Đồ án về giấu tin trong ảnh

Môn: An toàn và bảo mật thông tin

Thành viên:

Nguyễn Đức Hạnh

Trần Xuân Bách

MỞ ĐẦU

Sự ra đời và tiến bộ vượt bậc của công nghệ thông tin được đánh giá là động lực chính của sự thay đổi, là bước ngoặt trong lịch sử phát triển của xã hội, đưa thế giới chuyển từ kỷ nguyên công nghiệp sang kỷ nguyên thông tin và phát triển nền kinh tế tri thức. Cuộc cách mạng thông tin kỹ thuật số đã đem lại những thay đổi sâu sắc trong cuộc sống của nhân loại. Hàng loạt máy móc và các thiết bị số hiện đại nhờ máy tính cá nhân, máy ảnh kỹ thuật số, máy quét, máy in, máy ghi âm kỹ thuật số đã ra đời đem lại nhiều tiện ích cho người sử dụng. Đi kèm theo những phần mềm xử lý tiện ích là vấn nạn vi phạm bản quyền, ăn cắp thông tin, truy cập trái phép ngày càng tinh vi và khó kiểm soát. Hiện đã có một số kỹ thuật được đề xuất để khắc phục những vấn đề trên ví dụ như mã hóa thông tin, chữ ký số, RSA, giấu tin trong các sản phẩm đa phương tiện. Kỹ thuật giấu tin được biết đến bởi hai lĩnh vực chủ yếu là Steganography (giấu tin mật) và Watermarking (thủy vân). Steganography là kỹ thuật giấu tin mật vào các dữ liệu truyền thông (Ảnh, văn bản, nhạc, phim..) để chuyển tải đến người nhận mà thứ ba không thể biết đến sự tồn tại của thông tin mật trong quá trình truyền. Kỹ thuật Steganography cũng làm thay đổi tới duy trong lĩnh vực bảo mật thông tin bởi tính khả thi của việc ẩn một lượng thông tin mật trong một dữ liệu thông thường mà khó bị phát hiện bằng giác quan của con người. Bên cạnh đó Watermarking được sử dụng chủ yếu trong lĩnh vực bảo vệ bản quyền sản phẩm số bằng cách đưa thông tin bản quyền như tên tác giả, logo.. vào sản phẩm. Với sự tồn tại của thông tin thủy vân nhà sản xuất có thể chứng minh được nguồn gốc của sản phẩm khi sản phẩm được phát tán không hợp pháp. Cả hai kỹ thuật được sử dụng với các mục đích khác nhau song chúng đều có đặc điểm chung là giấu thông tin vào sản phẩm số sao cho không bị phát hiện bởi người thứ ba trong quá trình trao đổi thông tin trên mạng. Hiện nay kỹ thuật giấu thông tin mật đã được quan tâm, nghiên cứu và được triển khai ứng dụng rộng rãi trong các cơ quan quân sự, ngoại giao, an ninh, giáo dục và cả các doanh nghiệp khi cần trao đổi các thông tin quan trọng.

Xuất phát từ những nhu cầu trên, học viên quyết định lựa chọn đề tài “Tìm hiểu các kỹ thuật giấu tin trong ảnh” nhằm nghiên cứu, đánh giá các kỹ thuật giấu tin trong ảnh và chọn lựa một vài các kỹ thuật giấu tin tốt để cài đặt thử nghiệm và so sánh đánh giá.

**CHƯƠNG 1: NHỮNG KHÁI NIỆM CƠ BẢN**

* 1. Giới thiệu chung:

Cuộc cách mạng thông tin kỹ thuật số đã đem lại những thay đổi sâu sắc trong xã hội và trong cuộc sống của chúng ta. Những thuận lợi mà thông tin kỹ thuật số mang lại cũng sinh ra những thách thức và cơ hội cho quá trình đổi mới. Sự ra đời những phần mềm có tính năng rất mạnh, các thiết bị mới như máy ảnh kỹ thuật số, máy quét chất lượng cao, máy in, máy ghi âm kỹ thuật số, v.v... , đã được sáng tạo trên cơ sở thoả mãn thế giới tiêu dùng rộng lớn, để xử lý và thưởng thức các dữ liệu đa phương tiện (multimedia data). Mạng Internet toàn cầu đã hình thành một xã hội ảo nơi diễn ra quá trình trao đổi thông tin trong mọi lĩnh vực chính trị, quân sự, quốc phòng, kinh tế, thương mại…Chính trong môi trường mở và tiện nghi như thế xuất hiện những vấn nạn, tiêu cực đang rất cần đến các giải pháp hữu hiệu cho vấn đề an toàn thông tin như nạn ăn cắp bản quyền, nạn xuyên tạc thông tin, truy nhập thông tin trái phép v.v... Tìm giải pháp cho những vấn đề nêu trên không chỉ tạo điều kiện đi sâu vào lĩnh vực công nghệ phức tạp đang phát triển rất nhanh này mà còn dẫn đến những cơ hội phát triển kinh tế.

Giải pháp nào cho những vấn đề trên ?

Trong một quá trình phát triển lâu dài, nhiều phương pháp bảo vệ thông tin đã được đưa ra, trong đó giải pháp dùng mật mã học là giải pháp được ứng dụng rộng rãi nhất. Các hệ mật mã đã được phát triển nhanh chóng và được ứng dụng rất phổ biến cho đến tận ngày nay. Thông tin ban đầu được mã hoá thành các ký hiệu vô nghĩa, sau đó sẽ được lấy lại thông qua việc giải mã nhờ khoá của hệ mã. Đã có rất nhiều những hệ mã phức tạp được sử dụng như DES, RSA, ... Các phương pháp này trong thực tế tỏ ra rất hiệu quả và được ứng dụng phổ biến. Tuy nhiên trong luận văn không đi sâu nghiên cứu về các hệ mật mã mà chỉ tiếp cận với một phương pháp đã và đang được nghiên cứu, phát triển ở nhiều nước trên thế giới, đó là phương pháp che giấu thông tin. Phương pháp này còn mới và phức tạp, ứng dụng trong an toàn và bảo mật thông tin, đang được xem như một công nghệ chìa khoá cho vấn đề bảo vệ bản quyền, nhận thực thông tin và điều khiển truy cập … Để bảo đảm an toàn cho nội dung của thông tin, người ta thường sử dụng phương pháp mã hoá thông tin, nhằm giấu đi ý nghĩa của nó. Để giữ bí mật cho thông tin, người ta tìm ra cách che giấu đi sự hiện diện của nó. Xu hướng hiện nay là kết hợp hai kỹ thuật: mã hóa thông tin sau đó che giấu thông tin. Mã hoá và che giấu thông tin có quan hệ chặt chẽ với nhau. Nhiều ý tưởng của kỹ thuật mật mã (Cryptography) rất hữu ích trong những công việc che giấu sự hiện hữu của thông tin. Nghiên cứu việc kết hợp hai kỹ thuật mật mã và che giấu dữ liệu, nhằm khắc phục những nhược điểm hoặc những hạn chế của từng loại, cho phép xây dựng những hệ thống bảo mật, an toàn cho việc chuyển tải dữ liệu trên các phương tiện thông tin đại chúng.

|  |
| --- |
| Cryptology  Ngành mật mã |

|  |
| --- |
| Cryptography Mật mã |

|  |
| --- |
| Steganography Giấu tin mật |

Hình 1. Phân cấp các lĩnh vực nghiên cứu của mật mã học

Giấu thông tin, xét theo khía cạnh tổng quát cũng là một hệ mã mật, nhằm đảm bảo tính an toàn thông tin. Phương pháp này ưu điểm là làm vô hình nội dung thông tin được chứa trong bức ảnh, đó chính là biện pháp hữu hiệu, hạn chế tối đa được sự phá hoại của những tên tin tặc (hacker). Việc gửi đi những bức ảnh thông thường sẽ không gây ra sự tò mò, chú ý của những tên tin tặc.

1.2. Giấu thông tin và vài nét về lịch sử của nó

* + 1. Vài nét về lịch sử của giấu thông tin

Từ Steganography bắt nguồn từ thời Hi Lạp cổ và được sử dụng cho tới ngày nay, nó có nghĩa là tài liệu được phủ (covered writing). Các câu chuyện kể về kỹ thuật giấu thông tin được truyền qua nhiều thế hệ. Có lẽ những ghi chép sớm nhất về kỹ thuật giấu thông tin (thông tin được hiểu theo nghĩa nguyên thủy của nó) thuộc về sử gia Hi Lạp Herodotus. Khi bạo chúa Hi Lạp Histiaeus bị vua Darius bắt giữ ở Susa vào thế kỷ thứ năm trước Công Nguyên, ông ta đã gửi một thông báo bí mật cho con rể của mình là Aristagoras ở Miletus. Histiaeus đã cạo trọc đầu của một nô lệ tin cậy và xăm một thông báo trên da đầu của người nô lệ ấy. Khi tóc của người nô lệ này mọc đủ dài người nô lệ được gửi tới Miletus.

Câu chuyện khác về thời Hi Lạp cổ đại cũng do Herodotus ghi lại. Môi trường để ghi văn bản chính là các viên thuốc được bọc trong sáp ong. Demeratus, một người Hi Lạp, cần thông báo cho Sparta rằng Xerxes định xâm chiếm Hi Lạp. Để tránh bị phát hiện, anh ta đã bóc lớp sáp ra khỏi các viên thuốc rồi khắc thông báo lên bề mặt các viên thuốc này, sau đó bọc lại các viên thuốc bằng một lớp sáp mới. Những viên thuốc được chuyển công khai và lọt qua mọi sự kiểm soát một cách dễ dàng.

1.2.2 Định nghĩa giấu thông tin

Giấu thông tin là một kỹ thuật nhúng (giấu) một lượng thông tin số nào đó vào trong một đối tượng dữ liệu số khác

Kỹ thuật giấu thông tin nhằm hai mục đích đảm bảo an toàn và bảo mật thông tin: một là bảo mật cho dữ liệu được đem giấu, hai là bảo mật cho chính đối tượng được dùng để giấu tin. Hai mục đích khác nhau này dẫn đến hai khuynh hướng kỹ thuật chủ yếu của giấu tin.

1.2 Những khái niệm cơ bản

1.2.1 Những quy ước.

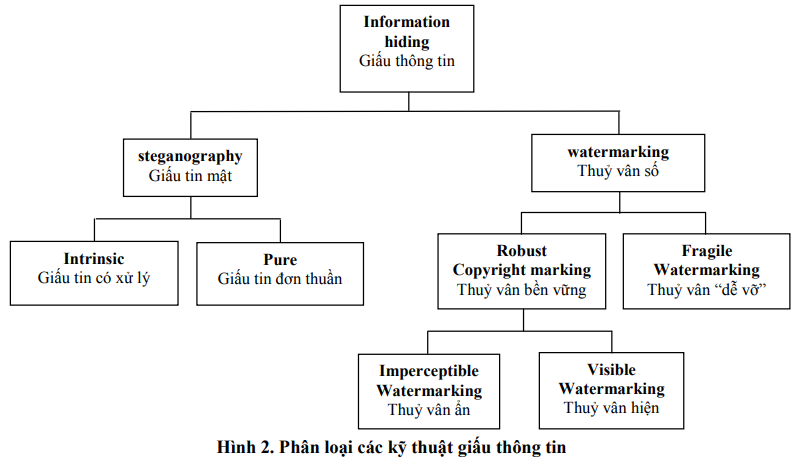
Ảnh môi trường hay đôi khi còn gọi là ảnh gốc (cover image) là ảnh (đối tượng) chứa mang thông tin nhúng trong đó. Nó có thể là ảnh đen trắng, ảnh màu hoặc ảnh đa cấp xám. Trong nghiên cứu này ảnh môi trường sẽ được ký hiệu là C. Nếu có nhiều ảnh môi trường, chúng là sẽ ký hiệu là C1, C2…

Ảnh stego (stego image) là ảnh có chứa thông tin mật trong đó. Ta thường gọi là ảnh có chứa thông tin ẩn và được ký hiệu là S. Nếu có nhiều stego image thì ta ký hiệu là S1, S2…

Để tiện cho việc trình bày, ta gọi hai người liên lạc với nhau là Alice vàBob còn người thứ 3 Wendy không biết được sự hiện hữu của thông điệptrong ảnh mà Alice và Bob trao đổi với nhau

1.2.2 Phân loại các kỹ thuật giấu tin:

Do kỹ thuật giấu thông tin số mới được hình thành trong thời gian gần đây nên xu hướng phát triển vẫn chưa ổn định. Nhiều phương pháp mới, theo nhiều khía cạnh khác nhau đang và sẽ được đề xuất, bởi vậy chưa thể có được một định nghĩa chính xác, một sự đánh giá phân loại rõ ràng. Một số tác giả đã đưa ra các cách đánh giá phân loại, thậm chí các định nghĩa, nhưng không lâu sau lại có các định nghĩa khác, một sự phân loại khác được đề xuất. Sơ đồ phân loại trên hình 2 được Fabien A. P. Petitcolas đưa ra năm 1999, sau hội nghị quốc tế lần thứ hai về giấu tin năm 1998 và đã được chấp nhận rộng rãi trong giới nghiên cứu.

Sơ đồ phân loại này như một bức tranh khái quát về ứng dụng và kỹ thuật giấu thông tin. Dựa trên việc thống kê sắp xếp khoảng 100 công trình đã công bố trên một số tạp chí, cùng với thông tin về tên và tóm tắt nội dung của khoảng 200 công trình đã công bố trên internet, có thể chia lĩnh vực giấu dữ liệu ra làm hai hướng lớn, đó là **watermarking** và **steganography**. Nếu như watermark quan tâm nhiều đến các ứng dụng giấu các mẩu tin ngắn nhưng đòi hỏi độ bền vững lớn của thông tin cần giấu (trước các biến đổi thông thường của tệp dữ liệu môi trường) thì steganography lại quan tâm tới các ứng dụng che giấu các bản tin đòi hỏi mật độ và dung lượng càng lớn càng tốt. Đối với từng hướng lớn này, quá trình phân loại theo các tiêu chí khác có thể tiếp tục được thực hiện, ví dụ dựa theo ảnh hưởng các tác động từ bên ngoài có thể chia watermark thành hai loại, một loại bền vững với các tác động sao chép trái phép, loại thứ hai lại cần tính chất hoàn toàn đối lập, phải dễ bị phá huỷ trước các tác động nói trên. Cũng có thể chia watermark theo đặc tính, một loại cần được che dấu để chỉ có một số những người tiếp xúc với nó có thể thấy được thông tin, loại thứ hai đối lập, cần được mọi người nhìn thấy.

|  |  |
| --- | --- |
| Steganography | Watermarking |
| - Tập trung vào việc giấu được càng nhiều thông tin càng tốt, ứng dụng trong truyền dữ liệu thông tin mật.  - Cố gắng làm nhỏ nhất những ảnh hưởng đến chất lượng của đối tượng vỏ để không bị chú ý đến dữ liệu đã được giấu trong đó.  - Thay đổi stego-object cũng làm cho dữ liệu giấu bị sai lệch (nhất là ứng dụng trong nhận thực thông tin) | Không cần giấu nhiều thông tin, chỉ cần lượng thông tin nhỏ đặc trưng cho bản quyền của người sở hữu.  - Trong trường hợp thuỷ vân nhìn thấy thì thuỷ vân sẽ hiện ra.  - Thuỷ vân phải bền vững với mọi tấn công có chủ đích hoặc không có chủ đích vào sản phẩm. |

**CHƯƠNG 2: STEGANOGRAPHY SECURITY (MỨC ĐỘ AN TOÀN CỦA GIẤU TIN MẬT)**

**2.1 Khái quát chung**

Để đánh giá một thuật toán giấu tin nào đó có đạt các yêu cầu đặt rahay không chúng ta cần đưa ra độ đo (measure) chất lượng của thuật toán đó.Ta ký hiệu phân bố xác suất của ảnh C là Pc và phân bố xác suất củaảnh stego S là tương ứng với một thuật toán được sử dụng nào đó là Ps. Khiđó khả năng phát hiện hệ stego dựa trên entropy giữa phân bố xác suất củaảnh môi trường C và ảnh stego S được “đo” dựa trên công thức:

D(Pc||Ps) =  (1)

Từ phương trình này, chúng ta thấy rằng D(Pc||Ps) tăng theo sự tăng củatỷ số  và do đó, độ tin cậy của việc phát hiện cũng tăng. Vì vậy, kỹ thuật stego được gọi là an toàn tuyệt đối nếu D(Pc||Ps)=0 (tức Pc=Ps) và

Về lý thuyết, người ta đã chứng tỏ được rằng có tồn tại thuật toán antoàn tuyệt đối mặc dù chúng không xẩy ra trong thực hành.

Ở đây chúng ta giả thiết rằng cover image và stego image là những vectơ ngẫu nhiên, độc lập cùng phân bố (independent, identically distributed-iid).

Như vậy để kiểm tra sự khác nhau giữa ảnh gốc và ảnh có giấu tintương ứng, người ta (Wendy ) sẽ kiểm tra sự khác biệt giữa tỷ số PsPc.Trong quá trình kiểm tra này, Wendy sẽ mắc phải hai sai lầm loại 1(type – I error) và sai lầm loại 2 (type –II error) với xác xuất lần lượt là α vàβ (0<α,β<1)

Sai lầm loại một là sai lầm sảy ra khi giả thiết là đúng nhưng anh ta lại bác bỏ nó. Còn sai lầm loại hai là sai lầm sảy ra do chấp nhận giả thiết sai( tức giả thiết là sai nhưng Wendy lại chấp nhận nó).

Chúng ta không thể đồng thời cực tiểu hóa cả hai sai lầm. Thông thường người ta cho cố định xác suất sai lầm loại một và xây dựng bài toán làm cực tiểu hóa sai lầm loại hai (ở đây là cực tiểu hóa β).

Ta ký hiệu d(α,β)= (3)

Khi đó, Entropy giữa hai phân bố Pc và Ps và entropy tương đối của hai phân bố với tham số (α,1-α) và (β,1-β) cần thỏa mãn:

d(α,β) ≤D(Pc||Ps) (2)

Như vậy, đối với hệ thống ε-an toàn chúng ta có:

d(α,β)≤ε(4)

Khi ε=0 hệ thống steganography được gọi là an toàn tuyệt đối.

2.2 Dung lượng chứa thông tin ẩn(steganography capacity).

Để đảm bảo tính chất không thể cảm nhận được (impercepcity) mỗipixel ảnh không được giấu quá một bit dữ liệu. Như vậy tỷ lệ giữa độ dài (quy ra bit) bức thông điệp cần nhúng với số các pixel ảnh môi trường là =12,5%. Nếu độ dài thông điệp cần nhúng so với số pixel ảnh môi trường mà vượt quá con số này thì khả năng vi phạm tiêu chuẩn ‘imperceptibility’ là rất lớn.

2.3 Các kỹ thuật giấu tin mật trong ảnh (image steganography )

Hiện nay, đã có một số thuật toán giấu đã được đề nghị. Các thuật toán này được nhúng trong hai miền: miền không gian (Spatial Domain) và miền biến đổi (Transform Domain).

2.3.1 Nhúng tin trong miền không gian (Spatial Domain Embedding)

Việc nhúng tin trong miền không gian thường được thực hiện với kỹ thuật giấu tin mật. Các thuật toán giấu tin mật nổi tiếng nhất hiện nay đều dựatrên sự thay đổi các bít ít ý nghĩa nhất (Least Significant Bit-LSB) của các điểm ảnh và được gọi là kỹ thuật LSB. Kỹ thuật LSB có nhược điểm là dễ bị phát hiện khi truyền trên băng thông thấp. Tuy nhiên nó có nhiều ưu điểm là nhúng được nhiều lượng thông tin và ít bị mất thông tin trong quá trìnhtruyền.

Chính vì vậy, các thuật toán steganography đều thực hiện trên miềnkhông gian là chủ yếu

2.3.2 Nhúng thông tin trong miền biến đổi(Transform Domain Embedding).

Những thuật toán nhúng trong miền biến đổi đều tận dụng độ dư thừa trong miền DCT (Diserete Cosine Transformatio), chủ yếu đối với ảnh nén JPEG.

Việc nhúng tin trong miền DCT được thực hiện bằng cách thay đổi các hệ số, chẳng hạn như là thay đổi bit ít ý nghĩa nhất của mỗi hệ số.

Một trong những hạn chế của việc nhúng thông tin trong miền DCT là ở chỗ có 64 hệ số bằng 0 (zero) và việc thay đổi hai hay nhiều con zero than hệ số khác không sẽ ảnh hưởng đến tỷ lệ nén. Vì vậy, số các bit 1 cần thiết nhúng trong miền DCT sẽ ít hơn nhiều so với số các bit 1 được nhúng bằng phương pháp LSB. Do đó khả năng chứa dữ liệu ẩn đối với miền DCT phụ

**3- Hình ảnh Steganography**

Với hàng tỷ hình ảnh di chuyển trên internet mỗi năm, có thể nói rằng steganography hình ảnh kỹ thuật số là mối quan tâm thực sự đối với nhiều người trong lĩnh vực bảo mật (11). Hình ảnh kỹ thuật số có thể được sử dụng cho một số loại mối đe dọa bảo mật khác nhau. Trong thế giới doanh nghiệp, việc gửi một tệp bitmap trông vô hại thực sự có thể che giấu những bí mật mới nhất của công ty. Các tệp JPEG có thể được sử dụng trong các tổ chức quốc phòng để che giấu và bảo vệ các bí mật sâu sắc (12). Việc sử dụng hình ảnh kỹ thuật số cho steganography tận dụng những điểm yếu trong hệ thống thị giác của con người. Hệ thống thị giác của con người có độ nhạy thấp đối với những thay đổi mẫu ngẫu nhiên và độ chói. Mắt người không có khả năng nhận ra những thay đổi nhỏ về màu sắc hoặc hoa văn và vì điểm yếu này, các tệp văn bản hoặc đồ họa có thể được chèn vào hình ảnh sóng mang mà không bị phát hiện. Mỗi hình ảnh đồ họa được tạo thành từ các pixel. Màu sắc của mỗi pixel được xác định bởi giá trị số mà nó được gán, nằm trong khoảng từ 0 đến 255. Hình ảnh kỹ thuật số điển hình được tạo thành từ các pixel 8 bit (256 màu) hoặc 24 bit (màu thật, mỗi pixel 8 bit cho màu đỏ, xanh lá cây và xanh lam) (13).

**3-1.Hình ảnh là gì?**

Hình ảnh là một tập hợp các con số tạo thành cường độ ánh sáng khác nhau ở các khu vực khác nhau của hình ảnh. Biểu diễn số này tạo thành một lưới và các điểm riêng lẻ được gọi là pixel (14). Các pixel trong một hình ảnh được hiển thị theo chiều ngang theo từng hàng. Số bit trong bảng màu, được gọi là độ sâu bit, đề cập đến số bit được sử dụng cho mỗi pixel. Độ sâu bit nhỏ nhất trong bảng màu hiện tại là 8, có nghĩa là có 8 bit được sử dụng để mô tả màu sắc của mỗi pixel (tương ứng với 2 ^ 8 = 256 màu). Hình ảnh đơn sắc và thang độ xám sử dụng 8 bit cho mỗi pixel và có thể hiển thị 256 sắc thái khác nhau của màu xám. Hình ảnh màu kỹ thuật số thường được lưu trữ trong các tệp 24 bit và sử dụng mô hình màu RGB, còn được gọi là màu thật (15).

**2-Hệ thống được đề xuất**

Nghiên cứu được đề xuất nhằm mục đích phát triển một phương pháp tiếp cận steganography được cải thiện là Phương pháp LSB thích ứng cho hình ảnh màu có khả năng / chất lượng cao hơn, dung lượng / tải trọng lớn và tốt hơn về độ bền / khả năng chống lại các cuộc tấn công.

Cũng như tin nhắn văn bản có thể được ẩn trong phương pháp mới đó là ẩn các vị trí của LSB bằng các bit văn bản bí mật ở một nơi khác của hình ảnh bìa và sử dụng một phím đặc biệt.

Điều quan trọng là vị trí đầu tiên mà tôi bắt đầu ẩn các vị trí của LSB bằng với các bit của văn bản bí mật .

Hệ thống được đề xuất bao gồm hai thành phần:

1. Mô-đun nhúng

2. Mô-đun trích xuất.

**3-1. Mô-đun nhúng**

Nhúng là quá trình chia ảnh bìa thành hai phần và bắt đầu so sánh các bit văn bản bí mật với LSB trong phần đầu tiên, nếu bằng nhau, hãy lưu các vị trí của LSB sau khi kết thúc so sánh, bắt đầu ẩn các vị trí trong phần thứ hai của hình ảnh và vị trí bắt đầu ẩn trong đó sẽ là chìa khóa và tôi đã chọn nó một cách ngẫu nhiên.

Các hoạt động của Nhúng

1. Nhập các tệp văn bản / hình ảnh bí mật sẽ được ẩn trong ảnh bìa.

2. Chuyển đổi văn bản bí mật trong mã nhị phân và lưu trong mảng1

3. Chọn ảnh bìa (tệp JPEG) từ danh sách các tệp Hình ảnh được lưu trữ và các tệp văn bản.

4. Chia hình ảnh thành hai phần và lưu phần 1 trong mảng 2 và lưu phần 2 trong mảng 3.

5. Bắt đầu so sánh giữa văn bản bí mật (mảng 1) và phần 1 từ hình ảnh (mảng 2)

a) So sánh hai bit đầu tiên với các bit có ý nghĩa nhỏ nhất của part1 (hình ảnh) trong mảng2 với hai bit đầu tiên của văn bản bí mật.

b) nếu hai bit văn bản bằng hai LSB của part1 (hình ảnh) trong mảng 2 thì lưu vị trí của LSB (hàng, cột) trong mảng4.

c) Nếu hai bit văn bản không bằng hai LSB của part1 (hình ảnh) trong mảng 2 thì hãy chuyển đến LSB trong pixel tiếp theo và tiếp tục so sánh cho đến khi tìm thấy LSB bằng hai bit ký tự.

d) Lấy hai bit mới từ văn bản bí mật và chuyển sang bước a cho đến khi kết thúc văn bản bí mật

5. Khi kết thúc hoạt động 4 trở thành tất cả các bit trong văn bản bí mật có LSB bằng với nó và vị trí của nó được lưu trong mảng4.

6. Chuyển đổi array4 sang mã nhị phân và ẩn nó trong array3 part2 của hình ảnh.

7. Vị trí đầu tiên sẽ bắt đầu ẩn trong đó sẽ là chìa khóa .

7. Sao lưu hai phần của hình ảnh lại với nhau và sao lưu ảnh bìa lưu vào đó các vị trí của LSB bằng với văn bản bí mật.

3-2. Mô-đun giải nén

Trích xuất là quá trình nhận thông điệp nhúng từ hình ảnh stego.

Thuật toán chính cho giai đoạn nhúng như sau:

1. Nhập khóa và stego-image .

2. Chia hình ảnh stego thành hai phần

3. Trích xuất các vị trí của LSB bằng văn bản bí mật từ phần 2 và bắt đầu bằng vị trí bằng khóa.

3- Chỉ lưu LSB mà các vị trí trong phần 2 trong mảng

1. Chuyển đổi mảng vị trí sang hệ thống Ascii sau đó thành ký tự (văn bản bí mật)

**5.1- Thuật toán của hệ thống được đề xuất**

Các bước sau đây mô tả thuật toán:

Thuật toán 1: Chuyển đổi văn bản bí mật thành mã nhị phân

Đầu vào : Văn bản bí mật

Đầu ra: Mảng mã nhị phân

Step1-nhập văn bản

Step2- chuyển đổi văn bản sang mã Ascii

Bước 3 - chuyển đổi mã Ascii sang mã nhị phân

Step4 – lưu trong mảng1

Bước 5- Kết thúc

Thuật toán 2: Chia nhỏ hình ảnh

Đầu vào : Hình ảnh

Đầu ra: Hai mảng mã nhị phân

Bước 1- Mở thao tác hình ảnh

Thao tác này sẽ mở tệp hình ảnh và lưu tiêu đề trong một tệp và lưu giá trị bảng màu của nội dung trong một tệp khác.

Bước 2- Chia phần thân của tệp hình ảnh thành hai phần

Thao tác này sẽ chia hình ảnh cơ thể thành các phần để sử dụng một phần so sánh và phần thứ hai ẩn nấp.

Step3- lưu giá trị pixel của part1 trong array2lưu giá trị pixel của part2 trong mảng3

Bước 4-Kết thúc

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Algorthim3: Tìm vị trí

Đầu vào: Array2, Array1

Đầu ra: Vị trí pixel bằng mã ASCII của hình ảnh văn bản

Bước 1-Tìm thao tác vị trí thao tác này được thực hiện bằng cách đọc phần 1 của mảng hình ảnh2

Bước 2- So sánh mảng 1 (nhị phân của văn bản bí mật)

Bước 3- cho I = 1 đến độ dài của mảng1 bước 2

Đọc mảng1

Cho j = 1 đến độ dài của mảng2 bước 8

Đọc mảng2

Nếu array1[I] = array2[j+6] và array1[I+1] = array2[j+7] thì

Mảng4[I] =j+6

Mảng4[I+1]= j+7

Kết thúc nếu

J kế tiếp

kế tiếp I

Bước 4- kết thúc

Thuật toán 4: Trạm biến áp

Đầu vào: Array4, Array3

Đầu ra: Array3 sau khi ẩn Array4 bên trong

Bước 1- cho i = 1 đến độ dài của mảng4

Đọc mảng4

Cho j = Key to length of array3 step 8

Array3[j+6]=array4[I]

Mảng3[j+7]=mảng4[I+1]

Kế tiếp j

Kế tiếp i

Bước 2- kết thúc

Thuật toán5: trích xuất stego-image

Đầu vào: Array2, Array3

Đầu ra: stego-image

Step1- quay lại mảng2 để tập tin

Step2- quay lại mảng3 vào một tệp khác

Bước 3- hợp nhất hai tệp để trích xuất bảng màu

Bước 4 - quay lại tiêu đề vào bảng màu để trích xuất stego-image

Bước 5- kết thúc

**6- Trích xuất văn bản bí mật**

**6-1. Hoạt động trích xuất**

1. Chia hình ảnh stego thành hai phần và lưu phần 1 trong mảng5 và lưu phần 2 trong mảng6.

2. Trích xuất vị trí từ Array6 và lưu trong Array7

3. Theo array7 trích xuất văn bản bí mật từ mảng5

**6.2- Các thuật toán trích xuất**

Các bước sau đây mô tả thuật toán:

Thuật toán 1: Chia nhỏ hình ảnh

Đầu vào : Hình ảnh

Đầu ra: Hai mảng mã nhị phân

Bước 1- Mở thao tác hình ảnh

Thao tác này sẽ mở tệp hình ảnh và lưu tiêu đề trong một tệp và lưu giá trị bảng màu của nội dung trong một tệp khác.

Bước 2- Chia phần thân của tệp hình ảnh thành hai phần

Thao tác này sẽ chia hình ảnh cơ thể thành hai phần.

Step3- lưu giá trị pixel của part1 trong mảng 5 và part2 trong mảng6

Bước 4-Kết thúc

Thuật toán 2: trích xuất văn bản bí mật từ ảnh bìa

Đầu vào: mảng6

Đầu ra: Hai mảng mã nhị phân

Bước 1- cho i = 1 đến độ dài của mảng6 bước 8

Cho j = 1 đến độ dài của văn bản bí mật

Mảng7[j] = mảng6[i+6]

Mảng7[j+1] = mảng6[i+7]

Kế tiếp j

Kế tiếp i

Step2- cho m = key to length array5

Đọc Array5[m] =

Đọc Array7[m]

Nếu Mảng5[m] = m thì

Cho n = 1 đến độ dài của văn bản bí mật

Mảng8[n] = Mảng7[m]

Tiếp theo n

M tiếp theo

Step3- chuyển đổi array8 sang asci code theto text

Bước 4 - in văn bản bí mật

Bước 5 - kết thúc

Step3- lưu giá trị pixel của part1 trong mảng 5 và part2 trong mảng6

Bước 4-Kết thúc

@ **Kết Luận**

Có một số loại thuật toán cho steganography. Mỗi loại thuật toán đều có những ưu điểm và hạn chế riêng. Không có phương pháp nào có thể cung cấp giải pháp hoàn toàn hoàn hảo. Mỗi loại giải pháp có tính mạnh mẽ đối với một số loại tấn công nhưng kém linh hoạt hơn đối với một số loại tấn công khác. Trọng tâm chính của nghiên cứu hiện tại ở các vị trí ẩn . Trong trường hợp áp dụng thực tế, việc lựa chọn loại giải pháp thực sự phụ thuộc vào bản chất của ứng dụng và yêu cầu. Phương pháp được đề xuất sử dụng Phương pháp LSB để tối ưu hóa sức mạnh của quá trình steganographic. Tính không thể chấp nhận và mạnh mẽ của phương pháp được đề xuất cho thấy hiệu suất tốt hơn so với các phương pháp khác trong thực tế. Độ chính xác của hệ thống được đánh giá là 94%, cho thấy sự cải thiện đáng kể so với các phương pháp hiện có.

**CHƯƠNG 3 KỸ THUẬT THỦY VÂN ẢNH SỐ SỬ DỤNG ROI VÀ EXTRACTION VOTING TRÊN MIỀN TẦN SỐ**

**I.Giới thiệu**

Ngày nay, các sản phẩm multimedia như video, âm thanh, hình ảnh, văn bản được sản xuất và truyền đi một cách nhanh chóng, phục vụ mục đích giải trí, thương mại. Chúng dễ dàng được sao chép, phát tán. Nhiều người không được cấp quyền cũng dễ dàng truy cập và kiếm lợi nhuận từ những sản phẩm không phải của mình. Hành vi sao chép, chia sẻ, sửa đổi khi không có sự cho phép chủ sở hữu hợp pháp là vi phạm bản quyền sản phẩm số. Hiện tượng này thường xuyên xảy ra trên mạng internet. Vì vậy, nhà sản xuất hoặc chủ sở hữu hợp pháp các sản phẩm multimedia có nhu cầu cao đối với việc bảo vệ bản quyền sản phẩm của họ.

Thủy vân số - Digital watermarking [1] là một giải pháp kỹ thuật đáp ứng nhu cầu bảo vệ bản quyền các sản phẩm số. Các thông tin mật (thông tin bản quyền, logo, thông tin tác giả, chữ ký số, . . . ) được mã hóa và nhúng vào dữ liệu đa phương tiện mà không làm ảnh hưởng đến chất lượng để làm căn cứ khi xảy ra tranh chấp hoặc xử lý vi phạm bản quyền. Khi xảy ra tranh chấp, thông tin mật sẽ được trích rút dựa vào khóa bí mật. Từ đó, chứng mình quyền sở hữu hợp pháp đối với sản phẩm số.

Thủy vân ảnh số [1] [2] là một trong các kỹ thuật thủy vân phổ biến, được nghiên cứu và ứng dụng rộng rãi trong lĩnh vực bảo vệ bản quyền ảnh số. Các nhà nghiên cứu đã phát triển nhiều lược đồ thủy vân trên miền không gian và miền tần số. Đặc biệt, kỹ thuật thủy vân ảnh số phát triển mạnh trên các miền tần số sử dụng các biến đổi DCT, SVD, DWT [2], [5]–[8], [11]hoặc các miền tần số kết hợp [9], [12]. Họ khai thác các đặc tính của mỗi miền dữ liệu để tăng cường tính bền vững [9], tính ẩn, cải thiện tốc độ tính toán cũng như khả năng ẩn giấu thông tin mật trong các lược đồ thủy vân.

Trong nhiều lược đồ thủy vân, các tác giả thường nhúng thông tin lên toàn bộ các vùng của ảnh mang thông qua các phép biến đổi [5], [7], [9], Kỹ thuật này đảm bảo nhúng được lượng thông tin lớn. Tuy nhiên, việc tác động lên toàn bộ các vùng dữ liệu đã phần nào làm giảm đi chất lượng của ảnh mang.

Qasim và cộng sự [3] đã áp dụng lược đồ thủy vân cho các ảnh chụp MR não trong y tế. Sử dụng ROI (Region Of Interest) để xác định vùng nhúng dữ liệu và vùng không nhúng dữ liêu. Vùng sử dụng nhúng dữ liệu là vùng ROI (nền) có độ mượt cao hơn (có độ chênh lệch giá trị màu giữa các điểm ảnh thấp) và vùng không quan tâm tới RONI (phần dữ liệu quan trọng của ảnh hoặc không được phép tác động vào). Ngưỡng T được dùng để xác định ROI được chọn bằng cách thử nghiệm trên một tập nhiều ảnh đa dạng và ước lượng giá trị thích hợp. Lược đồ thủy vân được các tác giả trình bày là lược đồ thủy vân thuận nghịch, nhằm phát hiện ra sự thay đổi của các bức ảnh y tế chụp não bộ của bệnh nhân.

Jeya và cộng sự [4] cũng đề xuất lược đồ thủy vân trên vùng dữ liệu RONI (Region-Of-None-Interest) áp dụng trên ảnh CT của bệnh viêm phổi trong y tế. Việc xác định RONI và nhúng thông tin mật lên đó để tránh làm thay đổi thông tin quan trọng của ảnh.

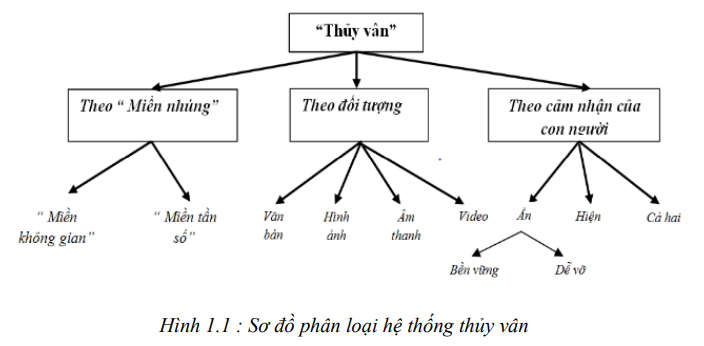
Trong kỹ thuật xử lý ảnh hoặc trong lĩnh vực thị giác máy tính, ROI có thể là bất cứ hình dạng nào, có thể hình vuông, chữ nhật, tròn, hoặc nhiều vùng nhỏ trên ảnh hoặc thậm chí là các pixel. Với các vùng ROI lớn, nếu sử dụng để nhúng thông tin thì sẽ đảm bảo các vùng khác không bị tác động, làm cho chất lượng ảnh ở đó không bị thay đổi, còn các vùng bị tác động chỉ thay đổi ở mức thấp nhất. Do đó, chất lượng ảnh tổng hợp sau khi nhúng thủy vân sẽ tốt hơn. Nhưng lại đối diện với nguy cơ mất thông tin ẩn giấu nhiều hơn khi bị tấn công thủy vân bởi các kỹ thuật như cắt ảnh, ghép ảnh. Do đó, việc xác định ROI là bước quan trọng. Vì vậy, ở trong thí nghiệm này, ảnh mang sẽ được chia thành các khối có kích thước 8x8 không trùng lặp. Trên mỗi khối sẽ tính giá trị phương sai [16] theo công thức (1) để xác định xem khối có thích hợp để nhúng thông tin hay không. Đồng thời lưu lại thông tin của các khối được chọn để làm chìa khóa cho quá trình trích rút thông tin. Bên cạnh đó, để tăng cường độ chính xác, thông tin sẽ được nhúng trên cả 3 kênh màu của ảnh đầu vào. Do đó, khi trích rút, ngoài việc dùng các khóa như trên, còn kết hợp thêm kỹ thuật bỏ phiếu (voting) để tăng cường độ chính xác.

Giá trị phương sai của khối các điểm ảnh có kích thước N×M được tính theo công thức sau [16]:



**II.PHÂN LOẠI THỦY VÂN**

Thủy vân và kỹ thuật thủy vân tùy theo từng tiêu chí phân loại mà có thể được chia thành nhiều loại khác nhau :

****

2.1.Phân loại thủy vân theo miền nhúng:

Một trong những tiêu chí để phân loại là “miền nhúng ” là nơi chứa thủy vân. Ví dụ, thủy vân có thể được thực hiện trong “miền không gian”. Một khả năng khác là thủy vân trong miền tần số.

2.2.Phân loại theo đối tượng được nhúng thủy vân :

Kỹ thuật thủy vân có thể được phân loại theo đối tượng đa phương tiện cần nhúng thủy vân như sau:

+ Thủy vân trên ảnh

+ Thủy vân trên video

+ Thủy vân trên âm thanh

+ Thủy vân trên văn bản

2.3.Phân loại thủy vân theo cảm nhận của con người

Tùy theo cảm nhận của con người, thủy vân có thể được chia ba loại khác nhau

+ Thủy vân hiện: hiển thị cho người xem thông tin về sản phẩm dưới dạng các hình mờ.

+ Thủy vân ẩn bền vững: được nhúng bằng cách thay đổi trên điểm ảnh sao cho hệ thống cảm giác của con người không thể nhận thấy và phải chịu được các thông tác xử lý tín hiệu thông thường “tấn công” và nó chỉ có thể được phục hồi với cơ chế giải mã thích hợp mà thôi. Xét theo tính bí mật của thủy vân bền vững được phân loại nhỏ hơn như sau :

-Lược đồ “thủy vân” bí mật : Cần tới ảnh gốc để trích xuất thủy vân. Có 2 loại lược đồ thủy vân bí mật :

Loại 1: yêu cầu cả ảnh bị biến đổi và ảnh gốc khi trích xuất thủy vân. Ảnh gốc được sử dụng để tìm kiếm vị trí thủy vân trong bức ảnh bị biển đối.

Loại 2: trong đó yêu cầu một bản sao của thủy vân trong quá trình trích xuất và kiểm tra mới có thể biết được thủy vân ở trong bức ảnh cần kiểm tra hay không

Trong hai loại trên khi trích xuất thủy vân cần đòi hỏi có chìa khóa bí mật. Đối với loại thứ nhất thì chìa khóa bí mật ở đây là ảnh gốc, còn đối với loại chìa khóa thứ 2 thì chìa khóa bí mật là dữ liệu bí mật được sử dụng để nhúng vào bức ảnh (hay nói cách khác là thủy vân )

- Lược đồ thủy vân nửa bí mật

Không sử dụng ảnh gốc trong quá trình xác định thủy vân. Tuy nhiên, lược đồ này chỉ đưa ra thông tin có sự hiện diện của thủy vân hay không .

- Lược đồ thủy vân mù Trong lược đồ này, không yêu cầu ảnh gốc lẫn thủy vân được nhúng trong quá trình trích thủy vân

- Lược đồ thủy vân khóa công khai

Còn được gọi là thủy vân bất đối xứng. Trong lược đồ này, chìa khóa để tìm kiếm và trích xuất thủy vân được công khai với mọi người trái ngược với thủy vân bí mật chìa khóa để tìm kiếm và trích xuất thủy vân là chìa khóa bí mật.

+ Thủy vân ẩn dễ vỡ : được nhúng theo cách mà bất kỳ biến đổi hay giả mạo đều làm thay đổi hay phá hủy “thủy vân”.

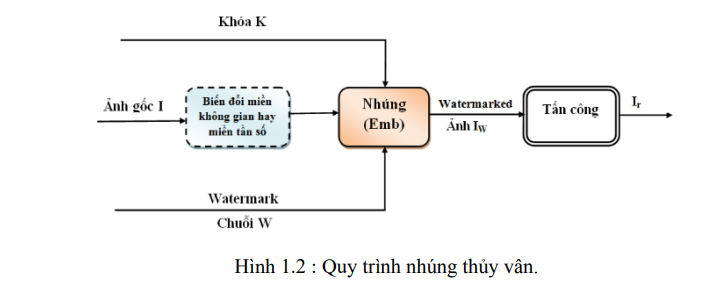
+ Thủy vân hiện và ẩn đồng thời : (dual watermark) là sự kết hợp giữa thủy vân ẩn và thủy vân hiện

2.4. MÔ HÌNH THỦY VÂN SỐ

2.4.1. Tạo thủy vân số

2.4.2 Quy trình nhúng thủy vân

Bộ nhúng thủy vân là chương trình được cài đặt những thuật toán thủy vân và được thực hiện với một khóa bí mật



Hình 1.2 trình bày và giải thích quá trình nhúng thủy vân cho ảnh tĩnh. Trong đó, Ảnh gốc được kí hiệu bằng I, “thủy vân” được kí hiệu bở W, hình ảnh chứa “thủy vân” là I w và K là khóa nhúng. Hàm nhúng EMB có đầu vào là ảnh gốc I, “thủy vân” W và khóa K và tạo ra một ảnh mới có chứa thủy vân mới thể hiện bằng Iw.

Khóa nhúng K là thực sự cần thiết cho việc nâng cao khả năng bảo mật của hệ thống “thủy vân”.

Đối với kỹ thuật biến đổi theo miền không gian :

Emb (I, W, K) = I w

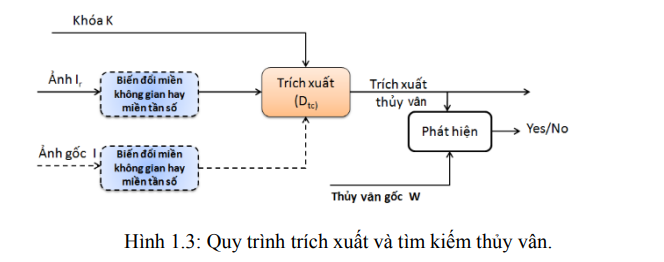
Đối với kỹ thuật biến đổi theo miền tần số :

Emb (f, W, K) = I w

Trong đó f là vecto hệ số cho phép biến đổi.

2.4.3.Trích xuất và tìm kiếm thủy vân

Quá trình tách thủy vân được thực hiện thông qua một bộ tách thủy vân tương ứng với bộ nhúng thủy vân cùng với khóa của quá trình nhúng.



Hình 1.3 trình bày và giải thích quy trình trích xuất và tìm kiếm thủy vân ở trong ảnh tĩnh. Một hàm phát hiện Dtc có đầu vào là hình ảnh Ir có chức năng xác định quyền sở hữu sản phẩm. Các hình ảnh I r có thể chứa thủy vân hoặc không chứa thủy vân. Trong trường hợp tổng quát, hình ảnh có thể bị biến đổi. Hàm phát hiện có khả năng khôi phục thủy vân We từ bức ảnh hoặc kiểm tra sự hiện diện của thủy vân W trong bức ảnh đã cho I r hay 10 không. Trong quá trình này hình ảnh gốc I cũng có thể yêu cầu, phụ thuộc vào lược đồ thủy vân được lựa chọn.

Biểu thức toán học cho thủ tục trích xuất mù (trích xuất không sử dụng ảnh gốc I) cụ thể như sau :

Dtc (Ir ,K) = W

Biểu thức toán học cho thủ tục trích xuất không mù (trích xuất có sử dụng ảnh gốc I) cụ thể như sau :

Dtc (Ir ,I, K) = We

Thuật toán phát hiện thủy vân mù có đầu ra là một giá trị nhi phân cho biết có sự hiện diện của thủy vân hay không.

2.5.CÁC HƯỚNG ỨNG DỤNG CỦA THỦY VÂN

\* Bảo vệ bản quyền ảnh số

Trong việc mua bán và trao đổi các tác phẩm số này nảy sinh các vấn đề cụ thể như sau:

+ Vấn đề thứ 1 là phải bảo đảm quyền tác giả.: Để bảo vệ được bản quyền của người sở hữu ảnh số thì ảnh số đó phải có những thông tin đặc biệt chứng minh nó là thuộc quyền sở hữu của minh.

+ Vấn đề thứ 2 là đảm bảo thông tin sẵn sang cho người dùng hợp pháp và chống phân phối bất hợp pháp nội dung tác phẩm: mua bán,…

+ Vấn đề thứ 3 lần vết thông tin phát hiện người phân phối sản phậm bất hợp pháp: khi vấn đề về vi phạm bản quyền xảy ra hoặc khi chủ sở hữu sản phẩm số nghi ngờ là có bản sao sản phẩm không hợp lệ.

\* Dấu vân tay hay dán nhãn : thủy vân trong những ứng dụng này được sử dụng để nhận diện người gửi hay người nhận một thông tin nào đó.

\* Điều khiển truy nhập: các thiết bị phát hiện thủy vân (ở đây sử dụng phương pháp phát hiện thủy vân đã giấu mà không cần thông tin gốc) được gắn sẵn vào trong các hệ thống đọc ghi, tùy thuộc vào việc có thủy vân hay không để điều khiển (cho phép/ cấm) truy cập. Ví dụ hệ thống quản lý sao chép DVD được ứng dụng ở nhật .

2.6 .ĐẶC TÍNH CỦA THỦY VÂN

Trước đây, đã có một số bài báo thảo luận về đặc tính của thủy vân. Một số thuộc tính thường được thảo luận như: tính phức tạp, tính trung thực hình ảnh, độ tin cậy phát hiện, tính bền vững, dung lượng, bảo mật,…Trong thực tế, không thể để thiết kế một hệ thống thủy vân đảm bảo được tất cả các thuộc tính trên. Do đó, việc đảm bảo cân bằng giữa các thuộc tính là thực sự cần thiết và vấn đề đảm bảo cân bằng phải dựa trên sự phân tích ứng dụng một cách cẩn thận.

• Độ trung thực

Độ trung thực nghĩa là người theo dõi không thể phát hiện ra dấu thủy vân hay nói cách khác dấu thủy vân không làm giảm chất lượng hình ảnh. Để tín hiệu thực sự là không thể cẩm thấy thì thông tin phải được nhúng vào những bít ít quan trọng. Tuy nhiên, tín hiệu lại dễ dàng bị loại bỏ trong quá trình nên có tổn thất thông tin.

• Tính bền vững

Hình ảnh được thủy vân có thể phải trải qua nhiều loại xử lý biến đổi khác nhau, ví dụ, tăng độ tương phản, lọc thông, làm mờ,…

Do vậy, dấu thủy ký phải có tính bền vững mới chịu được các phép biến đổi ảnh cũng như biến đổi tín hiệu số thành tín hiệu tương tự, tương tự thành số và nén. Ngoài ra, ảnh chứa thủy vân phải chịu được các phép biến đổi hình học như di chuyển vị trí, co dãn kích thước và cắt xén.

• Tính dễ hỏng Là thuộc tính đối ngược hoàn toàn với tính bền vững của thủy vân. Thuộc tính này thường được ứng dụng trong lược đồ thủy vân vỡ. Với lược đồ này yêu cầu đặt ra là dấu thủy ký hoặc bị phá hủy bởi bất cứ phương pháp sao chép nào ngoại trừ các phương pháp sao chép hợp pháp

• Tỉ lệ lỗi sai dương Tỉ lệ lỗi sai dương là xác suất hệ thống phát hiện nhầm: xác định một mẩu dữ liệu không mang dấu thủy ký là mang dấu thủy ký. Tùy theo ứng dụng mà ảnh hưởng của lỗi là khác nhau, trong một số ứng dụng có thể là rất nghiêm trọng. Do đó, trong ứng dụng, người ta phát tính toán trước sao cho tỷ lệ lỗi sai dương nhỏ hơn mức cho phép.

• Tính dư thừa

Tính dư thừa liên quan đến một thực tế là thủy vân được lặp lại ở những vùng tấn số khac nhau, do đó nếu có một lỗi trên một vùng tần số thì vẫn có thể được khôi phục thông điệp từ các dải tần khác. Tính dư thừa ánh xạ đến tính bền vững, có nghĩa là thủy vân có thể được khôi phục ngay cả khi nó bị biến đổi ở độ nhất dịnh do sự vô ý hay tấn công có chủ ý.

• Độ phức tạp tính toán Cũng như bất cứ công nghệ nào sử dụng trong thương mại, độ phức tạp tính toán của lược đồ thủy vân đều rất quan trọng. Điều này, đặc biệt đúng khi sử lý với các dữ liệu thời gian thực.

2.1.3.Thuật toán LBS

Về cơ bản, kỹ thuật thủy vân LBS dựa trên tần suất xuất hiện của các bit 0 và 1 trong file ảnh gốc và trong thông điệp cần mã hóa, từ đó đưa ra sự thay thế các bit này để thực hiện việc giấu tin .

Cụ thể hơn, trong kỹ thuật thủy vân LSB, bit cuối cùng của mỗi byte đuợc đặt giá trị 0, sau đó tùy thuộc vào giá trị 0 hoặc 1 của dữ liệu mà thay đổi. Nếu bit của dữ liệu là 0 thì giữ nguyên, còn nếu bit của dữ liệu là 1 thì sẽ đổi giá trị này trên ảnh thành 1.

Để thực hiện kỹ thuật thủy vân này, cần một ảnh gốc, hay còn gọi là cover image. Do phương pháp này sử dụng những bits của từng pixcel trong ảnh, nó đòi hỏi một định dạng nén không mất thông tin. Khi ta sử dụng ảnh màu 24 bit, từng bit của mỗi màu thành phần R, G, B đều có thể được sử dụng, như vậy có thể giấu được 3 bit trong mỗi điểm ảnh, đồng nghĩa với việc nhúng được nhiều thông tin hơn.

• Dữ liệu vào

+ Ảnh gốc

+ Dữ liệu thủy vân

+ Khóa bí mật

• Dữ liệu ra

+ Ảnh mang: có chứa thông tin thủy vân. Ảnh mang có sự thay đổi không đáng kể so với ảnh gốc.

+ Để có thể lưu trữ lượng thông tin lớn và sự thay đổi màu sắc của ảnh là không đáng kể, chúng ta sử dụng file bitmap 24 bit . Cụ thể hơn, một pixel 26 của ảnh được biểu diễn 3 màu đỏ, xanh lá cây và xanh da trời (R|, G, B), mỗi màu sử dụng 8 bit. Tuy nhiên trong lược đồ trình bày dưới đây, chúng ta chỉ sử dụng bit cuối cùng màu xanh da trời để giấu thông tin.

• Thuật toán nhúng thủy vân:

+ Bước 1 : Tính tổng số bytes cần dùng để nhúng thủy vân. Giá trị này được lưu trong biến int S.

+ Bước 2: Đọc các kí tự từ file text sau đó chuyển giá trị ASCII của chúng sang dạng nhị phân 8 bit, lưu giữ trong một mạng từ A [7] về A [0] (A[0] là LSB)

+ Bước 3: Tính toán xem có tất cả bao nhiêu bit 0 và 1 xuất hiện trong mỗi byte, lưu tổng các giá trị này lần lượt là i 0 và i1.

+ Bước 4: Lặp lại từ bước 1 -> 4 cho đến khi kết thức toàn bộ văn bản EOF.

+ Bước 5 : Từ ảnh gốc , đọc giá trị RGB của mỗi pixel.

+ Bước 6: Đọc giá trị bit cuối cùng của mỗi pixels. Với ảnh RGB 24 bit thì ta sẽ đọc bit cuối cùng trong số 8 bit của màu xanh da trời.

+ Bước 7: Kiểm tra xem bit này có giá trị 0 hay l, sau đó tính tổng số lần xuất hiện các bit này trong S pixcels, lưu lần lượt vào hai biến c 0 và c 1.

+ Bước 8: Lặp lại từ bước 5 cho đến bước 7 [8\*S] lần. Đây là số pixel cần đọc để có thể giấu toàn bộ các byte thông điệp +

Bước 9: Nếu [ (c0 > c1) và (i0 > i1) ] và [ (c1> c0) và (i1>i0)], đặt gí trị cho flag = 0, ngược lại đạt giá trị flag = 1.

+ Bước 10 : Ghi giá trị của flag vào phía bên trái của bit cuối cùng của pixel đầu tiên trong ảnh giấu.

+ Bước 11: Mở ảnh gốc ở chế độ đọc. Tạo một ảnh mang giống ảnh gốc ở chế độ ghi. 27

+ Bước 12: đọc Header của file gốc. Ghi thông tin header này lên ảnh mang. Từ ảnh gốc, đọc giá trị RGB của mỗi pixel.

+ Bước 13: đọc bit stream của dữ liệu. Nếu giá trị của cờ là 0 thì giữ nguyên giá trị bit của dữ liệu, sau đó ghi đè lên bit cuối cùng của màu xanh của pixel, ngược lại, nếu giá trị cờ là 1 thì đảo lại bit dữ liệu rời mới ghi lên pixcel (0 thành 1 hoặc 1 thành 0). Ghi pixel này vào ảnh stego.

+ Bước 14: Nếu toàn bộ các LSB đã được sửa đổi thành công, thì ghi nốt các bit còn lại của các pixel vào ảnh mang. Ngược lại, quay lại bước 13.

• Thuật toán tách thủy vân

+ Bước 1: Mở ảnh Stego dưới chế độ đọc

+ Bước 2: Đọc bit liền kề bit cuối của pixel đầu tiên trong ảnh. Dựa trên giá trị của nó, đặt giá trị flag là 0 hoặc 1.

+ Bước 3: Đọc từng pixel của ảnh Stego.

+ Bước 4: Nếu flag là 0 thì đọc bit cuối cùng của mỗi pixel và đưa vào một mảng. Ngược lại nếu flag = 1 thì đảo bit rồi mới chuyển vào mảng.

+ Bước 5: Đọc mỗi 8 pixel theo cách trên, sau đó chuyển nội dung của mỗi 8 phần tử của mạng sang hệ thập phân, dây chính là giá trị ASCII của kí tự.

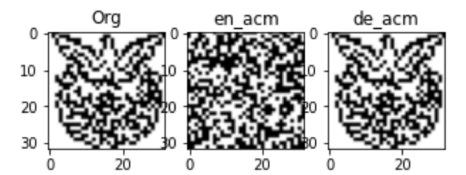
+ Bước 6: Nếu chưa gặp giá trị EOF thì in kí tự và quay lại bước 3

**III. ĐỀ XUẤT LƯỢC ĐỒ THỦY VÂN**

A. Biến đổi

Arnold logo Trong lược đồ thủy vân, để tăng cường bảo mật, thông tin mật là logo bản quyền được mã hóa sử dụng kỹ thuật xáo trộn Arnold [8], [12], [15], [18]. Xét ảnh đầu vào có kích thước NxN được mã hóa bằng cách xáo trộn các điểm ảnh theo phương trình sau [15], [18]:

 mod N

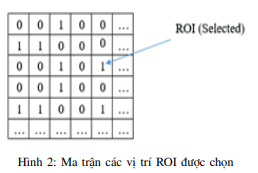
Trong đó, x, y tương ứng là tọa độ của ảnh gốc; x ′, y ′ tương ứng là tọa độ của điểm ảnh bị xáo trộn. Biến đổi Arnold thực hiện hoán đổi vị trí của các pixel trong ảnh nhiều lần. Với N=32 thì sau 24 lần lặp, hình ảnh xáo trộn sẽ bằng với hình ảnh ban đầu. Do đó, để mã hóa logo thông tin mật bằng biến đổi Arnold, khóa “key wm” là một số nguyên sẽ được truyền vào và thực hiện xáo trộn “key wm” lần.

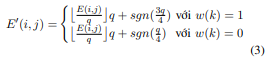
Hình 1: Biến đổi Arnold với khóa "key wm" Từ trái sang phải: Logo gốc, Logo mã hóa, Logo giải mã

B. Xác định vùng dữ liệu nhúng thông tin ROI

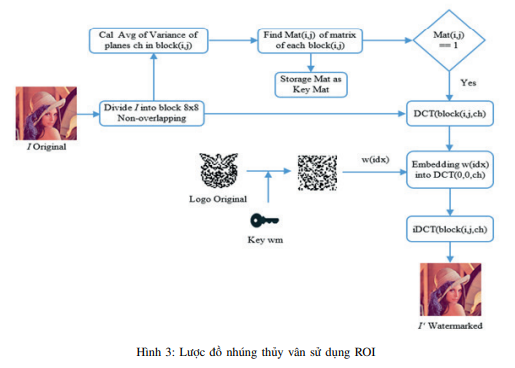
Vùng quan tâm ROI là vùng dữ liệu được sử dụng cho mục đích nhúng thông tin mật mà không làm giảm chất lượng ảnh mang. Ở đây, ảnh mang được chia thành các khối 8x8 riêng biệt. Trên mỗi khối, tính giá trị phương sai theo công thức (1). Để đánh giá khối có phù hợp để nhúng thông tin không, thực hiện so sánh giá trị phương sai với ngưỡng T. Các khối thỏa mãn sẽ được lưu lại thông tin vào ma trận M at để làm khóa cho quá trình trích rút thông tin mật và được gán giá trị là 1. Ngược lại, các khối không được chọn để nhúng thông tin được gán giá trị là 0. Thông tin vị trí các khối được lưu lại dưới dạng ma trận nhị phân M at. Ma trận M at được mã hóa (AES) và gửi cho cơ quan quản lý, chứng thực bản quyền CA (Certificate Authority). Khi có tranh chấp xảy ra, CA sẽ cung cấp “bộ khóa” trong đó có ma trận M at chứa thông tin vị trí nhúng, để trích rút thông tin bản quyền.

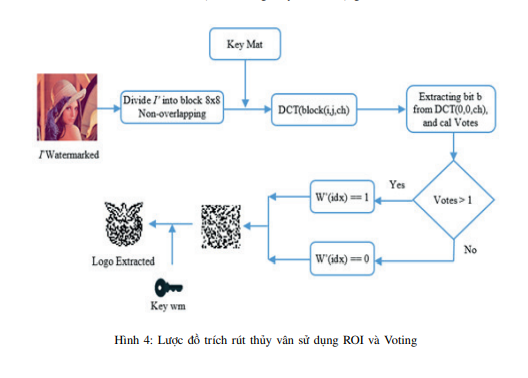
Ngưỡng T là giá trị để đánh giá xem khối có phù hợp để nhúng thông tin mật. Giá trị của T được xác định thông qua quá trình thực nghiệm trên tập hợp gồm nhiều ảnh khác nhau và có thể điều chỉnh phù hợp với từng tập dữ liệu. Tuy nhiên T phải đảm bảo là lớn hơn giá trị phương sai nhỏ nhất và nhỏ hơn giá trị phương sai lớn nhất của ảnh. Việc xác định giá trị T cũng cần phải căn cứ vào tương quan kích thước giữa logo bản quyền (thông tin mật) và kích thước của ảnh mang.

C. Tiến trình nhúng thủy vân

Trong quá trình nhúng thủy vân, sử dụng biến đổi DCT [6], [7], [9], [11] và kỹ thuật QIM [7], [14] để nhúng thông tin. Khối được chọn sẽ được biến đổi DCT 2 chiều, chọn phần tử DCT[0, 0] chứa nhiều năng lượng ảnh nhất, để tác động theo kỹ thuật QIM với trọng số nhúng q = 0.01. Phương trình nhúng cụ thể:

Trong đó, E(i, j) là giá trị DCT tại vị trí [0, 0, ch]; E′ (i, j) là giá trị DCT tại vị trí [0, 0, ch] nhận được sau khi nhúng thông tin; w(k) là giá trị bit nhị phân tại ví trí thứ k của chuỗi thông tin mật.





Các bước thực hiện của quá trình nhúng thủy vân như sau:

Bước 1: Chia ảnh mang I thành các block 8x8 không trùng lặp

Bước 2: Tính giá trị trung bình avg phương sai của 3 plane trên mỗi block

Bước 3: Tính ma trận M at:

Nếu avg >= T & count(M at[i, j] == 1) <= wmlen) thì: M at[i, j] = 1

Ngược lại: M at[i, j] = 0

Bước 4: Sử dụng Arnold Transform để mã hóa logo w với key wm

Bước 5: Duyệt ma trận M at, nếu M at[i, j] == 1 thì thực hiện:

Trên mỗi plane ch:

- Biến đổi DCT block[i, j, ch]

- Sử dụng phương trình (3) để nhúng bit w(idx) vào DCT[0, 0, ch]

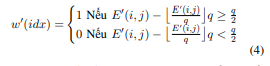
- Biến đổi iDCT block[i, j, ch]

Bước 6: Khôi phục I ta thu được ảnh I ′ đã được nhúng thông tin mật

Sau khi thực hiện nhúng thủy vân, ảnh số được phân phối tới người dùng hợp pháp. Các giá trị “key wm”, ma trận thông tin ROI M at, trọng số nhúng q được lưu trữ thành bộ khóa K(key wm, M at, q). Bộ khóa K được mã hóa (AES) và gửi lên cơ quan quản lý bản quyền CA (Certificate Authority). Khi xảy ra tranh chấp, CA sẽ cung cấp bộ khóa và tiến hành trích rút thông tin để chứng minh quyền sở hữu hợp pháp.

D. Tiến trình trích rút thủy vân sử dụng Voting

Khi xảy ra tranh chấp, CA sẽ giải mã và cung cấp bộ khóa K(key wm, M at, q). Ma trận M at được sử dụng để xác định các vị trí đã giấu thông tin mật; trọng số q dùng để trích rút thông tin mật theo phương trình (4); khóa “key wm” được dùng để giải mã thông tin mật thu được. Để trích rút thông tin, sử dụng phương trình (4) như sau:



Trong đó, E′ (i, j) là giá trị DCT tại vị trí [0, 0, ch], q là trọng số nhúng, w ′ (idx) là giá trị thứ idx được trích rút từ E′ (i, j).

Khi thực hiện trích rút thông tin mật b trên từng plane ch, so sánh với các giá trị 0 và 1 để tính toán Vote. Sau đó thực hiện voting để tìm ra bit w ′ (idx) thích hợp nhất. Các bước thực hiện trích rút thủy vân cụ thể:

Bước 1: Chia ảnh mang I ′ thành các block 8x8 không trùng lặp

Bước 2: Duyệt ma trận M at, nếu M at[i, j] == 1 thì thực hiện:

Bước 2.1: V ote = 0

Bước 2.2: Trên mỗi plane ch:

- Biến đổi DCT block[i, j, ch]

- Sử dụng phương trình (4) để tính ra bit b

- Nếu b == 1 thì:

V ote = V ote + 1

Bước 2.3: Nếu V ote > 1.5 thì:

w ′ (idx) = 1

Ngược lại:

w ′ (idx) = 0

Bước 3: Khôi phục w ′ sử dụng key wm thu được ảnh logo w ′

@ **Kết Luận**

Qua thí nghiệm và các kết quả thu được, việc áp dụng kỹ thuật ROI để xác định vùng dữ liệu ảnh theo lược đồ thủy vân đã đề xuất cho kết quả tương đối tốt. Đối với độ đo PSNR đều > 49db, các ảnh mang nhúng thủy vân đảm bảo tính trong khó nhận biết bằng phương pháp thông thường. Khi tấn công bằng các phép biến đổi ảnh phổ biến, việc trích rút thông tin đã nhúng cho kết quả cao đối với các phép nén, thay đổi màu, lọc trung bình, lọc trung vị, lọc Gaussian. Tuy nhiên, kết quả chưa được tốt khi thực hiện cắt ảnh. Do ở đây khi thực hiện cắt ¼ ảnh là tương đối lớn, nhưng vẫn có khả năng trích rút được một phần quan trọng của thông tin nhúng. Vì vậy, nếu sử dụng các kỹ thuật nhúng phân tán ảnh logo ra nhiều vị trí khác thì có thể cải thiện hiệu quả của kỹ thuật nhúng.

**TÀI LIỆU THAM KHẢO**

[1] A. Tefas, N. Nikolaidis, and I. Pitas, “The Essential Guide to Image Processing: Chapter 22 – Image Watermarking: Techniques and Applications”. 2st ed. Elsevier, pp. 597–648. 2009.

[2] Roy S, Pal AK, “An indirect watermark hiding in discrete cosine transform–singular value decomposition domain for copyright protection”. R. Soc. Open, sci.4:170326. 2017.

[3] Qasim, A.F., Aspin, R., Meziane, F. et al, “ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images”. Multimed Tools Appl 78, 16433–16463. 2019.

[4] Jeya, I. , Suganthi, J., “RONI Based Secured and Authenticated Indexing of Lung CT Images”. Computational and Mathematical Methods in Medicine. 2015. 1-9. 10.1155/2015/830453. 2015.

[5] Ta Minh Thanh, “Distributed Redundant Image Storages and Reconstruction Algorithm to Contents Verification”, The International Conference of National Foundation for Science and Technology Development (NAFOSTED). 2017.

[6] M. Saidi, H. Hermassi, R. Rhouma, and S. Belghith, “A new adaptive image steganography scheme based on DCT and chaotic map”. pp. 13493–13510, 2017. 2017.

[7] Ta Minh Thanh, Nguyen Trung Thanh, “Extended DCT domain for improving the quality of watermarked image”, The Seventh International Conference on Knowledge and Systems Engineering (KSE 2015). VIETNAM, 2015.

[8] D. Asatryan and M. Khalili, “Colour spaces effects on improved discrete wavelet transform-based digital image watermarking using Arnold transform map”. IET Signal Process., vol. 7, no. 3, pp. 177–187. 2013.

[9] A. K. Abdulrahman and S. Ozturk, “A novel hybrid DCT and DWT based robust watermarking algorithm for color images”. Multimed. Tools Appl. 2019.

[10] Katharotiya Anil, Patel Swati, Mahesh Goyani. “Comparative Analysis between DCT and DWT Techniques of Image Compression”. Journal of Information Engineering and Applications. 1. 9-17. 2011.

[11] Lai, C. C, “An improved SVD-based watermarking scheme using human visual characteristics”. Optics Communications, 284(4), 938–944, 2011.