đã chỉnh sửa video, học viên sẽ tiến hành tách các khung hình từ video đã chọn và chỉnh sửa. Đồng thời, trong quá trình tách các khung hình từ video, ta sẽ tiến hành chuyển đổi các khung hình sang dạng ảnh xám để làm đầu vào cho quá trình xử lý ở module tiếp theo. Mỗi khung hình sẽ được lưu lại dưới định dạng ảnh *.png tại thư mục cố định. Module này được xây dựng trên cơ sở áp dụng các hàm trong thư viện OpenCV của Python - một thư viện rất mạnh và quen thuộc trong xử lý đa phương tiện.

3.2.2. Xử lý tìm điểm cắt ghép trong từng khung hình

Quá trình xử lý tìm điểm cắt ghép trong từng khung hình của video được thực hiện theo một quy trình tuần tự. Cụ thể:

1) Chia ảnh xám đã được Module xử lý thực hiện thành các khối vuông điểm ảnh chồng nhau với kích thước cố định, ở đây ta áp dụng kích thước mỗi khối là 8x8 pixel. Các khối vuông điểm ảnh này được lấy từ góc trên màn hình bên trái theo thứ tự từ trên xuống dưới, từ trái qua phải đến góc dưới màn hình bên phải tạo thành các khối xếp chồng lên nhau. Mỗi khối được biểu diễn là điểm bắt đầu của hàng và cột tương ứng trong mỗi khung ảnh.

2) Áp dụng phép biến đổi Cosin rời rạc (DCT) cho mỗi khối điểm ảnh đã được trích xuất ra. Các khối điểm ảnh sẽ được gán vào trong một mảng 2 chiều, làm dữ liệu đầu vào cho hàm chức năng của phép biến đổi Cosin rời rạc.

3) Trích chọn đặc trưng từ các khối vuông đã áp dụng DCT. Thông qua phép biến đổi Cosin rời rạc, ta sẽ được một ma trận gồm các trọng số của DCT. Ma trận trọng số có thể được sắp xếp theo hình zig-zag để lấy các thông tin được lưu trong các khối đã hiển thị. Qua phép biến đổi Cosin rời rạc, mỗi khối sẽ được gán một trọng số nhất định làm cơ sở để trích chọn các đặc trưng mỗi khối điểm ảnh.

4) Dùng thuật toán học máy là phương pháp phân tích thành phần chính PCA với hàm lõi Gaussian RBF để tối ưu hóa đặc trưng, giảm thiểu chiều dữ liệu, xác định số lượng nhóm nhất định kết hợp áp dụng phương pháp phân cụm K-nearest neighbor (viết tắt là KNN). Kế thừa các nghiên cứu trong và ngoài nước về phân cụm dữ liệu trong xử lý ảnh, học viên đã áp dụng tính toán khoảng cách Ecudid giữa các cụm dữ liệu đã được xử lý; với một ngưỡng đã được nghiên cứu, tính toán cho trước, áp dụng tính toán khoảng cách giữa các cụm dữ liệu, tiến hành phân loại thành các cụm có đặc điểm tương đồng, để đưa ra những cụm dữ liệu có đặc điểm giống nhau, như một kết quả cho việc dự đoán các vùng bị chỉnh sửa trong mỗi khung ảnh của video.

5) Cuối cùng ta sẽ tiến hành xóa bỏ những khối có kích thước nhỏ, nằm rời rạc nhau để đưa ra vùng dự đoán cuối cùng.

Đây là module quan trọng nhất trong chương trình phát hiện điểm cắt ghép trong video. Các tham số được sử dụng trong thuật toán gồm tham số lượng tử hóa, ngưỡng khoảng cách Ecudid giữa các khối, ngưỡng khoảng cách Ecudid giữa các pixel. Với tùy từng tham số trên, hiệu suất tính toán cũng như kết quả đưa ra sẽ có sự khác nhau.

3.3. Kết quả thực nghiệm

Thuật toán được thử nghiệm trên 5 bộ video từ nguồn InVID Fake Video Corpus là những video đã được chỉnh sửa lấy trên nguồn các trang mạng xã hội Youtube và Facebook có dung lượng từ 10-30s. Thiết bị thử nghiệm chạy thuật toán là máy tính có cấu hình Intel Core i7 4700MQ 2.4GHz, ổ cứng SSD 512GB, RAM 8GB để thử nghiệm khả năng tính toán của thuật toán. Ta sẽ thử nghiệm các bộ dữ liệu trên các giá trị khác nhau của phương pháp áp dụng bộ lọc Cosin rời rạc gồm: Kích thước các khối điểm ảnh: B ; Khoảng cách Eculid giữa các khối điểm ảnh: d_B ; Khoảng cách Eculid giữa các pixel: d_p ; Giới hạn vector dịch chuyển: Vec_{limit} ;

Các kết quả thử nghiệm như sau:

- Kết quả thử nghiệm $B=16 \times 16$, $d_B=5$, $d_p=20$, $Vec_{limit}=20$

Bảng 3.1. Kết quả thực nghiệm khối 16×16 pixels

Bộ video	Độ chính xác	TPR	FPR
Video 1	0.683	0.821	0.073
Video 2	0.667	0.805	0.082
Video 3	0.635	0.763	0.080
Video 4	0.598	0.725	0.092
Video 5	0.572	0.689	0.112

- Kết quả thử nghiệm $B=24 \times 24$, $d_B=5$, $d_p=20$, $Vec_{limit}=20$

Bảng 3.2. Kết quả thực nghiệm khối 24×24 pixels

Bộ video	Độ chính xác	TPR	FPR
----------	--------------	-----	-----

Video 1	0.723	0.841	0.064
Video 2	0.687	0.725	0.092
Video 3	0.612	0.689	0.098
Video 4	0.598	0.693	0.106
Video 5	0.552	0.632	0.131

Qua hai lần thử nghiệm với hai bộ khối điểm ảnh khác nhau nhận thấy độ chính xác áp dụng mô hình trên có sự thay đổi. Ngoài sự ảnh hưởng việc thay đổi kích thước khối điểm ảnh trong thuật toán, độ chính xác còn phụ thuộc vào chất lượng video rất nhiều. Video càng có độ sắc nét cao thì khả năng chính xác cao hơn tuy nhiên thời gian thực hiện thuật toán lại khá lớn (trung bình 4 phút/khung ảnh). Tuy nhiên việc thực hiện thuật toán trên vẫn đảm bảo được việc phát hiện các video bị chỉnh sửa mặc dù độ chính xác vẫn chưa cao và thời gian thực hiện còn lớn.

3.4. Nhận xét

Qua nghiên cứu, thực nghiệm phương pháp trên, từ kết quả cho thấy một số vấn đề cần phải giải quyết:

- Về phương pháp, đây là phương pháp đơn giản áp dụng việc xử lý từng khung ảnh trong video để tiến hành tìm ra vùng bị chỉnh sửa. Phương pháp sử dụng các biện pháp xử lý ảnh cơ bản và nâng cao, các thư viện sẵn có giúp cho việc lập trình dễ dàng hơn. Ngoài ra, phương pháp không đòi hỏi độ phức tạp trong xử lý dữ liệu, cũng như áp dụng các biện pháp học máy nâng cao để xây dựng module, do vậy giảm bớt được nhiều công đoạn tính toán phức tạp. Tuy nhiên, mặt hạn chế của phương pháp đó là thời gian xử lý mỗi khung ảnh khá dài, đặc biệt là với những video có chất lượng cao.

Các vùng dự đoán đưa ra còn bị ảnh hưởng bởi nhiều, phải chọn lọc các tham số đầu vào cho từng video để cho kết quả tốt nhất.

- Về kết quả thử nghiệm, bộ dữ liệu video thử nghiệm được lựa chọn là những video đã bị cắt ghép, chỉnh sửa. Qua quá trình thử nghiệm với các thông số khác nhau đã đánh giá được hiệu suất và khả năng của phương pháp được áp dụng. Độ chính xác của phương pháp giảm dần theo thời gian của mỗi video, video có thời gian càng ngắn thì độ chính xác càng cao và ngược lại. Đồng thời những video có độ phân giải cao sẽ có thời gian thực hiện lâu hơn, tuy nhiên kết quả lại chính xác hơn. Các tham số đầu vào của phương pháp cũng ảnh hưởng nhiều tới kết quả. Vì áp dụng việc chia các khối điểm ảnh do vậy với kích thước các khối điểm ảnh sẽ cho kết quả tính toán khác nhau, từ đó ảnh hưởng tới hiệu quả của phương pháp.

KẾT LUẬN

Sự xuất hiện của mạng xã hội nói chung và sự lan tỏa mạnh mẽ của các video đã làm thay đổi thế giới, hình thành một "thế giới ảo" đan xen với thế giới thực, tạo ra tương tác tối đa trong mọi quan hệ xã hội, vượt qua mọi khoảng cách về không gian và thời gian. Chính sự phát triển mạnh mẽ, đa dạng của các loại hình mạng xã hội tạo ra mạng lưới truyền thông đa phương tiện ngày càng hiện đại, để mọi người có thể trao đổi, tiếp cận thông tin nhanh chóng trên khắp thế giới, tạo điều kiện phát triển mọi mặt của đời sống xã hội; đồng thời, cũng tạo điều kiện thuận lợi để các thế lực phản động và tội phạm sử dụng vào các hoạt động vi phạm pháp luật, gây ra nguy cơ mất an ninh, an toàn thông tin, đặc biệt nguy hiểm khi chúng sử dụng để tuyên truyền, xuyên tạc thông tin, kích động biểu tình. Đáng chú ý, các đối tượng phạm tội đã và đang gia tăng các hoạt động giả mạo video để bịa đặt, xuyên tạc thông tin gây mất uy tín cá nhân, tổ chức, chính quyền; cũng như phục vụ các hành vi lừa đảo chiếm đoạt tài sản. Tại nước ta, công tác giám định video phục vụ xác thực tính chính xác của thông tin lan truyền trên internet và công tác điều tra,

đầu tranh tội phạm còn nhiều hạn chế, việc ứng dụng các kỹ thuật, công nghệ hiện đại chưa đạt được nhiều thành tựu và phục vụ nhu cầu thực tiễn của xã hội. Do đó, yêu cầu cấp thiết đặt ra hiện nay là phải nghiên cứu, xây dựng các giải pháp để phát hiện điểm cắt, ghép trong video phục vụ công tác giám định hình sự.

Trong phạm vi nghiên cứu của luận văn, học viên đã trình bày một số vấn đề liên quan đến giám định video, trong đó tập trung nghiên cứu một số phương pháp tiếp cận giải quyết vấn đề bài toán đặt ra, từ đó đề xuất xây dựng chương trình thực nghiệm bằng biện pháp biến đổi Cosin rời rạc, thu được một số kết quả nhất định. Qua đó, học viên nhận thấy nghiên cứu, phát triển các giải pháp trong phát hiện điểm cắt ghép trong video là một hướng nghiên cứu mới, đáp ứng được tình hình thực tế trong thời kỳ khoa học công nghệ phát triển, các công cụ chỉnh sửa hình ảnh, video ngày càng phát triển, công nghệ chỉnh sửa tinh vi và hiện đại. Tuy nhiên, quá trình nghiên cứu đòi hỏi thời gian và thử nghiệm nhiều phương pháp khác nhau để đưa ra hiệu quả tốt nhất.

Trong tương lai, hướng nghiên cứu sâu hơn về việc phát hiện các điểm cắt ghép trong video nói riêng và phát hiện chỉnh sửa trong video nói chung cần xem xét đến sự thành công hiện có của các phương pháp tiên tiến hiện đại bắt nguồn từ thị giác máy tính, khai thác dữ liệu lớn và khoa học máy tính. Học viên sẽ tập trung nghiên cứu các phương pháp khác làm tăng hiệu suất mô hình như các phương pháp biến đổi không giới hạn về mô-men hình học (biến đổi Fourier - Mellin) và các đặc trưng dựa trên cấu trúc video.

Bám sát mục tiêu, nhiệm vụ, sử dụng đúng đắn các phương pháp nghiên cứu khoa học, luận văn đã thu được một số thành công và về cơ bản đã giải quyết tốt mục tiêu, nhiệm vụ nghiên cứu đặt ra. Luận văn là công trình nghiên cứu công phu, nghiêm túc, song do đây là vấn đề mới, rất khó và phức tạp, phạm vi nghiên cứu rộng, cộng thêm những khó khăn khách quan, kiến thức của học viên còn hạn chế nên chắc chắn còn nhiều khiếm khuyết. Học viên rất mong

nhận được sự quan tâm, góp ý của các nhà khoa học, nhà hoạt động thực tiễn và đồng nghiệp. Cuối cùng, học viên xin chân thành cảm ơn các đơn vị liên quan, các đồng chí, đồng nghiệp, đặc biệt là thầy Phó Giáo sư Hà Hải Nam, người hướng dẫn khoa học đã tận tình giúp đỡ để học viên hoàn thành luận văn này./.