LỜI CAM ĐOAN

Tôi xin cam đoan đồ án này là do chính tôi thực hiện.

Tất cả những tham khảo từ các nghiên cứu liên quan đều được nêu nguồn gốc một cách rõ ràng từ danh mục tài liệu tham khảo trong đồ án. Trong đồ án, không có việc sử dụng tài liệu, công trình nghiên cứu của người khác mà không chỉ rõ về tài liệu tham khảo. Các chương trình thực nghiệm đều là do tôi thực hiện mà có được, không sao chép từ bất cứ nguồn nào.

|  |
| --- |
| Hà Nội, ngày 12 tháng 01 năm 2019 |
| Sinh viên thực hiện: |
| Nguyễn Thành Đồng |

**LỜI MỞ ĐẦU**

Bước vào thế kỷ XXI, loài người đã chứng kiến một cuộc bùng nổ thật sự, sau cuộc cách mạng phát minh ra máy tính điện tử, đó là Internet. Chỉ trong vòng có vài năm ngắn ngủi, Internet đã nhanh chóng len lỏi vào từng ngóc ngách của đời sống, và càng ngày càng đóng những vai trò quan trọng trong cuộc sống của con người.

Từ khi loài người bước ra khỏi thời kỳ công xã nguyên thủy, và chủ nghĩa cá nhân xuất hiện, thì nhu cầu bảo vệ dấu ấn cá nhân trong cộng đồng ngày càng trở nên bức bách. Các nghệ sỹ đua nhau sáng tạo ra những trường phái mới, chỉ với mục đích làm cho tác phẩm của mình không bị lẫn lộn với những tác phẩm của người khác. Những người giàu có sẵn sàng bỏ ra nhiều triệu đô la, chỉ để sở hữu một chiếc điện thoại di động không có gì độc đáo về mặt tính năng nhưng là duy nhất trên thế giới.

Bước vào thời kỳ kinh tế tri thức, khi tri thức ngày càng trở nên đắt giá, đồng thời với đó, các tài liệu được lưu ở dạng số hóa ngày càng nhiều và phổ biến, thì vấn đề bảo vệ bản quyền cho tri thức của con người ngày càng trở nên quan trọng, bởi những đặc trưng sau của tài liệu số:

**Dễ dàng sao chép**: nhờ sự phát triển của máy vi tính, các tài liệu số hóa có thể được sao chép một cách đơn giản và nhanh chóng. Chỉ cần một vài thao tác đơn giản, như click chuột, một cuốn tiểu thuyết dày hàng nghìn trang, hay một tác phẩm trị giá nhiều triệu đô la của danh họa Picasso có thể được sao chép chỉ trong vài giây. Điều quan trọng hơn nữa là khi sao chép tài liệu số thì chất lượng bản sao chép được giữ nguyên so với bản gốc.

**Dễ dàng phát tán**: những năm gần đây, Việt Nam chứng kiến cuộc chay đua về nâng cấp hạ tầng đường truyền Internet giữa các nhà cung cấp dịch vụ. Điều này giúp cho khách hàng có được dịch vụ tốt hơn, nhưng đồng nghĩa với đó là việc phát tán thông tin cũng như tài liệu qua mạng Internet càng trở nên dễ dàng hơn bao giờ hết. Ngày nay, chỉ sau vài phút tìm kiếm trên mạng, người sử dụng có thể dễ dàng tìm và tải về những bộ phim mới nhất còn chưa được trình chiếu ở rạp. Cùng với đó, một người sử dụng bình thường có thể trở thành một nguồn phát tán tài liệu cũng rất dễ dàng, thông qua các tin nhắn tức thời (IM – Instant Message), email hay các dịch vụ chia sẻ file trực tuyến (online file sharing service). Điều này có thể gây nên một hiệu ứng lan truyền như hiệu ứng virus, khiến tài liệu được phát tán nhanh và rộng khắp. Ví dụ, đoạn clip về Susan Boyle được đưa lên mạng vào tháng 4.2009 vừa qua đã đạt được hơn 200 triệu người xem chỉ trong có 2 tuần.

**Dễ dàng lưu trữ**: dung lượng ổ cứng ngày càng lớn, giá thành các thiết bị lưu trữ ngày càng rẻ đã khiến cho việc lưu trữ các tài liệu số hóa trở nên đơn giản hơn bao giờ hết. Một chiếc đĩa CD mà giờ đây ngày càng trở nên lỗi thời, đã có thể chứa hàng trăm cuốn sách điện tử. Một thiết bị nghe nhạc MP3 giá vài trăm nghìn đồng có thể lưu

được hàng vạn bài hát. Những ổ cứng với dung lượng lên tới hàng trăm GB, thậm chí hàng TB khiến cho một máy tính cá nhân dễ dàng biến thành một rạp chiếu phim. Tất cả những điều đó đã khiến cho việc lưu trữ các tài liệu số hóa trở nên dễ dàng hơn bao giờ hết.

Ở Việt Nam, tình hình vi phạm bản quyền vẫn đang diễn biến rất nghiêm trọng, mặc dù đã có sự cố gắng của nhiều cơ quan chức năng, cũng như của cộng đồng. Tình trạng sử dụng phần mềm không có bản quyền, download những file video lậu trên các trang web chia sẻ file, tìm nghe những album mới nhất trên những diễn đàn mà không muốn mua đĩa, v.v… là những hiện tượng hết sức phổ biến.

Để đảm bảo cho nhu cầu giữ bí mật thông tin liên lạc cũng như đảm bảo an toàn dữ liệu, từ lâu con người đã phát minh ra một công cụ hết sức hiệu quả, đó là mã hóa. Mã hóa là một công cụ mạnh, và có lịch sử lâu đời, đã có nhiều kết quả nghiên cứu thành công và có ứng dụng rất lớn trong việc đảm bảo an toàn thông tin liên lạc. Tuy nhiên, mã hóa gặp phải một vấn đề rất lớn, đó là bản thân công cụ mã hóa không gắn liền với tài liệu được bảo vệ, mà chỉ như một dạng vỏ bọc (cover) của tài liệu mà thôi. Do đó, mã hóa không thể được sử dụng như là một công cụ an toàn để bảo vệ bản quyền tài liệu số khi phát hành trên mạng, vì sau khi kẻ gian đã giải mã được tài liệu, thì có toàn quyền đối với tài liệu đó.

Watermarking là một ứng dụng đã có từ lâu đời để bảo vệ bản quyền cho các cuốn sách. Tuy nhiên, thủy vân ảnh số (digital watermarking) lại là một lĩnh vực mới, đang nhận được nhiều sự quan tâm cũng như nghiên cứu của chuyên gia trên thế giới. Sử dụng thủy vân số có thể thay đổi và tác động vào chất lượng của tài liệu số như ý muốn, đồng thời với đó là thủy vân số có thể gắn liền với tài liệu, đảm bảo tài liệu được bảo vệ bản quyền cho tới khi bị hủy hoại. Chính vì vậy, trong nội dung này, em xin phép chọn đề tài “**Thuật toán thủy vân ảnh và áp dụng cho ảnh văn bản hành chính.**”

Mục lục

[LỜI MỞ ĐẦU 1](#_Toc535405263)

[LỜI CẢM ƠN 7](#_Toc535405264)

[CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ ĐỂ TÀI 8](#_Toc535405265)

[1.1. Lí do chọn đề tài 8](#_Toc535405266)

[1.2. Mục tiêu của đề tài 8](#_Toc535405267)

[1.3. Giới hạn và phạm vi của đề tài 8](#_Toc535405268)

[1.4. Kết luận chương 1 8](#_Toc535405269)

[1.5. Khái niệm thủy vân số 9](#_Toc535405270)

[1.6. Các ứng dụng của thủy vân số 13](#_Toc535405271)

[1.7. Các đặc tính của thủy vân 14](#_Toc535405272)

[1.8. Phân loại thủy vân 15](#_Toc535405273)

[1.9. Quy trình thực hiện thủy vân 16](#_Toc535405274)

[CHƯƠNG 2: KỸ THUẬT THỦY VÂN ẢNH VÀ THUẬT TOÁN NHÚNG THỦY VÂN 17](#_Toc535405275)

[2.1. Các toán tử tần số 17](#_Toc535405276)

[2.2. Thuật toán thủy vân DCT trên miền tần số: 23](#_Toc535405277)

[2.3. Một số thuật toán khác: 28](#_Toc535405278)

[CHƯƠNG 3: CÀI ĐẶT VÀ ĐÁNH GIÁ HỆ THỐNG 40](#_Toc535405279)

[3.1. Cài đặt hệ thống: 40](#_Toc535405280)

[3.2. Kết quả thực nghiệm: 44](#_Toc535405281)

[3.3. Đánh giá, so sánh kết quả thực nghiệm: 47](#_Toc535405282)

[Kết luận và hướng nghiên cứu tương lai 47](#_Toc535405283)

[TÀI LIỆU THAM KHẢO 48](#_Toc535405284)

**Danh mục hình vẽ**

[*Hình 1. 1: Thủy vân trên đồng dollar của Mỹ* 10](#_Toc534983268)

[*Hình 1. 2: Một số phương pháp phân loại thủy vân tiêu biểu* 15](#_Toc534983269)

[*Hình 2. 1: Lọc theo tần số thấp và cao* 19](#_Toc534972706)

[*Hình 2. 2: Quá trình phân tích và tổng hợp tín hiệu* 20](#_Toc534972707)

[*Hình 2. 3: Phân tích DWT hai mức* 21](#_Toc534972708)

[*Hình 2. 4: Quá trình nhúng thủy vân* 24](file:///C:\Users\thef4ck\Desktop\thực%20tập.docx#_Toc534972709)

[*Hình 2. 5: Quá trình giải mã* 26](file:///C:\Users\thef4ck\Desktop\thực%20tập.docx#_Toc534972710)

[*Hình 2. 6: Quy trình nhúng thủy vân* 29](file:///C:\Users\thef4ck\Desktop\thực%20tập.docx#_Toc534972711)

[*Hình 2. 7: Quy trình tách thủy vân* 30](#_Toc534972712)

[*Hình 2. 8: Quy trình nhúng thủy vân* 31](#_Toc534972713)

[*Hình 2. 9: Quy trình tìm lại thủy vân* 32](#_Toc534972714)

**LỜI CẢM ƠN**

Việc trao đổi thông tin, xuất bản thông tin trên Internet có nhiều nguy cơ không an toàn do thông tin có thể bị lộ hay bị sửa đổi hay bị vi phạm bản quyền. Nói chung, để bảo vệ các thông tin trên khỏi sự truy cập, sử dụng trái phép cần phải kiểm soát được những việc chính sau: thông tin được tạo ra, lưu trữ và truy nhập như thế nào, ở đâu, bởi ai và vào thời điểm nào. Như vậy việc quản lý bản quyền số đang là bài toán không dễ dàng của nhà quản lý.

Trên thực tế, nhu cầu về bảo vệ quyền tác giả, quyền sở hữu cho các thông tin số là rất lớn. Phương pháp Thủy Vân số (digital watermarking) hứa hẹn là một phương pháp hiệu quả bởi vì nó cho phép chủ sở hữu nội dung số có thể nhúng và giấu những bằng chứng về bản quyền của mình, từ đó có thể xác định được quyền sở hữu,phát hiện ra việc sử dụng trái phép mà vẫn không làm ảnh hưởng đến nội dung của nội dung số. Với các tính chất đặc thù của mình Digital watermarking rất thích hợp với việc bảo vệ bản quyền tác giả.

Xuất phát từ yêu cầu của thực tế, nhằm mục đích hướng tới một phần nhiệm vụ nhiệm vụ bảo vệ bản quyền thông tin số, đồ án có tựa đề: “**Thuật toán thủy vân ảnh và áp dụng cho ảnh văn bản hành chính**.”

Em xin bày tỏ lòng biết ơn chân thành nhất đến cô Ts.Ngô Hoàng Huy đã tận tâm chỉ bảo và hướng dẫn tận tình trong suốt thời gian em thực hiện đề tài này!

Đây là một đề tài rất hay mang tính thiết thực cao. Em đã nghiên cứu và cố gắng hoàn thành bằng tất cả khả năng của mình.Tuy đã cố gắng hết sức song chắc chắn đề tài này không tránh khỏi những thiết sót. Em rất mong nhận được sự thông cảm và chỉ bảo tận tình của quý Thầy cô và các bạn.

# CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ ĐỂ TÀI

## 1.1. Lí do chọn đề tài

Ngày nay, cùng với sự phát triển của công nghệ thông tin là sự phát triển mạnh mẽ của các phương tiện dữ liệu số (như các bức ảnh, các audio, video...), việc đánh dấu bản quyền cho các dữ liệu này cũng ngày một cải tiến và nâng cao. Tuy nhiên, công nghệ thông tin phát triển giúp cho việc chỉnh sửa, sao chép các dữ liệu số trở lên dễ dàng điều này kéo theo một thực trạng là số lượng các bản sao chép bất hợp pháp của các dữ liệu số ngày một nhiều. Xác nhận bản quyền tác giả, phát hiện sự xuyên tạc thông tin là một nhu cầu thiết yếu nhằm bảo vệ bản quyền cho các hương tiện dữ liệu số. Kỹ thuật thủy vân số (Digital Watermarking) là một trong những giải pháp đưa ra để giải quyết vấn đề .

## 1.2. Mục tiêu của đề tài

- Nắm được các kiến thức về các kỹ thuật thủy vân để bảo vệ bản quyền ảnh số.

* Thử nghiệm kỹ thuật thủy vân DCT và áp dụng bảo vệ bản quyền cho văn bản hành chính.

## 1.3. Giới hạn và phạm vi của đề tài

Tập trung nghiên cứu các kỹ thuật thủy vân trên ảnh số. Ứng dụng xây dựng hệ thống nhúng và kiểm tra thủy vân dựa vào kỹ thuật thủy vân DCT nhằm xác thực nội dung thông tin và bảo về bản quyền ảnh văn bản hành chính.

## 1.4. Kết luận chương 1

**1.4.1 Nội dung thực hiện:**

* Tìm hiểu khái niệm về thủy vân ảnh số
* Sử dụng kỹ thuật thủy vân trong bảo vệ bản quyền ảnh văn bản hành chính, tìm hiểu một số định dạng ảnh,các phép biến đổi ảnh như (Fourier , Cosine , Biến đổi sóng rời rạc ) , thuật toán thủy vân trên miền DWT , thuật toán thủy vân trên miền DCT, thuật toán bit nhị phân ít ý nghĩa nhất.
* Áp dụng thuật toán thủy vân trên miền DCT thực hiện chương trình thủy vân và kiểm tra thủy vân cho ảnh văn bản hành chính

**1.4.2 Phương pháp nghiên cứu:**

* Nghiên cứu tài liệu trên sách vở, internet….

**1.4.3 Kết quả đạt được:**

* Nội dung
* Tìm hiểu một số kỹ thuật thủy vân bảo vệ bản quyền ảnh văn bản hành chính.
* Kết quả
* Bản báo cáo về các kỹ thuật thủy vân để bảo vệ bản quyền ảnh văn bản hành chính.
* Chương trình thử nghiệm kỹ thuật thủy vân DCT bảo vệ bản quyền ảnh văn bản hành chính.

## 1.5. Khái niệm thủy vân số

**1.5.1. Lịch sử thủy vân số**

Khái niệm thủy vân đã ra đời từ lâu. Năm 1282, thủy vân đã được các công nhân nhà máy giấy sử dụng ở Italia. Các tờ giấy sẽ mỏng hơn và có hoa văn trên đó. Điều này giúp các xưởng sản xuất giấy đánh dấu bản quyền trên tờ giấy của họ làm ra. Đến thế kỷ XVIII, thủy vân đã có nhiều ứng dụng ở châu Âu và Mỹ trong việc xác thực bản quyền hay chống tiền giả. Thuật ngữ thủy vân bắt nguồn từ một loại mực vô hình và chỉ hiện lên khi nhúng vào nước.



*Hình 1. 1: Thủy vân trên đồng dollar của Mỹ*

Thủy vân số (digital watermarking) là một công cụ giúp đánh dấu bản quyền hay những thông tin cần thiết vào tài liệu điện tử.

Thuật ngữ thủy vân số được cộng đồng thế giới chấp nhận rộng rãi vào đầu thập niên 1990. Khoảng năm 1995, sự quan tâm đến thủy vân bắt đầu phát triển nhanh. Năm 1996, hội thảo về che dấu thông lần đầu tiên đưa thủy vân vào phần trình nội dung chính. Đến năm 1999, SPIE đã tổ chức hội nghị đặc biệt về Bảo mật và thủy vân trên các nội dung đa phương tiện. Cũng trong khoảng thời gian này, một số tổ chức đã quan tâm đến kỹ thuật watermarking với những mức độ khác nhau. Chẳng hạn CPTWG thử nghiệm hệ thống thủy vân bảo vệ phim trên DVD. SDMI sử dụng thủy vân trong việc bảo vệ các đoạn nhạc. Hai dự án khác được liên minh châu Âu ủng hộ, VIVA và Talisman đã thử nghiệm sử dụng thủy vân để theo dõi phát sóng.

Vào cuối thập niên 1990, một số công ty đưa thủy vân vào thương trường, chẳng hạn các nhà phân phối nhạc trên internet sử dụng Liqid Audio áp dụng công nghệ của Verance Corporation. Trong lĩnh vực thủy vân ảnh, Photoshop đã tích hợp một bộ nhúng và bộ dò thủy vân tên là Digimarc.

**1.5.2. Quá trình nghiên cứu thủy vân số**

Thủy vân số được coi là ra đời từ năm 1954, với bằng sáng chế của Emile Hembrooke. Tuy nhiên, nghiên cứu thủy vân vẫn chưa được đặt ra như là một lĩnh vực nghiên cứu độc lập cho tới những năm 1980.

Tuy nhiên khái niệm thủy vân chỉ được hoàn thiện vào giữa những năm 90 của thế kỷ trước.

Những nghiên cứu đầu tiên về thủy vân đều tập trung vào nghiên cứu "thủy vân mù" (blind watermark). Thủy vân mù là thủy vân được nhúng mà không cần quan tâm tới nội dung của môi trường nhúng. Tương tự như vậy, các thuật toán tách thủy vân

mù đều độc lập với những thành phần dữ liệu không chứa thủy vân. Có thể ví thủy vân mù như chữ ký tay, nội dung của thủy vân không thay đổi với các môi trường nhúng khác nhau.

Vào năm 1999, đã có một sự thay đổi lớn diễn ra. Trong một bài báo đăng trên IEEE, Cox và các đồng nghiệp đã nhận ra, chất lượng thủy vân sẽ tốt hơn rất nhiều nếu như thủy vân có quan tâm tới nội dung của môi trường nhúng. Các thủy vân này được gọi là các thủy vân giàu (informed watermark), khi đó nội dung của thủy vân được hiểu là một hàm của nội dung môi trường nhúng. Có thể so sánh ý tưởng này với ý tưởng về chữ ký điện tử.

Đi xa hơn nữa, vào năm 2000, hai đội ngũ tác giả B.Chen, G.W.Wornell và J. Chou, Pradhan, Ghaoui, Ramchandran đã phát triển từ bài báo của M.Costa năm 1983 "Writing on diry paper" để phát triển một hướng nghiên cứu rất mới. Ý tưởng chính của Costa là, có hai loại nhiễu sẽ tác động lên nội dung bản tin truyền đi. Loại nhiễu thứ nhất, là loại nhiễu xảy ra tại bên gửi, do các tác vụ biến đổi và xử lý tài liệu. Loại nhiễu này có thể kiểm soát. Loại nhiễu thứ hai là loại nhiễu xảy ra trên đường truyền, và chúng ta không thể kiểm soát được chúng. Costa lý luận rằng, các thuật toán thủy vân trước đây chỉ cố gắng nhúng thủy vân vào trong loại nhiễu thứ nhất, cho nên dung lượng tin giấu được là rất nhỏ. Costa cũng đã chỉ ra, dung lượng tin cần giấu là độc lập với loại nhiễu thứ nhất. Do đó, nếu ta coi toàn bộ tài liệu số là nhiễu thứ nhất, chúng ta sẽ có một phương pháp để nhúng một lượng thông tin rất lớn vào tài liệu.

Thủy vân có một ứng dụng rất quan trọng là bảo vệ sự toàn vẹn của tài liệu và chống xuyên tạc. Để thỏa mãn được yêu cầu này của thủy vân, các nghiên cứu trước kia đều cố gắng áp dụng một mô hình tổng quát lên toàn bộ tài liệu. Tuy nhiên, vào năm 1995, Cox và các đồng nghiệp đã nhận ra, họ có thể sử dụng mô hình tri giác (perceptual model) để giảm dung lượng thủy vân cần giấu. Thay vì cố gắng áp dụng một mô hình tổng quát lên toàn bộ tài liệu, thực ra chỉ cần áp dụng thủy vân lên một số phần quan trọng của tài liệu mà thôi. Đây có thể coi là một dạng đặc biệt của mô hình thủy vân giàu, vì nội dung thủy vân cũng bị phụ thuộc vào tài liệu.

Như một chân lý của cuộc sống, luôn tồn tại sự thống nhất và đấu tranh giữa các mặt đối lập. Với sự ra đời của thủy vân, thì khoảng từ năm 1990 trở về sau, đã có nhiều nghiên cứu về tấn công cũng như chống tấn công đối với thủy vân. Những

nghiên cứu này đã thúc đẩy quá trình nghiên cứu thủy vân đạt được nhiều kết quả mới.

Thủy vân sử dụng công nghệ trải phổ (spread spectrum) được giới thiệu cùng thời điểm với mô hình tri giác, là một nỗ lực nhằm cân bằng giữa tính bền vững (robustness) và tính tin cậy (fidelity) của thủy vân số. Công nghệ trải phổ sẽ trải một băng tần hẹp vào một băng tần rông hơn, do đó tỷ lệ nhiễu trên mỗi tần số trở nên rất nhỏ. Phía bên người gửi sẽ tổng hợp lại các tín hiệu này, và lúc này nhiễu trở nên lớn. Công nghệ trải phổ là một hướng đi có nhiều triển vọng của kỹ thuật thủy vân.

Chất lượng tài liệu điện tử sau khi giấu tin phải không được thay đổi nhiều, để cho con người khó có thể nhận ra bằng các giác quan thông thường. Thủy vân số là một lĩnh vực nghiên cứu mới, có nhiều triển vọng. Những năm gần đây lĩnh vực này có được sự quan tâm đáng kể của các nhà nghiên cứu.

## 1.6. Các ứng dụng của thủy vân số

**Bảo vệ quyền sở hữu (Copyright Protection)**: Digital watermark có thể được dùng để bảo vệ quyền sở hữu đối với các sản phẩm digital media. Nội dung của các digital media này sẽ chứa thêm các thông tin về người sở hữu. Khi các digital media này được sử dụng bất hợp pháp thì ta có thể dùng bộ watermark detector để phát hiện.

**Chống sao chép bất hợp pháp (Copy Protection)**: Các sản phẩm có chứa digital watermark biểu hiện cho việc sản phẩm này không được sao chép, vì nếu sao chép sẽ phạm luật. Nhà sản xuất sẽ trang bị cho các phương tiện dùng để nhân bản (như CD writer…) khả năng phát hiện xem digital media có chứa thủy vân hay không, nếu có thì sẽ từ chối không sao chép. Một số phần mềm cũng có chức năng này, ví dụ như phần mềm xem phim của hãng DivX sẽ từ chối chiếu các bộ phim chứa thủy vân.

**Theo dõi quá trình sử dụng (Tracking)**: Digital watermarking có thể được dùng để theo dõi quá trình sử dụng của các digital media. Mỗi bản sao của sản phẩm được chứa bằng một watermarked duy nhất dùng để xác định người sử dụng là ai. Nếu có sự sao chép bất hợp pháp, ta có thể truy ra người vi phạm nhờ vào watermark được chứa bên trong digital media.

**Chống giả mạo (Tamper Proofing)**: Digital water marking có thể được dùng để chống sự giả mạo. Nếu có bất cứ sự thay đổi nào về nội dung của các digital media thì watermark này sẽ bị huỷ đi. Do đó rất khó làm giả các digital media có chứa watermark.

**Theo dõi truyền thông (Broadcast Monitoring):** Các công ty truyền thông và quảng cáo có thể dùng kỹ thuật digital watermarking để quản lý xem có bao nhiêu khách hàng đã dùng dịch vụ cung cấp.

**Truyền tin bí mật (Concealed Communication)**: bởi vì digital watermarking là một dạng đặc biệt của việc che dấu dữ liệu (steganography) nên người ta có thể dùng để truyền các thông tin bí mật.

## 1.7. Các đặc tính của thủy vân

Để đảm bảo được những khả năng trên, thì thủy vân phải thỏa mãn được các yêu cầu sau:

**Tính ẩn**: tính ẩn là khả năng khó bị nhận ra của thủy vân sau khi đã nhúng vào tài liệu điện tử, mà chủ yếu là các giác quan của con người. Nói cách khác, các tài liệu điện tử phải chịu ít sự thay đổi về mặt chất lượng khi nhúng thủy vân.

**Tính bền vững**: tính bền vững được hiểu tùy vào mục đích của từng loại thủy vân, ví dụ với thủy vân dùng để bảo vệ bản quyền, thì thủy vân phải bền với các phép tấn công hay biến đổi, trong khi với thủy vân dùng để chống xuyên tạc hoặc đảm bảo toàn vẹn dữ liệu, thì thủy vân phải bị phá hủy ngay khi có sự tác động hoặc tấn công.

**Tính bảo mật**: sau khi thủy vân số đã được nhúng vào tài liệu, thì yêu cầu chỉ cho những người có quyền mới có thể chỉnh sửa và phát hiện thủy vân.

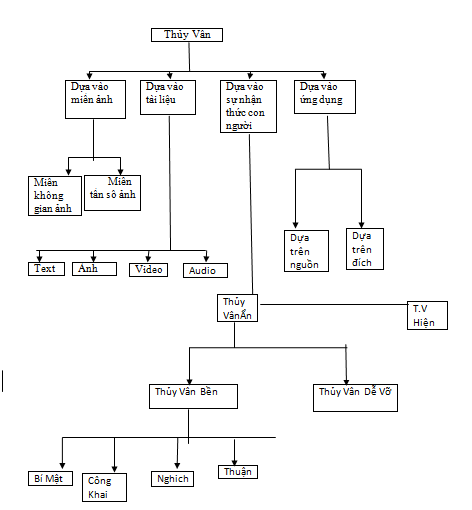
**Tính hiệu quả**: yêu cầu thuật toán thủy vân phải làm việc được với một vùng lớn các ảnh có thể.

**Dung lượng giấu**: thuật toán thủy vân cho phép giấu càng nhiều thông tin càng cao.

Tuy nhiên, các yêu cầu trên thường là trái ngược nhau, và người ta phải cân đối giữa các yêu cầu để phù hợp với từng bài toán cụ thể.

## 1.8. Phân loại thủy vân

Có nhiều phương pháp để phân loại thủy vân, dưới đây trình bày những phương pháp phân loại phổ biến nhất:



*Hình 1. 2: Một số phương pháp phân loại thủy vân tiêu biểu*

Dựa vào miền tác động, chúng ta có thể phân loại thủy vân thành tác động lên miền không gian ảnh (spatial domain) và tác động lên miền tần số ảnh (frequency domain).

Dựa vào kiểu tài liệu được nhúng thủy vân, chúng ta có thủy vân được nhúng vào ảnh, vào audio, video hay text.

Dựa vào tác động tới thị giác con người, chúng ta có thủy vân hiện (visible watermark) hoặc thủy vân ẩn (invisible watermark). Thủy vân ẩn lại được chia thành thủy vân bền (robust watermark) và thủy vân dễ vỡ (fraglie watermark).

Thủy vân hiện có ưu điểm là nhìn được bằng mắt thường, khiến cho tất cả người sử dụng đều biết được bản quyền của ảnh. Tuy nhiên, nó sẽ tác động tới chất lượng ảnh

## 1.9. Quy trình thực hiện thủy vân

Quy trình thực hiện thủy vân được trải qua bốn bước :

* Tạo thủy vân.
* Nhúng thủy vân.
* Tách thủy vân.
* Kiếm tra thủy vân.

**CHƯƠNG 2:** **KỸ THUẬT THỦY VÂN ẢNH VÀ THUẬT TOÁN NHÚNG THỦY VÂN**

## 2.1. Các toán tử tần số

**2.1.1. Phép biến đổi Fourier**

Biến đổi Fourier do nhà toán học Joseph Fourier (1768 - 1830) đưa ra. Ông đã chứng minh rằng bất cứ một hàm số tuần hoàn nào cũng có thể biến đổi thành tổng vô hạn hoặc hữu hạn các hàm sin (hoặc cos).

Do ảnh số chỉ là một phần của tín hiệu số, và bản chất tín hiệu ảnh số là tín hiệu rời rạc do quá trình lấy mẫu từ môi trường, nên ta sử dụng biến đổi Fourier rời rạc (Discrete Fourier Transform - DFT) trong xử lý ảnh.

Đồ án chỉ trình bày về phép biến đổi Fourier rời rạc với các tín hiệu hai chiều, tương ứng là các ảnh số, các kiến thức chuyên sâu hơn có thể tìm được ở các tài liệu tham khảo.

DFT hai chiều của một ảnh M x N : {u(m,n) } là một biến đổi tách được và được định nghĩa :

Biến đổi Fourier rời rạc của một ảnh M × N: {u(m,n)} được định nghĩa như sau:



với 0 <= l, k <= N-1

và biến đổi ngược:



với 0 <= m, n <= N-1

Như vậy, các hàm cơ sở trong ma trận ảnh cơ sở của biến đổi Fourier là các hàm cosine và hàm sine. Theo tính toán trên, ta thấy biến đổi Fourrier biểu diễn ảnh trong không gian mới theo các hàm sin và cos.

Như vậy, các hàm cơ sở trong ma trận ảnh cơ sở của biến đổi Fourier là các hàm cosine và hàm sine. Theo tính toán trên, ta thấy biến đổi Fourrier biểu diễn ảnh trong không gian mới theo các hàm sin và cos.

**2.1.2. Phép biến đổi Cosin rời rạc**

Biến đổi Cosin rời rạc viết tắt là DCT (Discrete Cosine Transform) do Ahmed và các đồng nghiệp của ông đưa ra vào năm 1974. Từ đó cho đến nay, nó được sử dụng rất phổ biến trong kỹ thuật xử lí ảnh số nói riêng và xử lí tín hiệu số nói chung. Trong các kỹ thuật thuỷ vân ảnh dựa trên phép biến đổi dữ liệu ảnh sang miền tần số thì phép biến đổi DCT được sử dụng nhiều nhất. Lí do ở đây là phép biến đổi DCT đã được dùng trong dạng chuẩn ảnh JPEG. Nếu áp dụng DCT thì cũng theo chuẩn của JPEG và do đó sẽ tránh được mất thuỷ vân do phép nén JPEG.

**Định nghĩa biến đổi cosine rời rạc hai chiều**

Biến đổi DCT hai chiều tổng quát là biến đổi trên khối hai chiều bất kì M×N, trong đó các khối kích thước 8×8, 16×16 được sử dụng nhiều nhất. Tuy nhiên, chúng ta sẽ chỉ tìm hiểu phép biến đổi DCT trên khối 8×8 được sử dụng trong chuẩn nén ảnh JPEG.

* Phép biến đổi thuận DCT 8\*8 được định nghĩa như sau:



I(u,v) được gọi là hệ số DCT và là số thực.

* Phép biến đổi ngược IDCT được định nghĩa như sau:



trong đó, 0 ≤ k, l, u, v ≤ 7 và ξ(u) = 

u=0

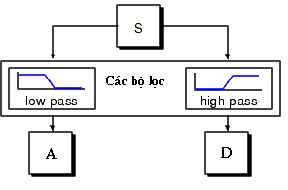
u>0

**2.1.3. Biến đổi sóng nhỏ rời rạc (Discrete Wavelet Transform-DWT)**

Đây là phép biến đổi mới nhất được áp dụng cho ảnh số. Ý tưởng của DWT cho tín hiệu một chiều như sau:

Tín hiệu được chia thành 2 phần, phần tần số cao và phần tần số thấp.Đối với nhiều tín hiệu, phần nội dung có tần số thấp chứa những thông tin quan trọng nhất - những đặc tính nhận dạng.Trái lại, phần nội dung có tần số cao chỉ mang sắc thái của tín hiệu.Bởi vậy, trong biến đổi sóng nhỏ, người ta thường nói về các phần gần đúng (approximations) và các phần tiểu tiết (details) của tín hiệu.Phần gần đúng là những thành phần có tỷ lệ cao (high-scale) nhưng có tần số thấp của tín hiệu.Phần tiểu tiết là những thành phần có tỷ lệ thấp nhưng có tần số cao của tín hiệu.

Để tách một tín hiệu thành hai phần như trên, người ta dùng hai bộ lọc bù nhau. Bộ lọc thứ nhất gọi là low-pass filter chỉ cho phép các thành phần có tần số thấp đi qua. Bộ lọc thứ hai gọi là high-pass filter chỉ cho phép các thành phần có tần số cao đi qua. Quá trình lọc ở mức cơ bản nhất có dạng như sau:



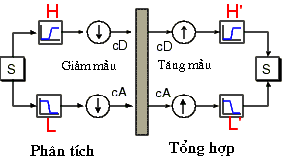
*Hình 2. 1: Lọc theo tần số thấp và cao*

Tín hiệu gốc S qua hai phép lọc được phân tích thành hai tín hiệu, A và D. Có một điều cần lưu ý là, nếu chúng ta thực hiện quá trình này trên một tín hiệu số thực sự thì sẽ thu được một số lượng dữ liệu gấp đôi số lượng ban đầu. Ví dụ, giả sử tín hiệu gốc S gồm 1000 mẫu dữ liệu thì mỗi tín hiệu được lọc ra cũng sẽ gồm 1000 mẫu. Tổng cộng sẽ có 2000 mẫu. Nhưng để đảm bảo có thể tổng hợp lại tín hiệu ban đầu, chúng ta thực sự chỉ cần sử dụng một nửa số tín hiệu được lọc. Do đó, nên có một qui trình giảm bớt một nửa số mẫu của A và D. Qui trình này được gọi là *dowsampling*. Sau khi giảm đi một nửa, chúng ta sẽ có hai tín hiệu tương ứng là cA và cD.Các dữ liệu của cA và cD được gọi là các *hệ số sóng nhỏ*.

Toàn bộ hai quá trình trên đây, *lọc* và *cắt giảm mẫu* được gọi là *phân tích* các tín hiệu hoặc *biến đổi sóng nhỏ rời rạc* (DWT). Đó là quá trình xuôi, còn ngược lại phải làm như thế nào để có thể lắp ghép các tín hiệu thu được thành tín hiệu gốc mà không bị mất thông tin. Quá trình này được gọi là *tổng hợp* các tín hiệu hoặc *biến đổi sóng nhỏ rời rạc ngược* (IDWT).

Để tổng hợp tín hiệu, người ta sẽ tăng số mẫu lên gấp đôi, được gọi là *upsampling* và sau đó lọc ngược trở lại. Các bộ lọc phân tích (L và H) và các bộ lọc tổng hợp (L’ và H’) cần phải được chọn lựa để thoả mãn một số điều kiện đảm bảo cho quá trình phân tích và tổng hợp là đồng nhất.Các bộ lọc này được gọi là đối xứng gương trực giao (quadrature mirror filters).

Quá trình phân tích và tổng hợp một tín hiệu có thể được mô tả qua hình vẽ dưới đây:



*Hình 2. 2: Quá trình phân tích và tổng hợp tín hiệu*

Trên đây chúng ta mới nói đến quá trình phân tích các tín hiệu thành hai tín hiệu xấp xỉ và chi tiết.Quá trình này được gọi là phân tích một mức.Có thể lặp lại quá trình này bằng cách tiếp tục phân tích tín hiệu xấp xỉ thành hai tín hiệu ở giải tần thấp hơn.Từ đây xuất hiện một khái niệm được gọi là phân tích nhiều mức.

Về mặt lý thuyết, quá trình này có thể lặp lại vô hạn.Tuy nhiên, trong thực tế, việc phân tích chỉ có thể tiếp diễn cho đến khi phần chi tiết chỉ còn một mẫu hay một điểm.Việc lựa chọn số mức phân tích hoàn toàn phụ thuộc vào bản chất của tín hiệu.

Tương tự như các tín hiệu một chiều, các tín hiệu hai chiều (tức là các bức ảnh) cũng có thể được phân tích tương tự bằng DWT nhưng theo cả hai chiều dọc và ngang. Một bộ lọc tần số thấp và một bộ lọc tần số cao sẽ được sử dụng kết hợp để lọc các tín hiệu. Đầu tiên, ảnh gốc sẽ được lọc thấp theo chiều dọc. Sau đó, nếu lọc lại bằng chính bộ lọc này theo chiều ngang ta sẽ có băng LL và nếu lọc lại bằng bộ lọc cao ta sẽ có băng LH. Tiếp theo, ảnh gốc sẽ được lọc cao theo chiều dọc và sau đó sẽ được lọc lại theo chiều ngang bằng lọc thấp để có băng HL và bằng lọc cao để có băng HH.

Quá trình trên lại có thể tiếp tục đối với băng LL để tạo ra phép biến đổi DWT hai mức như minh hoạ trong hình dưới đây:

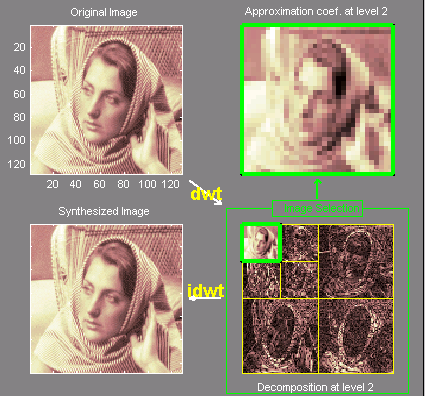
|  |  |  |
| --- | --- | --- |
| LL2 | HL2 | HL1 |
| LH2 | HH2 |
| LH1 | | HH1 |

*Hình 2. 3: Phân tích DWT hai mức*

Quá trình tổng hợp ảnh gốc từ các bức ảnh xấp xỉ và chi tiết cũng được làm tương tự như tổng hợp tín hiệu một chiều.Quá trình này bao gồm hai phần, đầu tiên là tăng mẫu và sau đó lọc tổng hợp.Nếu ảnh được phân tích thành nhiều mức thì sẽ tổng hợp từ mức cao lên mức thấp hơn.

Đến đây chúng ta đã có thể thấy ứng dụng của DWT trong thuỷ vân số ở chỗ nào. Hầu hết các lược đồ thuỷ vân đều tìm cách thay đổi các hệ số của một hoặc một số trong bốn băng tần để nhúng thuỷ vân trước khi tổng hợp.

Qua những trình bày trên đây, chúng ta thấy phân tích và tổng hợp tín hiệu bằng DWT bao gồm nhiều công đoạn và đòi hỏi rất nhiều tính toán phức tạp. Nhưng cũng rất may cho các nhà nghiên cứu trong lĩnh vực nén và thuỷ vân ảnh là tất cả những phức tạp này đã được MATLAB giải quyết giúp. MATLAB cho phép thực hiện các biến đổi DWT bằng một họ các hàm sóng nhỏ bao gồm hàm Harr, các hàm Daubechies và các biorthogonal.Ngoài ra, MATLAB còn giúp kiểm tra rất nhanh và chính xác các thực nghiệm về độ bền vững và chất lượng của thuỷ vân trên ảnh.Hình 10 dưới đây là mô tả công cụ phân tích và tổng hợp DWT của MATLAB.



*Hình 2. 4: Công cụ phân tích và tổng hợp DWT của MATLAB*

Biến đổi sóng có rất nhiều lợi thế so với các biến đổi khác, đó chính là:

* + Biến đổi sóng nhỏ là một mô tả đa độ phân giải của ảnh. Quá trình giải mã có thể được xử lý tuần tự từ độ phân giải thấp cho đến độ phân giải cao.
  + Biến đổi DWT gần gũi với hệ thống thị giác người hơn biến đổi DCT. Vì vậy, có thể nén với tỉ lệ cao bằng DWT mà sự biến đổi ảnh khó nhận thấy hơn nếu dùng DCT với tỉ lệ tương tự.

Các thuật toán thực hiện thủy vân hiện trên ảnh là tương đối dễ dàng, và đã được nghiên cứu nhiều trong môn xử lý ảnh số. Các yêu cầu về tính thẩm mỹ được đề cao. Tất nhiên cũng cần có yêu cầu là khó sử dụng các công cụ xử lý ảnh để loại bỏ thủyvân.

Như đã trình bày ở trên, các thuật toán thủy vân ẩn trên ảnh có thể chia ra làm hai loại chính, đó là thủy vân dựa vào biến đổi miền không gian ảnh, và biến đổi miền tần số ảnh.

## 2.2. Thuật toán thủy vân DCT trên miền tần số:

Thuật toán dưới đây sử dụng phương pháp trải phổ trong truyền thông để nhúng thuỷ vân. Giải tần được sử dụng để chứa tín hiệu thuỷ vân là miền tần số giữa của một khối biến đổi cô sin rời rạc DCT (discret cosne transformation) 8×8. Trong đó, các khối DCT 8×8 là những khối ảnh cùng kích thước đã được chọn ra ngẫu nhiên từ ảnh ban đầu và được áp phép biến đổi cosin rời rạc DCT để chuyển sang miền tần số. Mỗi tín hiệu thuỷ vân sẽ được chứa trong một khối.

**2.2.1. Mô tả thuật toán:**

* ***Input:*** 
  + Watermark: Một chuỗi các bit B
  + Một ảnh F
* ***Output:*** 
  + Một ảnh sau khi thuỷ vân, F’
  + Khoá để giải mã K

**2.2.2. Quá trình watermarking:**

* + Chia ảnh có kích thước m × n thành (m×n)/64 khối 8×8, mỗi bít sẽ được giấu trong một khối.
  + Chọn một khối bất kì B và biến đổi DCT khối đó thu được B’
  + Chọn hai hệ số ở vị trí bất kì trong miền tần số ở giữa của khối DCT, giả sử đó là b’(i,j) và b’(p,q).

Ta tính:

d = || b’(i,j)|- |b’(p,q)|| mod a

trong đó a là một tham số thoả mãn điều kiện: a=2(2t+1), t là một số nguyên dương.

Bít si sẽ được nhúng sao cho thoả mãn điều kiện sau:

− Nếu d<2t+1 và si = 1 thì một trong hai hệ số DCT b’(i,j) hoặc b’(p,q) có trị tuyệt đối lớn hơn sẽ bị thay đổi để d>=2t +1 theo công thức sau:

max(|b’(i,j)|, |b’(p,q)|) + (INT(0,75 \*a) - d)

Với hàm max(|b’(i,j)|, |b’(p,q)|) là hàm chọn ra hệ số có trị tuyệt đối lớn hơn, hệ số được chọn sẽ được cộng thêm một lượng là (INT(0,75 \*a) - d).

Hoặc cũng có thể biến đổi một trong hai hệ số theo công thức:

min(|b’(i,j)|, |b’(p,q)|) - (INT(0,25 \*a) + d)

Với hàm min(|b’(i,j)|, |b’(p,q)|) là hàm chọn ra hệ số có trị tuyệt đối nhỏ hơn, hệ số được chọn sẽ bị trừ đi một lượng là (INT(0,25 \*a) + d)

INT() là hàm làm lấy phần nguyên của một số thực.

− Tương tự, nếu d >= 2t+1 và si = 0 thì một trong hai hệ số DCT b’(i,j) hoặc b’(p,q) có trị tuyệt đối lớn hơn sẽ được thay đổi để thoả mãn d<2t+1 như sau:

max(|b’(i,j)|,|b’(p,q)|) - (d - INT(0,25\*a))

Hàm max(|b’(i,j)|, |b’(p,q)|) là hàm chọn ra hệ số có trị tuyệt đối lớn hơn, hệ số được chọn sẽ bị trừ đi một lượng là (d - INT(0,25 \*a))

Hoặc

min(|b’(i,j))|,|b’(p,q)|) + INT(1,25\*a) – d

*Hình 3. 5: Quá trình nhúng thủy vân*

**2.2.3. Quá trình giải nhúng để lấy lại thông tin:**

Đọc khối DCT từ ảnh chứa thuỷ vân và vị trí hai hệ số đã biến đổi, sau đó tính:

Ảnh gốc

{ Các khối ảnh }

{Các khối DCT}

Watermark

{Các khối đã nhúng thuỷ vân }

{ Các khối đã biến đổi ngược IDCT }

Ảnh đã nhúng thuỷ vân

tách khối

Biến đổi DCT

Nhúng thuỷ vân

Biến đổi IDCT

Ghép ảnh

d = ||b’(i,j)|-|b’(p,q)|| mod a với (a =2(2t+1))

Nếu d >= 2t+1 thì gán si =1

Nếu d< 2t+1 thì gán si =0

Ảnh đã nhúng thuỷ vân

Lấy ra các khối đã chọn

Khối đã biến đổi DCT

Watermarks

tách khối

biến đổi DCT

Giải mã

*Hình 2. 6: Quá trình giải mã*

**2.2.4. Chứng minh tính đúng đắn của thuật toán:**

Xét các trường hợp sau đây:

− Nếu d < 2t +1 với si = 0 và d >= 2t+1 với si =1 thì sẽ không thay đổi gì hệ số của khối DCT, và vì DCT là phép biến đổi thuận nghịch nên khi giải mã thì ta cũng thu được kết quả chính xác.

− Nếu d < 2t +1 với si = 1

Ta biến đổi một trong hai hệ số b’(i,j) và b’(p,q) như sau:

max(|b’(i,j)|,|b’(p,q)|) + (INT(0,75\*a) -d)

Khi đó giá trị d mới là:

d = (|| b(i, j) |-| b(p,q) || + (INT(0.75\* a) - d)) mod a

⬄ d’ = (||b’(i,j)|- |b’(p,q)|| mod a)+ (INT(0,75\*a) mod a) - (d mod a)

⬄ d’ = d + INT(0,75\*a) - d = INT(0,75 \*a) > 0,5 \*a = 2t +1 (dpcm)

Hoặc ta sử dụng cách biến đổi hai hệ số theo kiểu khác:

min(|b’(i,j)| - |b’(p,q)|) - (INT(0,25 \*a) +d))

Tính lại d:

d’ = (||b’(i,j)|-|b’(p,q)|| - (INT(0,25\*a) +d)) mod a

⬄ d’ = (||b’(i,j)|- |b’(p,q)|| mod a) - (INT(0,25\*a) mod a)

- (d mod a)

⬄ d’= d-(INT(0,25\*a) mod a) – d

= - INT(0,25\*a) mod a = INT(0,75\*a) >2t +1

− Trường hợp d>=2t +1 và si = 0

Ta sẽ biến đổi một trong hai hệ số DCT b’(i,j) hoặc b’(p,q) như sau: max(||b’(i,j)|,|b’(p,q)||) - (d- INT(0,25\*a))

Giá trị mới của d sẽ là:

d’ = (||b’(i,j)|-|b’(p,q)|| - (d-(INT(0,25\*a)) mod a

⬄ d’= ((||b’(i,j)|-|b’(p,q)||) mod a)-(d mod a) + (INT(0,25\*a) mod a)

⬄ d’= d-d + 0,25\*a = 0,25\*a < 0,5\*a = 2t +1(dpcm)

Hoặc ta sử dụng cách biến đổi khác đối với hai hệ số DCT:

min(||b’(i,j)|- |b’(p,q)||) + INT(1,25\*a) – d

Khi đó tính lại d ta được:

d’ = (||b’(i,j)|-|b’(p,q)|| + INT(1,25\*a) -d) mod a

⬄ d’= (||b’(i,j))|-|b’(p,q)|| mod a) +(INT(1,25\*a) mod a) - (d mod a)

⬄ d’ = d +INT(0,25\*a) -d = INT(0,25\*a) < 0,5\*a = 2t+1

Vậy với các phép biến đổi trên, ta luôn thoả mãn được điều kiện giấu tin.

**2.3. Một số thuật toán khác:**

**2.3.1. Thuật toán giấu thủy vân vào các bit ít ý nghĩa nhất:**

Một ý tưởng tự nhiên của thủy vân với ảnh số, cũng giấu như giấu tin, đó là sẽ sử dụng các bit có trọng số thấp (Least Significant Bit - LSB) để giấu thủy vân.

Các bit có trọng số thấp được hiểu là các bit mà nếu thay đổi giá trị của chúng sẽ ít làm thay đổi đến chất lượng ảnh.

Ví dụ, với ảnh bitmap 256 màu, màu của mỗi điểm ảnh được biểu diễn bằng 8 bit, nếu ta thay đổi giá trị bit thứ tám của mã màu, thì mã màu cũng chỉ thay đổi giá trị có 1 đơn vị, nên nhìn chung thì cả bức ảnh không bị ảnh hưởng nhiều.

Ta có thể minh họa thuật toán như sau: Xét thủy vân là chuỗi bit 0111.

Xét bức ảnh là chuỗi bit: 11001101 11000001 11110000 11110010.

Để nhúng thủy vân vào bức ảnh, ta sẽ chia bức ảnh thành các khối 8 bit, và đặt giá trị bit cuối cùng của khối bằng giá trị của bit thủy vân tương ứng.

Với minh họa trên, chúng ta có bức ảnh sau khi nhúng thủy vân là:

11001100 11000001 11110001 11110011

Để tách thủy vân, đơn giản ta chỉ làm ngược lại quy trình trên, tức là tách ra các bit cuối của từng khối 8 bit, ta sẽ thu được thủy vân ban đầu.

Muốn tăng tính an toàn của hệ thống, có thể nhúng liên tiếp thủy vân vào các khối 8 bit liền nhau, bởi thường thì dung lượng bức ảnh sẽ lớn hơn nhiều lần so với độ dài của thủy vân.

Ưu điểm của thuật toán trên là đơn giản, và dung lượng giấu cao.

Tuy nhiên, nhược điểm là do quá đơn giản nên rất dễ bị tấn công. Kẻ tấn công chỉ cần thay đổi ngẫu nhiên giá trị của các bit có trọng số thấp là thủy vân đã bị phá hủy.

* **Mô tả thuật toán**

Cách tiếp cận này khá đơn giản. Để nhúng thuỷ vân vào ảnh chủ, trước hết cần phải trải rộng ảnh thuỷ vân sang một miền ảnh có cỡ bằng cỡ của ảnh chủ. Sau đó thay thế các bit ít ý nghĩa nhất của ảnh chủ bằng các bit của thuỷ vân mở rộng. Việc chỉ thay thế các bit ít ý nghĩa nhất trong cách tiếp cận này đã đảm bảo chất lượng của ảnh thuỷ vân. Tuy nhiên, độ bền vững của thuỷ vân sẽ không đảm bảo đối với các phép biến đổi ảnh có sử dụng đến các bit ít ý nghĩa nhất.

* **Input:**
* **Một ảnh nhỏ được sử dụng làm thuỷ vân**
* **Một ảnh F cần nhúng thuỷ vân để bảo vệ bản quyền**
* **Output:**
* **Một ảnh sau khi thuỷ vân, F’**
* **Qui trình nhúng thuỷ vân**

Đọc ảnh chủ

Đọc ảnh thuỷ vân

Trải ảnh thuỷ vân trên một miền ảnh   
có cỡ bằng cỡ của ảnh chủ

Thay thế LSB của ảnh chủ bằng   
các bit của thuỷ vân trải rộng

*Hình 2. 7: Quy trình nhúng thủy vân*

* **Qui trình tách thuỷ vân**

Đọc ảnh chủ đã thuỷ vân

Lấy các LSB của ảnh chủ đã thuỷ vân

Thu nhỏ miền ảnh của các LSB để có ảnh thuỷ vân

*Hình 2. 8: Quy trình tách thủy vân*

**2.3.2. Thuật toán thủy vân ảnh trên miền DWT:**

* **Mô tả thuật toán**

**Input:**

* Một ảnh gốc sử dụng làm thuỷ vân
* Một ảnh F cần nhúng thuỷ vân để bảo vệ bản quyền
* Một khoá bí mật để nhúng và giải thuỷ vân (khóa k)

**Output:**

* Một ảnh sau khi thuỷ vân, F’
* **Quy trình nhúng thủy vân:**

Đọc ảnh chủ  
Đọc ảnh thuỷ vân

Phân tích DWT mức 1 cho ảnh chủ

Đặt khoá K làm trạng thái cho bộ sinh chuỗi giả ngẫu nhiên

Tại những điểm thuỷ vân bằng 0  
cộng thêm cho giá trị tương ứng của băng cH1 và cH2 một số giả ngẫu nhiên tương ứng trong hai dãy

Tổng hợp ảnh chủ bằng IDWT từ các băng cA1, cD1 và các băng đã nhúng thuỷ vân cH1 và cH2   
Ghi ảnh đã thuỷ vân lên đĩa

*Hình 2. 9: Quy trình nhúng thủy vân*

* **Quy trình tìm lại thủy vân:**

Đọc ảnh chủ đã thuỷ vân  
Đọc ảnh thuỷ vân gốc

Phân tích DWT mức 1 cho ảnh chủ đã thuỷ vân

Đặt khoá K làm trạng thái cho bộ sinh chuỗi giả ngẫu nhiên

Tính các hệ số đồng tương quan giữa các băng   
cH1, cH2 với các dãy số giả ngẫu nhiên

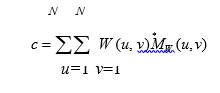
Xây dựng vector thuỷ vân thông qua   
các hệ số đồng tương quan

Hiệu chỉnh và hiện ảnh thuỷ vân đã tìm thấy

*Hình 2. 10: Quy trình tìm lại thủy vân*

**2.3.3. Thuật toán phép biến đổi Fourier rời rạc:**

Trong cả hai trường hợp trên, sự hiện diện của thủy vân được phát hiện bằng cách sử dụng sự tương quan



Trong đó N×N là kích thước của hình ảnh, M \* (u, v) là độ lớn của hệ số DFT của ảnh thủy vân.Nguyên tắc xác định sự tồn tại của thủy vân như sau:

*W*

H0: hình ảnh được nhúng thủy vân W nếu c >T

H1: hình ảnh không được nhúng thủy vân W nếu c < T

Ngưỡng T tính theo công thức

giá trị kỳ vọng của hàm mật độ xác suất Gauss liên quan đến các giả thuyết H0 và H1 tương ứng.

Tác giả đã mở rộng ý tưởng thủy vân bằng cách chèn hai thủy vân hình vòng trong miền DFT. Mặc dù việc chèn thủy vân hình vòng đơn giản và thuận tiện, độ lớn sửa đổi của các hệ số rõ ràng trong miền biến đổi nhưng lại đưa ra các thông tin có giá trị về vị trí tấn công của thủy vân

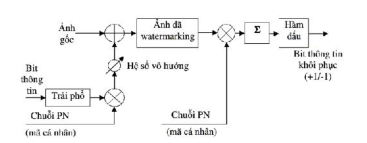
**Mô tả thuật toán**

* **Quá trình nhúng thủy vân**
* Tính toán DFT của ảnh chủ kích thước N×N
* Có được độ lớn của hệ số DFT
* Phân chia ma trận kích thước N×N thành 4 ma trận MUL, MUR, MLL, MLR
* có kích thước (N/2)×(N/2)
* Nhúng vào ma trận phía trên trái MUL
* Xác định dải tần số cao và dải tần số thấp dựa trên các hệ số DFT
* Chọn một khối có kích thước *k* và một chuỗi *PRN* chiều dài *m*, tạo thành khối có kích thước *k*×*m*
* Tạo ra thủy vân *W* bằng cách đặt một thành phần của chuỗi *PRN* trong mỗi khối băng tần, vị trí trong mỗi khối lựa chọn ngẫu nhiên
* Chọn hệ số co giãn *l*  cho các băng tần số thấp và giá trị lớn hơn hệ số co giãn *h* cho băng tần số cao.
* Thêm *W* vào ma trận *M*UL
* Sao chép các hệ số điều chỉnh ở ma trận phía trên trái MUL vào phía dưới phải MLR ma trận
* Nhúng vào ma trận phía trên phải MUR
* Sử dụng các dải tần
* Sử dụng một khối có kích thước như trong bước 4(b), và một chuỗi *PRN* khác có chiều dài tương tự như chuỗi *PRN* ở bước 4(b)
* Tạo một ma trận thủy vân *W* bằng cách đặt một thành phần của chuỗi *PRN* trong mỗi khối băng tần. Vị trí trong mỗi khối được lựa chọn ngẫu nhiên.
* Chọn hệ số co giãn *l* cho các băng tần số thấp và chọn hệ số co giãn *h* cho băng tần số cao
* Thêm *W* vào ma trận *M*UR
  1. Sao chép hệ số biến đổi ma trận trên phải *M*UR cho ma trận thấp trái *M*LL.
  2. Tính được các hệ số DFT của hình ảnh bằng cách lấy giá trị trong các ma trận *M*UL, *M*UR, *M*LL, *M*LR và các góc tương ứng.

Áp dụng DFT ngược để có được hình ảnh thủy vân

* **Quy trình tách thủy vân**

**



Tiêu chuẩn dùng để đánh giá chất lượng của ảnh được thuỷ vân và các ảnh bị tấn công dựa vào tỷ số giữa giá trị cực đại của tín hiệu và độ lớn của nhiễu nền PSNR (Peak-Signal-To-Noise Ratio). Tỷ số này được xác định bằng công thức PSNR = 20log10(255/RMSE), trong đó RMSE () là căn bậc hai của sai số trung bình bình phương giữa ảnh gốc và ảnh bị tác động. Trong kỹ thuật nén ảnh, người ta thường chấp nhận tỷ số này trong khoảng từ 20 đến 40 dB.

Thuỷ vân được nhúng sau khi giải mã sẽ được so sánh để kiểm định, chứng thực thuỷ vân. Có những thuỷ vân nhìn thấy được và mang ý nghĩa nhận biết thì công việc trở nên quá đơn giản chẳng hạn như thuỷ vân là một chuỗi mã kĩ tự ASCII mang thông tin nào đó như tên tác giả, ngày tháng… thì khi giải mã ta cũng dễ dàng nhận biêt thông tin. Hay như thuỷ vân là một ảnh nào đó chẳng hạn thì giải mã ra cũng được một cái ảnh tượng tự và ta có thể nhìn thấy sự khác biệt giữa hai ảnh.

Chất lượng các thuỷ vân lấy ra từ những ảnh bị tấn công được đo bằng Tỷ số Tương tự (Similarity Ratio) SR = S/(S+D), trong đó S là số giá trị điểm ảnh giống nhau, và D là số giá trị điểm ảnh khác nhau giữa thuỷ vân gốc và thuỷ vân lấy ra từ các ảnh bị tấn công. Chẳng hạn, nếu ta nhúng một thuỷ vân có độ dài là 1000 bit, khi giải mã thuỷ vân bị sai lệch mất 1 bit so với ban đầu vậy thì tỷ số tương tự sẽ là 999/1000 = 99.9% .

Tuy nhiên, trong một số trường hợp thuỷ vân là một chuỗi số ngẫu nhiên thì công việc nhận diện thuỷ vân sẽ không đơn giản. Trong trường hợp này người ta sẽ đưa ra một ngưỡng để xác định một ảnh có giấu thuỷ vân nào đó hay không.

**2.3.4. Thủy vân dễ vỡ:**

Thủy vân dễ vỡ (fragile watermark) là một dạng thủy vân đặc biệt, được sử dụng để bảo đảm tính toàn vẹn thông tin của ảnh số. Đặc điểm của thủy vân dễ vỡ là, chỉ cần ảnh bị thay đổi thì thủy vân sẽ bị phá hủy, do đó chống lại sự xuyên tạc nội dung của ảnh.

Thuật toán thủy vân dễ vỡ đầu tiên được Yeung và Mintzer đưa ra vào năm 1997. Thuật toán chỉ hoạt động trên các ảnh xám, với thủy vân là một chuỗi bit. Phía bên người nhận sẽ đọc từng điểm ảnh, trích xuất thủy vân và so nó với thủy vân được công bố. Nếu có sự khác biệt, thì độ xám của điểm ảnh sẽ được thay đổi tới khi hai thủy vân thu được là giống nhau. Thuật toán của Yeung và Mintzer làm việc và bảo toàn tính toàn vẹn dữ liệu cho từng điểm ảnh, cũng như hỗ trợ khả năng khôi phục ảnh gốc.

Một thuật toán khác, được Wong Ping Wah đưa ra năm 1998, kết hợp giữa thủy vân số và mật mã hóa công khai. Trong thuật toán của Wong, mức xám của LSB trong ảnh gốc sẽ được đặt bằng 0. Sau đó, ảnh gốc sẽ được chia thành các khối (block) có kích thước bằng kích thước thủy vân. Kích thước của ảnh cùng với mỗi khối đó sẽ được băm, và kết quả thu được sẽ được XOR với thủy vân. Kết quả của phép XOR sẽ được mã hóa bằng hệ mã hóa RSA, sau đó nhúng vào LSB của ảnh gốc.

Phía bên nhận, sẽ làm công việc ngược lại. Đầu tiên, ảnh cũng sẽ lại được chia thành các khối, và thông tin nhúng trong LSB sẽ được thu hồi và giải mã cũng bằng hệ mã hóa RSA. Cùng với đó, bên nhận cũng sẽ băm các khối của ảnh thu được cùng kích thước ảnh. Hai kết quả đó được XOR với nhau để thu lại thủy vân, và so sánh thủy vân thu được với thủy vân gốc được lưu trong cơ sở dữ liệu của bên gửi.

Cả hai thuật toán trên đều có nhược điểm, là không tận dụng sự tương quan giữa các khối ảnh liền nhau, và dung lượng giấu tin thấp.

Để cải tiến nhược điểm này, năm 2003, hai tác giả Li Chang Tsun và Yang Fong Man ở Đại học Havard đã đề xuất một thuật toán thủy vân mới. Thuật toán này sẽ thay đổi mức xám của mỗi điểm ảnh đi một lượng nhất định, lượng này phụ thuộc vào bản thân thủy vân và một điểm ảnh lân cận của điểm ảnh đang xét. Do điểm ảnh lân cận này là bí mật, cho nên sẽ làm tăng độ an toàn của thủy vân. Thuật toán của Li và Yang là thuật toán thủy vân miền không gian ảnh.

Cũng trong năm 2003, ba tác giả Hsieh Tsung Han, Li Chang Tsun và Wang Shuo đã đề xuất một thuật toán thủy vân trên miền tần số.

Ảnh gốc X được DCT và chia thành các khối 8x8. Một ảnh nhị phân A có kích thước bằng X được tạo ra từ khóa bí mật. Ảnh nhị phân B được tạo ra theo luật: tất cả các điểm ảnh tương ứng với các hệ số khác 0 thì có giá trị 1, ngược lại có giá trị 0. Thủy vân W được tạo ra bằng cách X = A XOR B. W cũng được chia thành các khối

8x8. Sau khi DCT mỗi khối Xi, bốn hệ số Xi(h), Xi(h - 1), Xi(h - 2) và Xi(h - 3) thỏa điều kiện là tần số thấp hơn hoặc bằng tần số giữa h sẽ được chọn là các hệ số thủy vân. Bốn hệ số này được điều chỉnh sao cho phương trình sau được thỏa mãn:

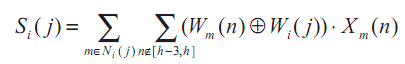
P (Si(j) \* Xi(j)) = Wi(j)

Trong đó:

P() là hàm số, trả về giá trị 1 nếu số lượng bit 1 trong tham số là lẻ, và bằng 0 nếu ngược lại.

\* là phép toán nối Xi(j) và Si(j).

Si(j) là tổng của các hệ số thủy vân không âm và Wi(j) thuộc tập Ni(j) – là tập chứa DCT của Xi và 8 khối xung quanh.



Quá trình được lặp lại cho tới khi tất cả các khối được thủy vân. Quá trình trích xuất thủy vân được thực hiện theo quy trình ngược lại.

***3.3.5. Thủy vân giòn:***

Thủy vân dễ vỡ có đặc điểm là rất nhạy cảm với thay đổi ảnh, tức chỉ cần có một thay đổi nhỏ trên ảnh là thủy vân đã bị phá hủy. Điều này không thích hợp với thực tế, khi ảnh truyền trên Internet có thể bị xử lý như nén, xoay chiều mà không làm thay đổi nội dung. Thủy vân giòn (semi-fragile watermark) ra đời nhằm cung cấp một khả năng tùy biến tốt hơn cho thủy vân ảnh. Thủy vân giòn chỉ bị hủy khi ảnh có sự thay đổi rất lớn, làm biến dạng ảnh.

Năm 1999, Kundur và Hatzinakos đã đưa ra một thuật toán thủy vân giòn trên ảnh sử dụng DWT.

Đầu tiên, hai tác giả định nghĩa hàm lượng tử hóa Q(f) như sau: Q(f) = 0, nếu [f/a.2l] chẵn.

Q(f) = 1 trong trường hợp ngược lại.

Với hàm [] là hàm lấy phần nguyên của một số. f là một hệ số, l là cấp DWT. Thực hiện DWT L cấp đối với ảnh. Trừ hệ số của băng tần có tần số thấp nhất, mỗi hệ số f(i) phải thỏa mãn phương trình sau, với qkey(i) là một biến boolean, phụ thuộc vào các lân cận của điểm ảnh i:

Q(f(i)) = w(i) XOR qkey(i)

Nếu phương trình trên không được thỏa, cần biến đổi f(i) như sau:

f(i) = f(i) – a.2l, nếu f(i) > 0

f(i) = f(i) + a.2l, nếu f(i) ≤ 0

Sau khi quá trình nhúng thủy vân kết thúc, thực hiện IDWT để thu lại ảnh đã nhúng thủy vân.

Ở phía nhận, sẽ thu lại thủy vân bằng công thức sau:

we(i) = Q (f(i)) XOR qkey(i)

Hệ số thay đổi ảnh (Tamper Assesment Function - TAF) được định nghĩa là:

TAF (we, w) = 1/Nw. ∑ (w(i) XOR we(i))

TAF đặc trưng cho sự biến đổi nhiều hay ít của ảnh, và quyết định có chấp nhận sự thay đổi đó hay không là tùy vào người dùng

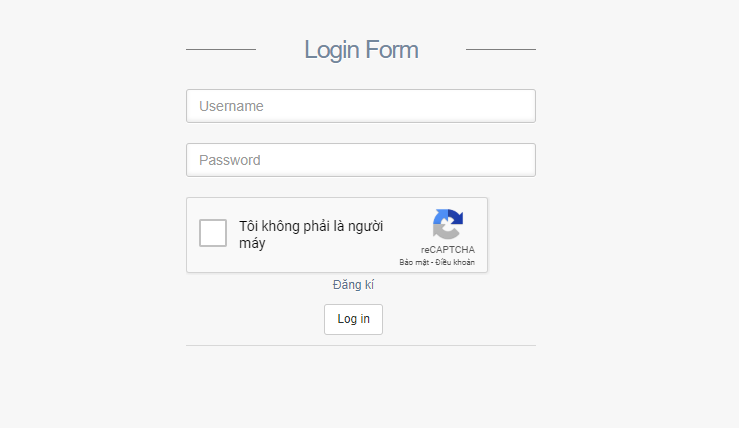
# CHƯƠNG 3: CÀI ĐẶT VÀ ĐÁNH GIÁ HỆ THỐNG

**3.1. Cài đặt hệ thống:**

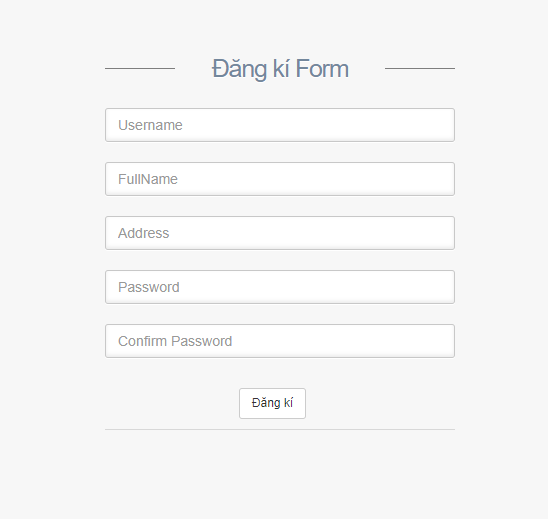
Phần mềm được viết bằng ngôn ngữ C# trên visual studio .

**Giao diện chương trình:**

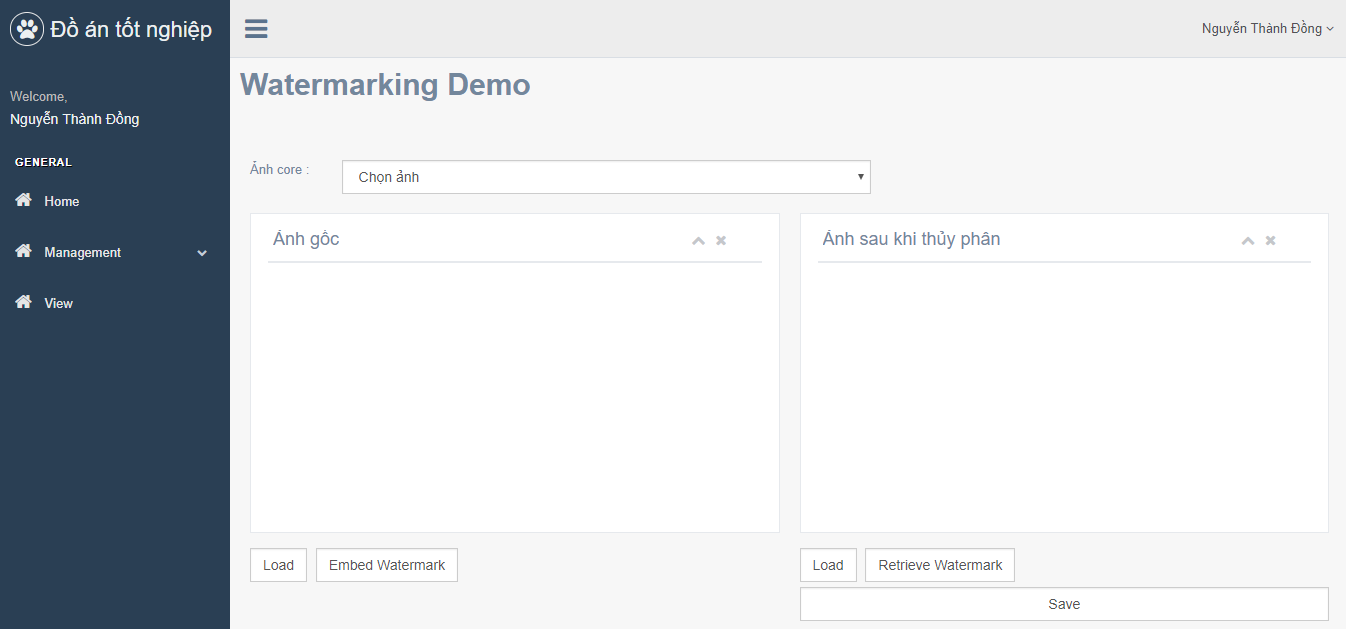
* Giao diện login:



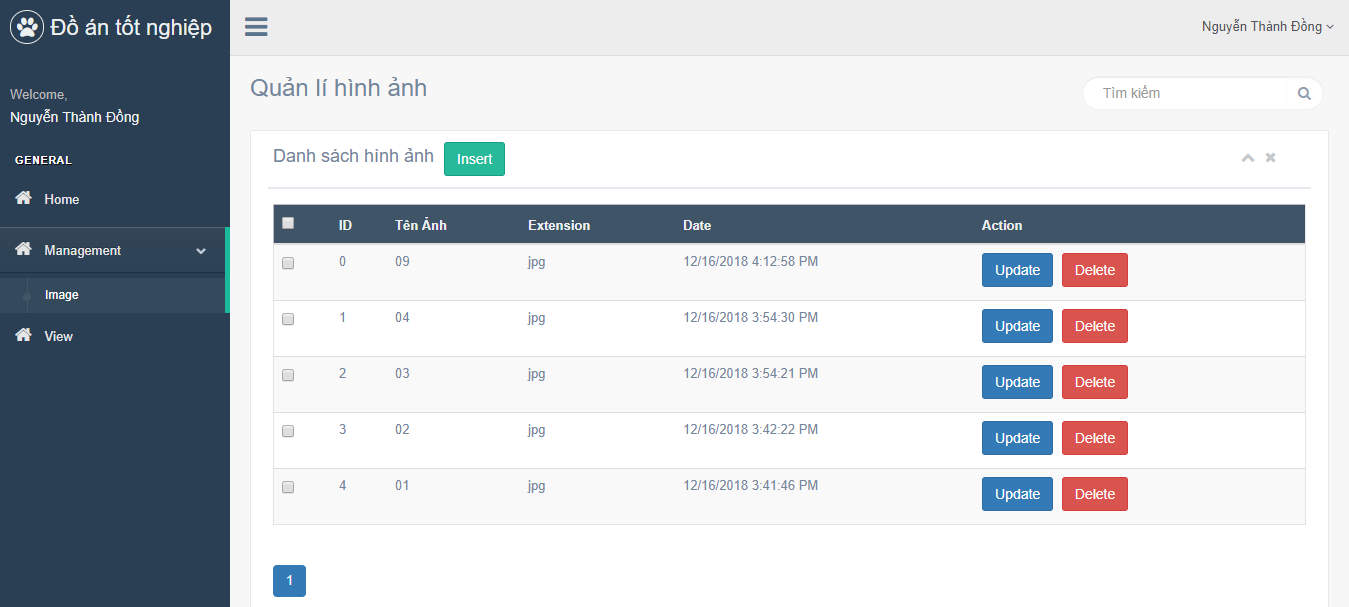
* Giao diện đăng kí:



* Giao diện chính là phần thủy phân và và tách thủy vân:



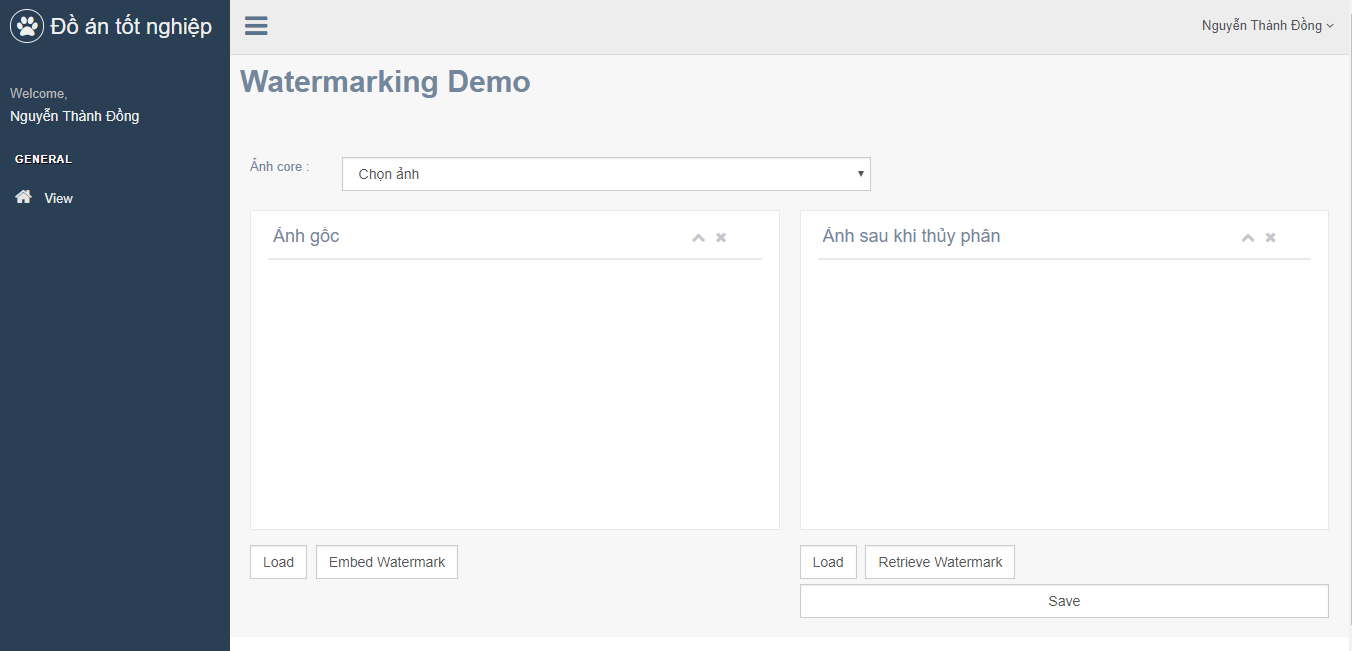
* Giao diện quản lý mã thủy vân



* Giao diện view ảnh đã xử lý thủy vân

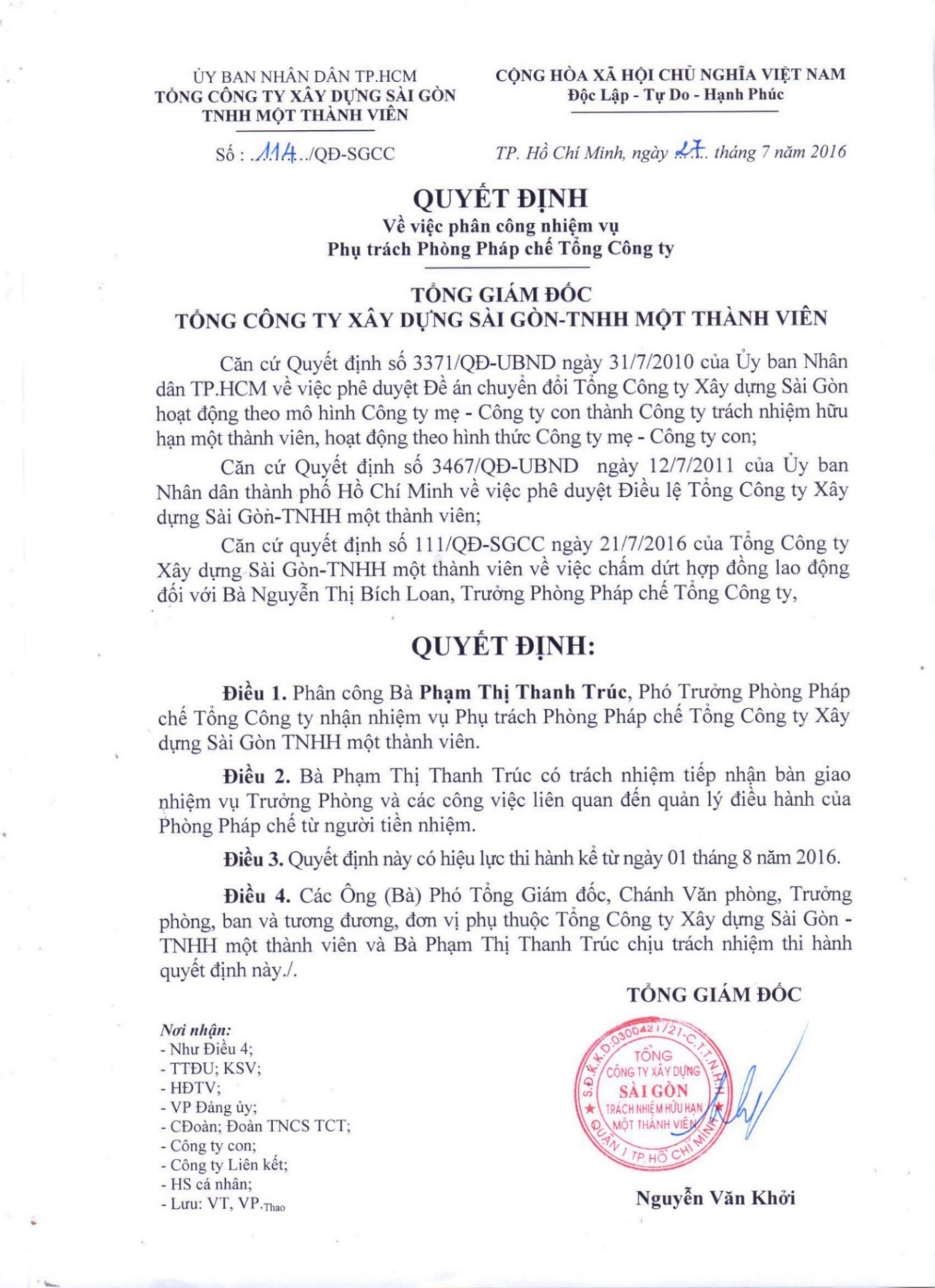


* Giao diện người dùng thường, chỉ có chức năng view ảnh:

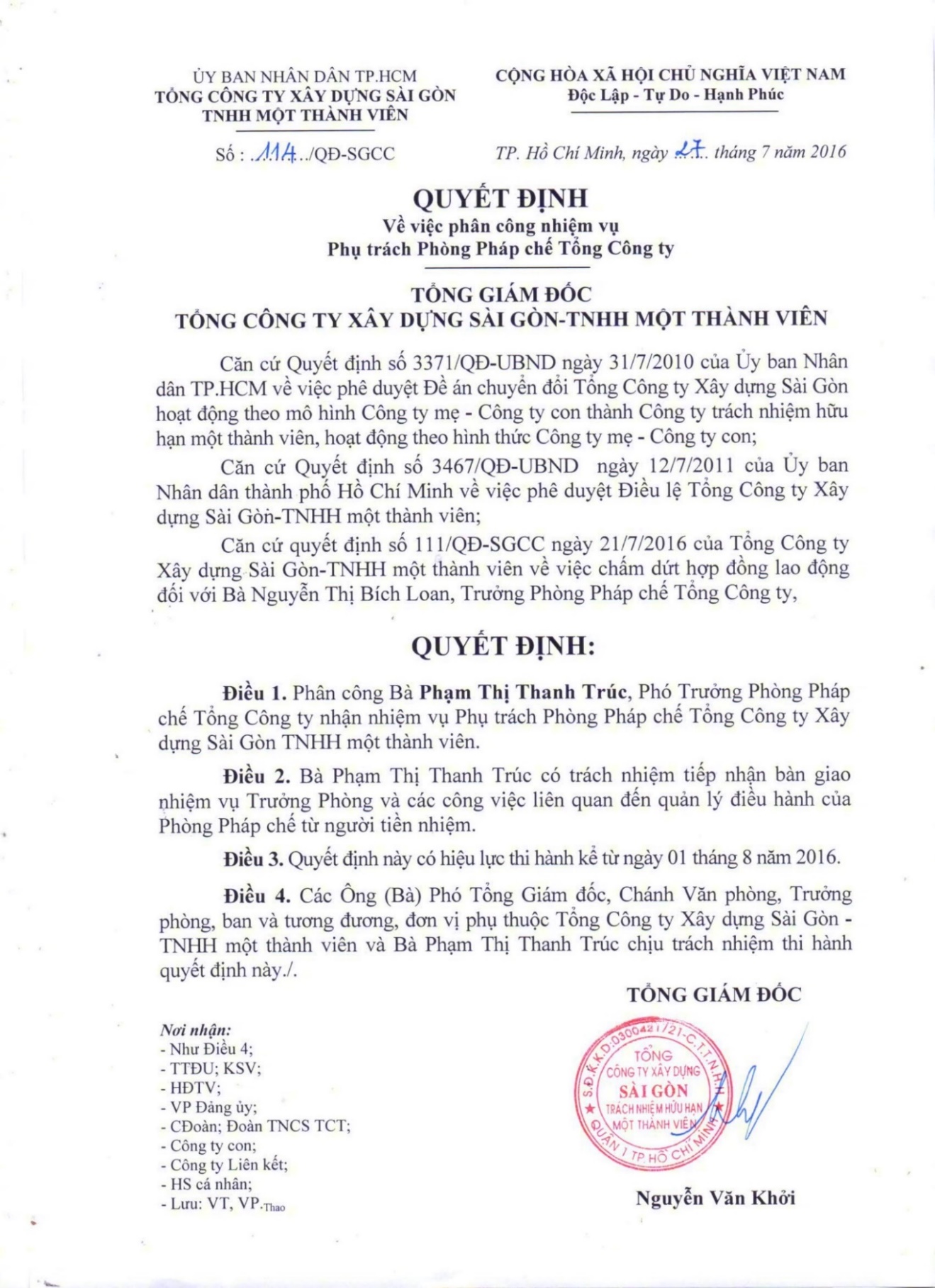


**3.2. Kết quả thực nghiệm:**

Ảnh ban đầu:



Ảnh sau khi xử lý:



Mắt thường gần như không thể phát hiện ra điểm khác nhau giữa 2 bức ảnh.

Tách thủy phân ra khỏi ảnh:



**3.3. Đánh giá, so sánh kết quả thực nghiệm:**

- Thực nghiệm có kết quả khá giống với nghiên cứu.

- Kết quả kiểm tra watermarking có tỉ lệ chính xác khá cao.

**Kết luận và hướng nghiên cứu tương lai**

* **Kết quả đạt được:**
* Rèn luyện được kỹ năng làm việc.
* Chuyển đổi code từ WDF sang nền web, kế thừa thuật toán thủy vân ảnh trên miền DCT.
* Phát triển thêm quản lý mã thủy vân và phân quyền tài khoản.
* **Hạn chế của đề tài:**
* Do chưa có nhiều kinh nghiệm trong việc bảo vệ bản quyền ảnh số nên đề tài của em vẫn chưa hoàn thiện theo ý muốn.
* Phần mềm chưa phát triển các phương pháp phòng chống các tấn công thủy vân.
* **Hướng phát triển của đề tài:**
* Nghiên cứu và hoàn thiện thêm đề tài, cố gắng giảm mức độ sai số xuống thấp nhất có thể.
* Phát triển, tích hợp thêm các phương pháp phòng chống tấn công thủy vân.

**TÀI LIỆU THAM KHẢO**

[1] Nguyễn Xuân Huy, Bùi Thế Hồng, Trần Quốc Dũng (2004), “*Kỹ thuật thuỷ vân số trong ứng dụng phát hiện xuyên tạc ảnh*”, Kỷ yếu Hội thảo quốc gia một số vấn đề chọn lọc của công nghệ thông tin lần thứ 7, Đà Nẵng, NXB Khoa học kỹ thuật, Hà Nội, tr. 183-187.

[2] Lê Tiến Thường, Nguyễn Thanh Tuấn (2004), *Giải pháp hiệu quả dùng kỹ thuật watermarking cho ứng dụng bảo vệ bản quyền ảnh số*, Tạp chí bưu chính viễn thông.

[3] Nguyễn Văn Tảo, Đỗ Trung Tuấn, Bùi Thế Hồng, *Một số thuật toán giấu tin và áp dụng giấu tin mật trong ảnh*, kỷ yếu hội thảo RDA 8.

[4] Nguyễn Văn Tảo, Bùi Thế Hồng, “*Nâng cao chất lượng ảnh trong kỹ thuật thuỷ vân sử dụng miền tần số giữa của phép biến đổi DWT-DCT*”, Tạp chí Tin học và Điều khiển học, số 4/2006.

[5] M.Arnold, M.Schmucker, S.D.Wolthusen, “*Techniques and applications of digital watermarking and content protection*”, Artech House, 2003.

[6] B. Barak, [O. Goldreich,](http://www.wisdom.weizmann.ac.il/~oded) [R. Impagliazzo, S. Rudich,](http://www-cse.ucsd.edu/~russell/) A. Sahay,

[S. Vadhan,](http://www.eecs.harvard.edu/~salil/) [K. Yang,](http://www.cs.cmu.edu/~yangke/) “*On the (Im)possibility of Obfuscating Programs*”, In Crypto '01, LNCS No. 2139, 2001, pp. 1-18.

[7] Emir Ganic, Scott D. Dexter, Ahmet M. Eskicioglu, *Embedding Multiple Watermarks in the DFT Domain Using Low and High Frequency Bands*, Department of Computer and Information Science.