MỤC LỤC

Trang

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH SÁCH CÁC TỪ VIẾT TẮT

BMP	Bitmap	Ảnh không nén Bitmap
DCT	Discrete Consine Transform	Phép biến đổi cosin rời rạc
GIF	Graphics Interchange Format	Định dạng ảnh đồ họa GIF
IMG	Image	Hình ảnh
JPEG	Joint Photographic Expert Group	Ảnh nén JPEG
LSBs	Least Significant Bits	Các bit ít quan trọng nhất
MSBs	Most Significant Bits	Các bit quan trọng
MSE	Mean squared error	Lỗi bình phương
PCX	Personal Computer Exchange	Ảnh xám PCX
PNG	Portable Network Graphics	Ảnh PNG
PSNR	Peak signal-to-noise ratio	Tỉ số tín hiệu cực đại trên nhiễu

LỜI MỞ ĐẦU

Với việc sử dụng internet để liên lạc ngày càng tăng, mối quan tâm chính đó là sự an toàn của truyền dữ liệu. Giấu tin mật là một nghệ thuật và khoa học về truyền thông vô hình. Nó ẩn thông tin mật trong các thông tin khác, do đó ẩn đi sự tồn tại của các thông tin truyền thông. Trong đồ án này em đã tìm hiểu một kỹ thuật giấu tin văn bản trong hình ảnh bằng cách sử dụng giấu tin mật trong hình ảnh. Kỹ thuật này sử dụng kết hợp giữa dữ liệu mật với các giá trị của điểm ảnh. Các bit có trọng số thấp của điểm ảnh được thay thế để đánh dấu sự hiện diện của dữ liệu bên trong điểm ảnh đó. Đối với việc lựa chọn các kênh để đánh dấu sự hiện diện của dữ liệu, một bộ tạo số giả ngẫu nhiên được sử dụng nên có thêm một lớp bảo mật cho kỹ thuật và làm cho việc khai thác các thông tin mật rất khó khăn cho những kẻ xâm nhập. Kết quả cho thấy kỹ thuật là bảo mật chống lại các cuộc tấn công trực quan, thống kê và cố gắng để có thể giấu nhiều dữ liệu hơn bằng cách sử dụng nhiều bit trên mỗi điểm ảnh.

Đồ án được tổ chức gồm bốn chương trong đó:

Chương 1. Tổng quan kỹ thuật giấu tin trong ảnh: Trình bày định nghĩa, mục đích, đặc trưng, tính chất, các phương pháp, mô hình giấu tin và phân loại các kỹ thuật giấu tin trong ảnh.

Chương 2. Cấu trúc chung của ảnh bitmap: Trình bày tổng quan về ảnh bitmap và cấu trúc của ảnh PNG.

Chương 3. Kỹ thuật giấu văn bản trong ảnh: Giới thiệu về kỹ thuật giấu văn bản trong ảnh và trình bày thuật toán giấu và tách văn bản trong ảnh.

Chương 4. Cài đặt thử nghiệm: Trình bày một số giao diện chính của chương trình và kết quả kiểm tra kỹ thuật giấu văn bản trong ảnh.

Chương 1. TỔNG QUAN KỸ THUẬT GIẤU TIN TRONG ẢNH

1.1. Định nghĩa giấu tin trong ảnh

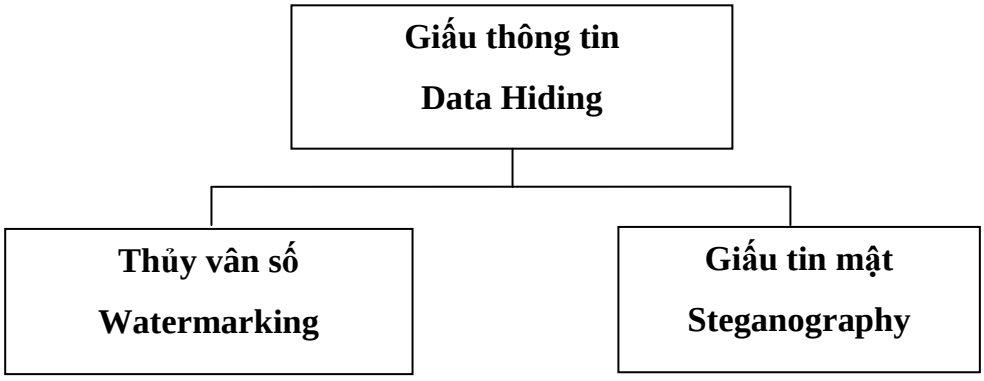
Giấu tin trong ảnh là một kỹ thuật giấu (nhúng) một lượng thông tin số nào đó vào trong một ảnh số [4].

1.2. Mục đích của giấu tin

Có 2 mục đích chính của giấu thông tin:

- Bảo mật cho những thông tin được giấu.
- Bảo mật cho chính các đối tượng giấu tin.

Có thể nhận thấy rằng sự khác biệt giữa hai mục đích. Trong thực tế hai mục đích này đã phát triển thành hai lĩnh vực với những yêu cầu và tính chất khác nhau.



Hình 1.1: Hai lĩnh vực chính của kỹ thuật giấu tin

➢ Kỹ thuật giấu tin mật (Steganography) [2]: Với mục đích đảm bảo an toàn và bảo mật thông tin được giấu. Các kỹ thuật giấu tin mật tập trung sao cho thông tin giấu được nhiều và người khác khó phát hiện ra thông tin có được giấu trong ảnh.

➢ Kỹ thuật thủy vân số (Watermarking): Với mục đích bảo mật cho chính các đối tượng giấu tin đánh dấu. Đảm bảo một số các yêu cầu như đảm bảo tính bền vững, khẳng định bản quyền sở hữu hay phát hiện xuyên tạc thông tin...

1.3. Các yêu cầu đối với giấu tin trong ảnh

Những yêu cầu cơ bản đối với giấu tin trong ảnh [1]:

- Tính ẩn của giấu tin được chèn vào ảnh: Sự hiện diện của giấu tin trong ảnh không làm ảnh hưởng tới chất lượng của ảnh đã chèn tin.

- Tính bền của giấu tin: Cho phép các tin có thể tồn tại được qua các phép biến đổi ảnh, biến dạng hình học hay các hình thức tấn công cố ý khác.
- Tính an toàn: không thể xoá được tin ra khỏi ảnh trừ khi ảnh được biến đổi tới mức không còn mang thông tin.

1.4. Đặc trưng và tính chất của kỹ thuật giấu tin trong ảnh

Giấu tin trong ảnh chiếm vị trí chủ yếu trong các kỹ thuật giấu tin. Các phương tiện chứa khác nhau thì cũng sẽ có các kỹ thuật giấu khác nhau. Đối tượng ảnh là một đối tượng dữ liệu tĩnh có nghĩa là dữ liệu tri giác không biến đổi theo thời gian. Dữ liệu ảnh có nhiều định dạng, mỗi định dạng có những tính chất khác nhau nên các kỹ thuật giấu tin trong ảnh thường chú ý những đặc trưng và các tính chất cơ bản sau đây:

- Phương tiện có chứa dữ liệu tri giác tĩnh

Dữ liệu gốc ở đây là dữ liệu tĩnh, dù đã giấu thông tin vào trong ảnh hay chưa thì khi ta xem ảnh bằng thị giác, dữ liệu ảnh không thay đổi theo thời gian, điều này khác với dữ liệu âm thanh và dữ liệu băng hình vì khi ta nghe hay xem thì dữ liệu gốc sẽ thay đổi liên tục với tri giác của con người theo các đoạn, các bài hay các cảnh…

- Kỹ thuật giấu phụ thuộc ảnh.

Kỹ thuật giấu tin phụ thuộc vào các loại ảnh khác nhau. Chẳng hạn đối với ảnh đen trắng, ảnh xám hay ảnh màu ta cũng có những kỹ thuật riêng cho từng loại ảnh.

- Kỹ thuật giấu tin lợi dụng tính chất hệ thống thị giác của con người

Giấu tin trong ảnh ít nhiều cũng gây ra những thay đổi trên dữ liệu ảnh gốc. Dữ liệu ảnh được quan sát bằng hệ thống thị giác của con người nên các kỹ thuật giấu tin phải đảm bảo một yêu cầu cơ bản là những thay đổi trên ảnh phải rất nhỏ sao cho bằng mắt thường khó nhận ra được sự thay đổi đó vì có như thế thì mới đảm bảo được độ an toàn cho thông tin giấu. Rất nhiều các kỹ thuật đã lợi dụng các tính chất của hệ thống thị giác để giấu tin chẳng hạn như mắt người cảm nhận về sự biến đổi về độ chói kém hơn sự biến đổi về màu hay cảm nhận của mắt về màu xanh da trời kém nhất trong ba màu cơ bản.

- Giấu thông tin trong ảnh tác động lên dữ liệu ảnh nhưng không thay đổi kích thước của hình ảnh.

Các thuật toán thực hiện công việc giấu tin sẽ được thực hiện trên dữ liệu của ảnh. Dữ liệu ảnh bao gồm phần header, bảng màu (có thể có) và dữ liệu ảnh. Do vậy mà kích thước ảnh trước hay sau khi giấu tin là như nhau.

- Đảm bảo chất lượng sau khi giấu tin

Đây là một yêu cầu quan trọng đối với giấu tin trong ảnh. Sau khi giấu tin bên trong, ảnh phải đảm bảo được yêu cầu không bị biến đổi để có thể bị phát hiện dễ dàng so với ảnh gốc. Yêu cầu này dường như khá đơn giản đối với ảnh màu hoặc ảnh xám bởi mỗi điểm ảnh được biểu diễn bởi nhiều bit, nhiều giá trị và khi ta thay đổi một giá trị nhỏ nào đó thì chất lượng ảnh thay đổi không đáng kể, thông tin giấu khó bị phát hiện, nhưng đối với ảnh đen trắng mỗi điểm ảnh chỉ là đen hoặc trắng, và nếu ta biến đổi một bit từ trắng thành đen và ngược lại mà không khéo thì sẽ rất dễ bị phát hiện. Do đó, yêu cầu đối với các thuật toán giấu thông tin trong ảnh màu hay ảnh xám và giấu thông tin trong ảnh đen trắng là khác nhau. Trong khi đối với ảnh màu thì các thuật toán chú trọng vào việc làm sao giấu được càng nhiều thông tin càng tốt thì các thuật toán áp dụng cho ảnh đen trắng lại tập trung vào việc làm thế nào để thông tin giấu khó bị phát hiện nhất.

- Thông tin trong ảnh sẽ bị biến đổi nếu có bất cứ biến đổi nào trên ảnh

Vì phương pháp giấu thông tin trong ảnh dựa trên việc điều chỉnh các giá trị của các bit theo một quy tắc nào đó và khi giải mã sẽ theo các giá trị đó để tìm được thông tin giấu. Theo đó, nếu một phép biến đổi nào đó trên ảnh làm thay đổi giá trị của các bit thì sẽ làm cho thông tin giấu bị sai lệch. Nhờ đặc điểm này mà giấu thông tin trong ảnh có tác dụng nhận thực và phát hiện xuyên tạc thông tin.

- Vai trò của ảnh gốc khi tách tin

Các kỹ thuật giấu tin phải xác định rõ ràng quá trình lọc ảnh để lấy thông tin giấu cần đến ảnh gốc hay không cần. Đa số các kỹ thuật giấu tin mật thì thường không cần ảnh gốc để giải mã. Thông tin được giấu trong ảnh sẽ được mang cùng với dữ liệu

ảnh, khi giải mã chỉ cần ảnh đã mang thông tin giấu mà không cần dùng đến ảnh gốc để so sánh đối chiếu.

1.5. Các phương pháp giấu tin

- Các phương pháp giấu tin trong ảnh hiện nay thuộc một trong ba nhóm [4]:

- Giấu tin trong miền không gian.

Phương pháp này thường nhúng thông tin vào các bit có trọng số thấp của ảnh hay được áp dụng trên các ảnh bitmap không nén, các ảnh dùng bảng màu. Ý tưởng chính của phương pháp này là lấy từng bit của tin mật rải nó lên ảnh gốc và thay đổi bit có trọng số thấp của ảnh bằng các bit của tin mật. Vì khi thay đổi các bit có trọng số thấp không ảnh hưởng đến chất lượng ảnh, và mắt người không cảm nhận được sự thay đổi của ảnh đã giấu tin.

- Các phương pháp dựa vào kỹ thuật biến đổi ảnh, ví dụ biến đổi từ miền không gian sang miền tần số.
- Các phương pháp sử dụng mặt nạ giác quan.

Dựa trên nguyên lý đánh lừa hệ thống giác quan của con người. "Mặt nạ" ở đây ám chỉ hiện tượng mắt người không cảm nhận được một tín hiệu nếu nó ở bên cạnh một tín hiệu nhất định nào đó.

- Nếu phân chia các phương pháp theo định dạng ảnh thì có hai nhóm chính:

- Nhóm phương pháp phụ thuộc định dạng ảnh: đặc điểm của nhóm này là thông tin giấu dễ bị "tổn thương" bởi các phép biến đổi ảnh. Trong nhóm này lại chia ra theo dạng ảnh, có các phương pháp cho: ảnh dựa vào bảng màu; ảnh JPEG.
- Các phương pháp độc lập với định dạng ảnh: đặc trưng của các phương pháp nhóm này là lợi dụng vào việc biến đổi ảnh để giấu tin vào trong đó, ví dụ giấu vào các hệ số biến đổi. Như vậy có bao nhiêu phép biến đổi ảnh thì cũng có thể có bấy nhiêu phương pháp giấu ảnh. Các phép biến đổi như:
  - Phương pháp biến đổi theo miền không gian
  - Phương pháp biến đổi theo miền tần số (DCT)
  - Các biến đổi hình học

Các phương pháp nhóm thứ hai có nhiều ưu điểm hơn về tính bền vững, nhưng lượng thông tin giấu được sẽ ít hơn và cài đặt cũng sẽ phức tạp hơn.

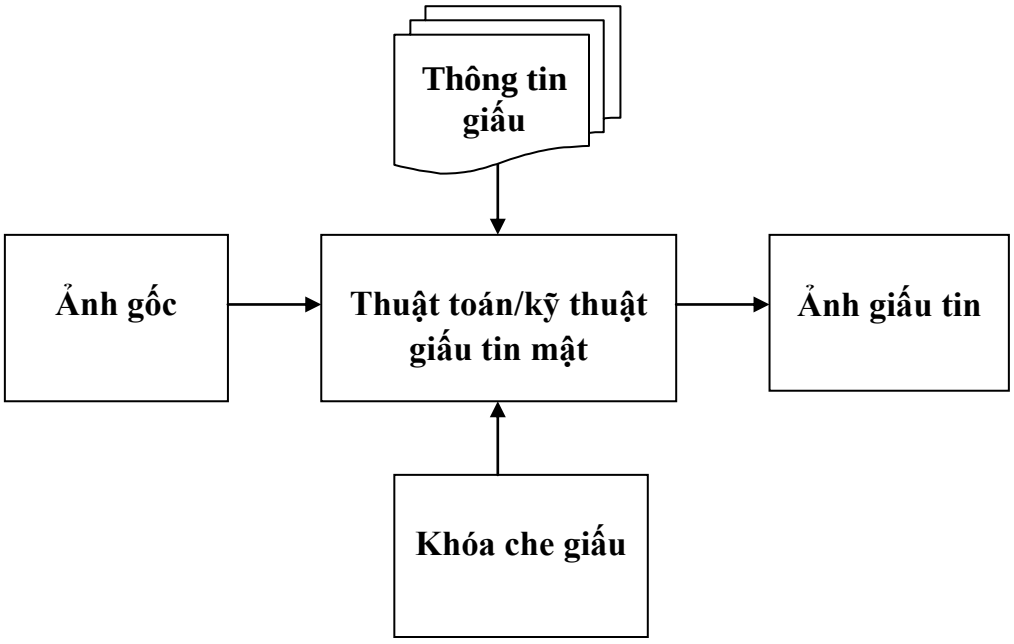
➢ Nếu phân chia các phương pháp theo đặc điểm kỹ thuật có:

- Phương pháp thay thế.
  - Thay thế các bit dữ liệu trong bản đồ bit.
  - Thay thế bảng màu.
- Phương pháp xử lý tín hiệu
  - Các phương pháp biến đổi ảnh.
  - Các kỹ thuật điều chế trải phổ.
- Các phương pháp mã hoá: Lượng hóa; mã hóa sửa lỗi.
- Các phương pháp thống kê - kiểm thử giả thuyết
- Phương pháp sinh mặt nạ.

1.6. Mô hình kỹ thuật giấu tin trong ảnh.

Kỹ thuật giấu tin trong ảnh bao gồm hai quá trình:

Quá trình 1: Giấu (nhúng) tin vào ảnh.



Hình 1.2: Mô hình cơ bản giấu tin mật trong ảnh.

Đầu vào:

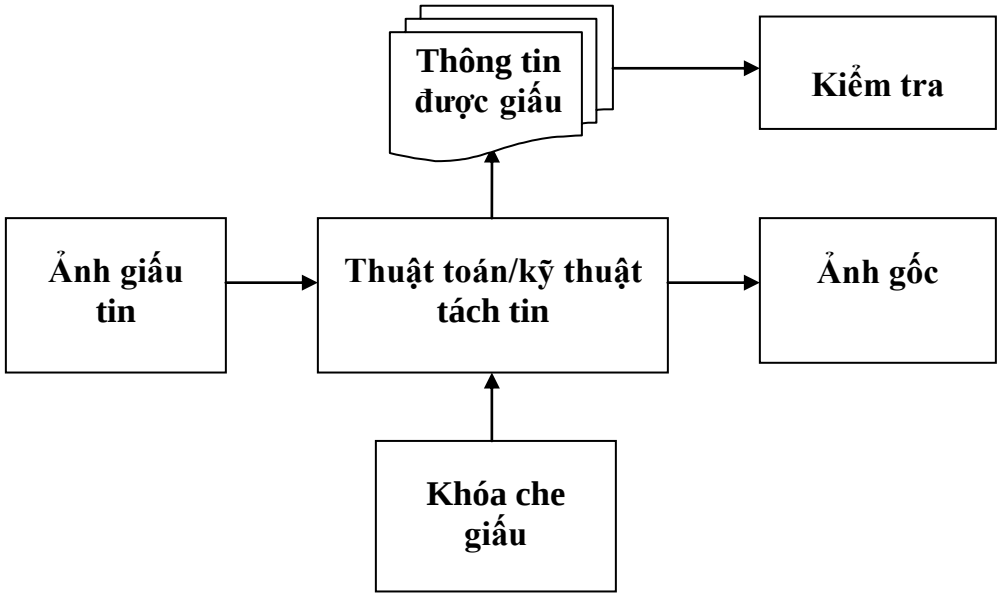
- Thông tin giấu: Tùy theo mục đích của người sử dụng mà thông tin giấu ở đây có thể là thông điệp, hình ảnh, video, âm thanh...

- Ảnh gốc: Là ảnh được chọn làm môi trường để giấu tin.

Đầu ra:

- Ảnh giấu đã được giấu tin

Quá trình 2: Tách tin từ ảnh giấu tin



Hình 1.3: Mô hình cơ bản tách tin mật

Đầu vào:

- Ảnh giấu tin.
- Khóa che giấu.

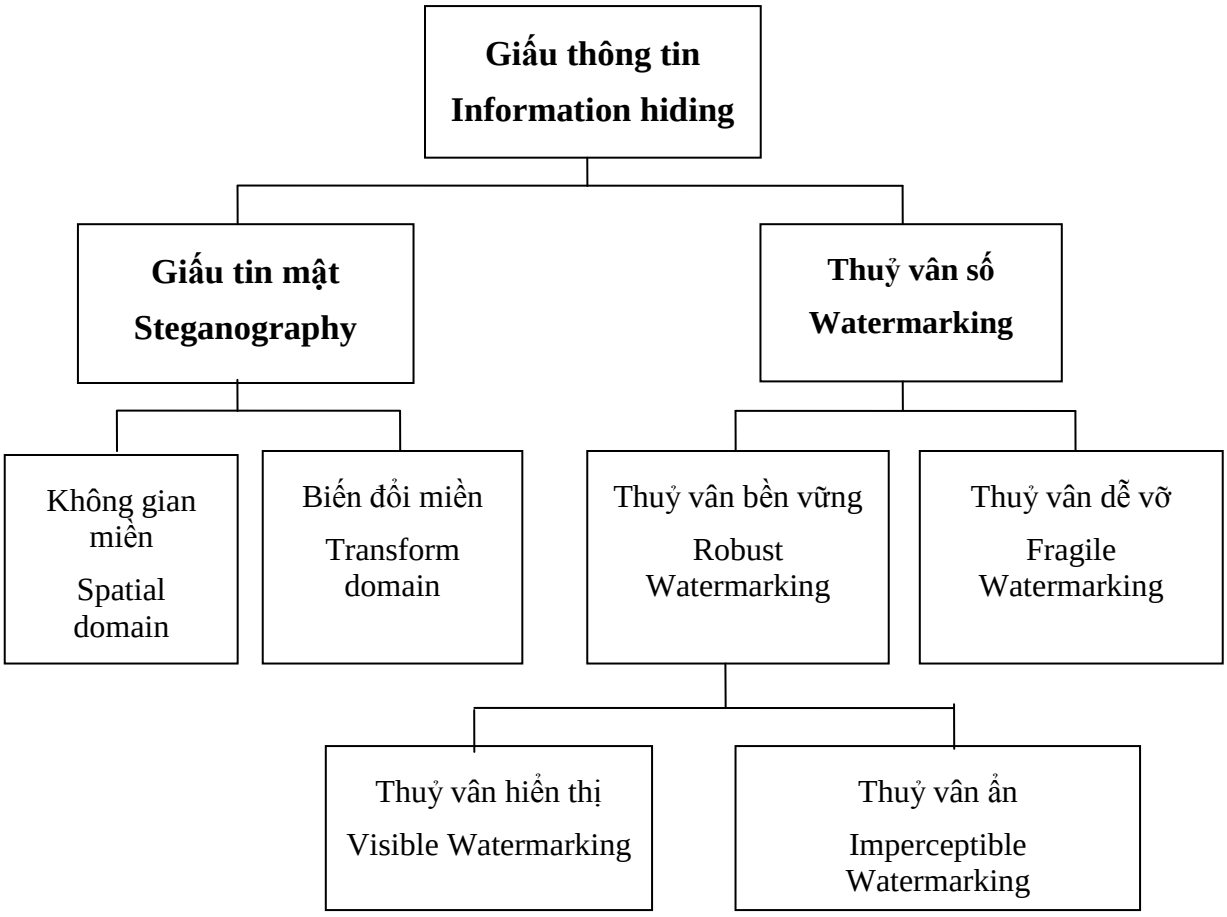
Đầu ra:

- Thông tin được giấu.
- Ảnh gốc.

Quá trình giải mã được thực hiện thông qua thuật toán/kỹ thuật tách tin tương ứng với thuật toán/kỹ thuật nhúng tin cùng với khoá che giấu của quá trình nhúng. Kết quả thu được gồm ảnh gốc và thông tin đã giấu. Thông tin đã giấu được kiểm tra so sánh với thông tin ban đầu.

1.7. Phân loại các kỹ thuật giấu tin trong ảnh

Có thể chia kỹ thuật giấu tin ra làm 2 loại lớn đó là thủy vân số và giấu tin mật.



Hình 1.4. Phân loại các kỹ thuật giấu tin

1.7.1. Giấu tin mật

Giấu tin mật có thể được định nghĩa là kỹ thuật để nhúng dữ liệu hoặc thông tin mật trong đối tượng gốc. Mục đích của giấu tin mật là thiết lập một đường truyền thông bí mật giữa hai bên, như vậy bất kỳ người nào ở giữa cũng không thể phát hiện sự tồn tại của dữ liệu - thông tin mật. Những kẻ tấn công không lấy được bất kỳ thông tin nào về dữ liệu – thông tin nhúng bằng cách nhìn đơn giản vào tập tin.

Ngày nay giấu tin được thực hiện bằng cách sử dụng phương tiện kỹ thuật số như văn bản, hình ảnh, âm thanh, video hoặc các phương tiện khác tùy thuộc vào yêu cầu và lựa chọn của người gửi. Trong số các phương tiện để giấu tin thì giấu tin mật

trong hình ảnh được sử dụng rộng rãi nhất. Vì hiện nay số thông tin dư thừa trong hình ảnh là lớn để có thể dễ dàng thay đổi và ẩn được nhiều thông tin mật bên trong hơn.

Một số kỹ thuật được đề xuất sử dụng tập tin hình ảnh làm đối tượng gốc. Những kỹ thuật này có thể được phân loại theo hai cách sau đây:

- Kỹ thuật không gian miền.
- Kỹ thuật thay đổi miền.

▪ Kỹ thuật không gian miền: Các thuật toán thuộc kỹ thuật không gian miền nhúng dữ liệu bằng cách lựa chọn thay thế một cách cẩn thận từ các điểm ảnh của hình ảnh gốc với các bit thông điệp mật. Các thuật toán giấu tin mật tốt nhất được biết đến là dựa trên việc sửa đổi lớp ít quan trọng của hình ảnh, do đó còn được gọi là kỹ thuật LSB. Kỹ thuật LSB được sử dụng rộng rãi nhất của giấu tin mật trong hình ảnh. Trong kỹ thuật này các bit ở các điểm ảnh gốc có trọng số thấp được thay thế bằng các bit tin.

▪ Kỹ thuật thay đổi miền: Các thuật toán thuộc kỹ thuật thay đổi miền nhúng dữ liệu bằng cách thay đổi miền hình ảnh gốc và sau đó giấu dữ liệu vào bên trong chúng. Thuật toán DCT là một trong những thuật toán thường được sử dụng chuyển đổi miền cho thể hiện ra dưới một dạng sóng như là một tổng hợp có trọng số của cosin. Các dữ liệu được giấu bằng cách thay đổi hệ số DCT của hình ảnh.

Một kỹ thuật giấu tin mật trong hình ảnh tốt nhằm ba mục tiêu

- Dữ liệu tối đa có thể được giấu bên trong hình ảnh.
- Tính không nhận thấy được tin giấu: tức là chất lượng của hình ảnh sau khi giấu tin. Bằng cách nhìn vào ảnh che giấu cũng không nhận thấy được hình ảnh có giấu tin.
- Bảo mật: An ninh phải mạnh mẽ để chống lại các cuộc tấn công của những kẻ tấn công.

1.7.2. Thủy vân số

Không cần giấu nhiều thông tin, chỉ cần lượng thông tin nhỏ đặc trưng cho bản quyền của người sở hữu, nhưng đòi hỏi độ bền vững cao của thông tin cần giấu.

Thủy vân bền vững: thường được ứng dụng trong bảo vệ bản quyền. Thủy vân được nhúng trong sản phẩm như một hình thức dán tem bản quyền. Trong trường hợp

này, thủy vân phải tồn tại bền vững cùng với sản phẩm nhằm chống việc tẩy xóa, làm giả hay biến đổi phá hủy thủy vân.

Thủy vân dễ vỡ: Là kỹ thuật nhúng thủy vân vào trong một đối tượng (sản phẩm) sao cho khi phân bố sản phẩm nếu có bất kỳ phép biến đổi nào làm thay đổi sản phẩm gốc thì thủy vân đã được giấu trong đối tượng sẽ không còn nguyên vẹn như trước khi giấu.

Thủy vân ẩn: Cũng giống như giấu tin, bằng mắt thường không thể nhìn được thủy vân ẩn.

Thủy vân hiện: Là loại thủy vân hiện ngay trên sản phẩm và mọi người đều có thể nhìn thấy được.

Chương 2. CẤU TRÚC CHUNG CỦA ẢNH BITMAP

2.1. Tổng quan về ảnh Bitmap

Để thực hiện việc giấu tin trong ảnh, trước hết ta phải nghiên cứu cấu trúc của ảnh và có khả năng xử lý được ảnh tức là phải số hoá ảnh. Quá trình số hoá các dạng ảnh khác nhau và không như nhau. Có nhiều loại ảnh đã được chuẩn hoá như: JPEG, PCX, BMP, GIF, IMG… Sau đây là cấu trúc ảnh bitmap (*.BMP)

Mỗi file ảnh BMP gồm 3 phần:

- BitmapHeader (54 byte)
- Palette màu (bảng màu)
- BitmapData (dữ liệu ảnh)

Cấu trúc cụ thể của ảnh:

- BitmapHeader (54 byte): Lưu trữ những thông tin cơ bản về tệp ảnh và thuộc tính cơ bản của ảnh.

Bảng 2.1 Bảng chi tiết những thông tin trong BitmapHeader.

Byte	Đặt tên	Ý nghĩa	Giá trị
1 - 2	ID	Nhận dạng file	‘BMP’ hay 19778
3 – 6	File_Size	Kích thước File	Kiểu Long trong turbo C
7 – 10	Reserved	Dành riêng	Mang giá trị 0
11 – 14	OffsetBit	Byte bắt đầu vùng dữ liệu	Offset của byte bắt đầu vùng dữ liệu
15 -18	Isize	Số byte cho vùng info	40 byte
19 – 22	Width	Chiều rộng của ảnh BMP	Tính bằng pixel
23 – 26	Height	Chiều cao của ảnh BMP	Tính bằng pixel
27 – 28	Planes	Số planes màu	Cố định là 1
29 – 30	bitCount	Số bit cho một pixel	Có thể là 1,4,6,16,24

Byte	Đặt tên	Ý nghĩa	Giá trị
31-34	Compression	Kiểu nén dữ liệu	0: Không nén 1: Nén 8bits/pixel 2: Nén 4bits/pixel
35 -38	ImageSize	Kích thước ảnh	Tính bằng byte
39 – 42	XpelsPerMeter	Độ phân giải ngang	Tính bằng pixel/metr
43 – 46	YpelsPerMeter	Độ phân giải dọc	Tính bằng pixel/metr
47 – 50	ColorsUsed	Số màu sử dụng trong ảnh	
51 – 54	ColorsImportant	Số màu được sử dụng khi hiện ảnh	

Thành phần bitCount của cấu trúc BitmapHeader cho biết số bit dành cho mỗi điểm ảnh và số lượng màu lớn nhất của ảnh. BitCount có thể nhận các giá trị sau:

1: Bitmap là ảnh đen trắng, mỗi bit biểu diễn 1 điểm ảnh. Nếu bit mang giá trị 0 thì điểm ảnh là đen, bit mang giá trị 1 điểm ảnh là điểm trắng.

4: Bitmap là ảnh 16 màu, mỗi điểm ảnh được biểu diễn bởi 4 bit.

8: Bitmap là ảnh 256 màu, mỗi điểm ảnh biểu diễn bởi 1 byte.

16: Bitmap là ảnh highcolor, mỗi dãy 2 byte liên tiếp trong bitmap biểu diễn cường độ tương đối của màu đỏ, xanh lá cây, xanh lơ của một điểm ảnh.

24: Bitmap là ảnh true color (2^{24} màu), mỗi dãy 3 byte liên tiếp trong bitmap biểu diễn cường độ tương đối của màu đỏ, xanh lá cây, xanh lơ (RGB) của một điểm ảnh.

Thành phần ColorUsed của cấu trúc BitmapHeader xác định số lượng màu của palette màu thực sự được sử dụng để hiển thị bitmap. Nếu thành phần này được đặt là 0, bitmap sử dụng số màu lớn nhất tương ứng với giá trị của BitCount.

- Palette màu (bảng màu): bảng màu của ảnh, chỉ những ảnh lớn hơn hoặc bằng 8 bit màu mới có Palette màu.

- BitmapData (thông tin ảnh): phần này nằm ngay sau phần palette màu của ảnh BMP. Đây là phần chứa giá trị màu của điểm ảnh trong ảnh BMP, các dòng ảnh được lưu từ dưới lên trên, các điểm ảnh được lưu từ trái sang phải. Giá trị của mỗi điểm ảnh là một chỉ số trỏ tới phần tử màu tương ứng của palette màu.

2.2. Cấu trúc ảnh PNG

Là một dạng hình ảnh sử dụng phương pháp nén dữ liệu mới – không làm mất đi dữ liệu gốc. PNG được hỗ trợ bởi thư viện tham chiếu libpng, một thư viện nền độc lập bao gồm các hàm của C để quản lý các hình ảnh PNG.

Một tập tin PNG bao gồm 8 – byte kí hiệu (89 50 4E 47 0D 0A 1A) được viết trong hệ thống có cơ số 16, chứa các chữ "PNG" và 2 dấu xuống dòng, ở giữa là xếp theo số lượng của các thành phần, mỗi thành phần đều chứa thông tin về hình ảnh. Cấu trúc dựa trên các thành phần được thiết kế cho phép định dạng PNG có thể tương thích với các phiên bản cũ khi sử dụng các "thành phần" trong tập tin.

PNG là cấu trúc như một chuỗi các thành phần, mỗi thành phần chứa kích thước, kiểu, dữ liệu, và mã sửa lỗi CRC ngay trong nó.

Chuỗi được gán tên bằng 4 chữ cái phân biệt chữ hoa chữ thường. Sự phân biệt này giúp bộ giải mã phát hiện bản chất của chuỗi khi nó không nhận dạng được.

Với chữ cái đầu, viết hoa thể hiện chuỗi này là cần thiết. Chuỗi này chứa thông tin cần thiết để đọc được tệp và nếu bộ giải mã không nhận dạng được chuỗi này việc đọc tệp được hủy.

Về cơ bản, định dạng PNG đem lại cho ta những ưu thế vượt trội hơn so với các định dạng phổ thông khác hiện nay như JPG, GIF, BMP…Những ưu thế tỏ rõ sức mạnh hơn khi được sử dụng trong môi trường đồ họa web.

- Giảm thiểu dung lượng: Trong tất cả các định dạng ảnh phổ thông hiện nay thì hình ảnh PNG có thể coi là dung lượng nhỏ nhất. Điều này rất quan trọng khi sử dụng PNG trong môi trường web.
- Độ sâu của màu: Ảnh PNG hỗ trợ đến true color 48bit màu. Trong khi đó ảnh gif chỉ ở mức 256 màu.

Chương 3. KỸ THUẬT GIẤU VĂN BẢN TRONG ẢNH SỐ

3.1. Giới thiệu.

Kỹ thuật giấu tin LSB là một kỹ thuật đáp ứng được tính bảo mật và giấu được nhiều dữ liệu hơn. Tuy nhiên hiện tại các kỹ thuật không sử dụng hết dữ liệu của ảnh gốc. Nhiều kỹ thuật đã được phát triển để sử dụng số lượng nhiều hơn và nhiều hơn nữa các bit/pixel để đạt được nhiều dữ liệu ẩn hơn [5-10]. Tháng 10/2010 hai tác giả Sukhpreet và Sumeet [3] đã đề xuất kỹ thuật giấu văn bản trong hình ảnh bằng cách sử dụng 7 bit/pixel để ẩn dữ liệu mà không có sự thay đổi của hình ảnh giấu tin. Văn bản được chuyển đổi thành mã ASCII và sau đó 7 bit mã ASCII của mỗi tin là lần xuất hiện với các giá trị điểm ảnh của ảnh gốc. Để đánh dấu sự có mặt của dữ liệu trong một điểm ảnh cụ thể tác giả đã sử dụng kỹ thuật LSB.

3.2. Kỹ thuật giấu văn bản trong ảnh.

Trong kỹ thuật này tác giả đã sử dụng hình ảnh Bitmap làm hình ảnh gốc. Bởi vì ảnh Bitmap có lượng thông tin dư thừa là lớn, nó dễ dàng thay đổi để nhúng được nhiều tin mật vào bên trong ảnh mà không có sự khác biệt nào của hình ảnh giấu tin. Điều này nhằm đáp ứng mục đích của giấu tin mật trong hình ảnh.

Như chúng ta đã biết mỗi pixel của hình ảnh BMP có ba byte: một cho Red, một cho Green, một cho Blue. Mỗi ký tự của văn bản sẽ được chuyển đổi sang mã ASCII tương ứng với 7 bit nhị phân. Với mỗi điểm ảnh ta so sánh lần lượt 7 bit của tin ẩn với 7 bit có trọng số cao (MBSs - Most Significant Bits) của cả ba kênh màu Red, Green, Blue. Nếu 7 bit của tin mật bằng 7 bit MBSs của một trong ba kênh thì ta sử dụng các bit LSBs để đánh dấu dữ liệu có trong điểm ảnh. Ta sẽ sử dụng hai kênh màu làm kênh chỉ báo. Và LSB của hai kênh đó sẽ đánh dấu sự có mặt của dữ liệu có trong bất kỳ một trong ba kênh màu.

Để quyết định các kênh sẽ hoạt động như các kênh chỉ báo thì đối với mỗi ký tự của tin mật, ta tạo ra một số giả ngẫu nhiên. Chuyển đổi số giả ngẫu nhiên thành chuỗi bit nhị phân. Đếm số lượng số 1 và số lượng số 0 và xác định tính chẵn lẻ của chuỗi bit. Tùy thuộc vào chuỗi bit nhị phân của số giả ngẫu nhiên ta sẽ chọn được một trường hợp dùng để thiết lập các kênh chỉ báo. Việc lựa chọn được hiển thị trong bảng 3.1.

Bảng 3.1. Tiêu chuẩn lựa chọn kênh chỉ báo

Trường hợp	Đặt chỉ báo kênh	Loại 1 (tính chẵn lẻ là chẵn)	Loại 2 (tính chẵn lẻ là lẻ)
Số các số 1 nhiều hơn số các số 0	RG	RG	GR
Số các số 0 nhiều hơn số các số 1	GB	GB	BG
Số các số 0 là bằng số các số 1	RB	RB	BR

Đầu tiên chúng ta ẩn chiều dài của tin trong hàng đầu tiên của ảnh sử dụng kỹ thuật LSB. Sau đó bắt đầu từ điểm ảnh đầu tiên của hàng thứ hai ta so sánh ký tự đầu tiên của tin với 7 MSBs của cả ba thành phần Red, Green, Blue của điểm ảnh. Nếu có tương ứng với bất kỳ kênh màu nào thì giá trị của các kênh chỉ báo được thiết lập theo các tiêu chí được đưa ra trong bảng 3.2. Thủ tục tương tự được lặp lại với các điểm ảnh kế tiếp của hình ảnh gốc.

Bảng 3.2. Tiêu chuẩn để đặt giá trị kênh chỉ báo

Kênh dữ liệu	Chỉ báo LSB 1	Chỉ báo LSB 2
Kênh Red	0	0
Kênh Green	0	1
Kênh Blue	1	0
Không có kênh	1	1

3.3. Thuật toán giấu văn bản trong ảnh

Đầu vào:

- Ảnh Bitmap 24bit.
- Văn bản cần giấu.

Đầu ra:

- Ảnh giấu tin.

Các bước thực hiện thuật toán:

Bước 1: Đọc ảnh gốc. Đọc tin mật, chuyển đổi thành mã ASCII. Lưu vào mảng C.

Bước 2: Tính chiều dài tin lưu nó vào biến L. Giấu chiều dài tin vào hàng đầu tiên của ảnh gốc bằng cách sử dụng kỹ thuật LSB.

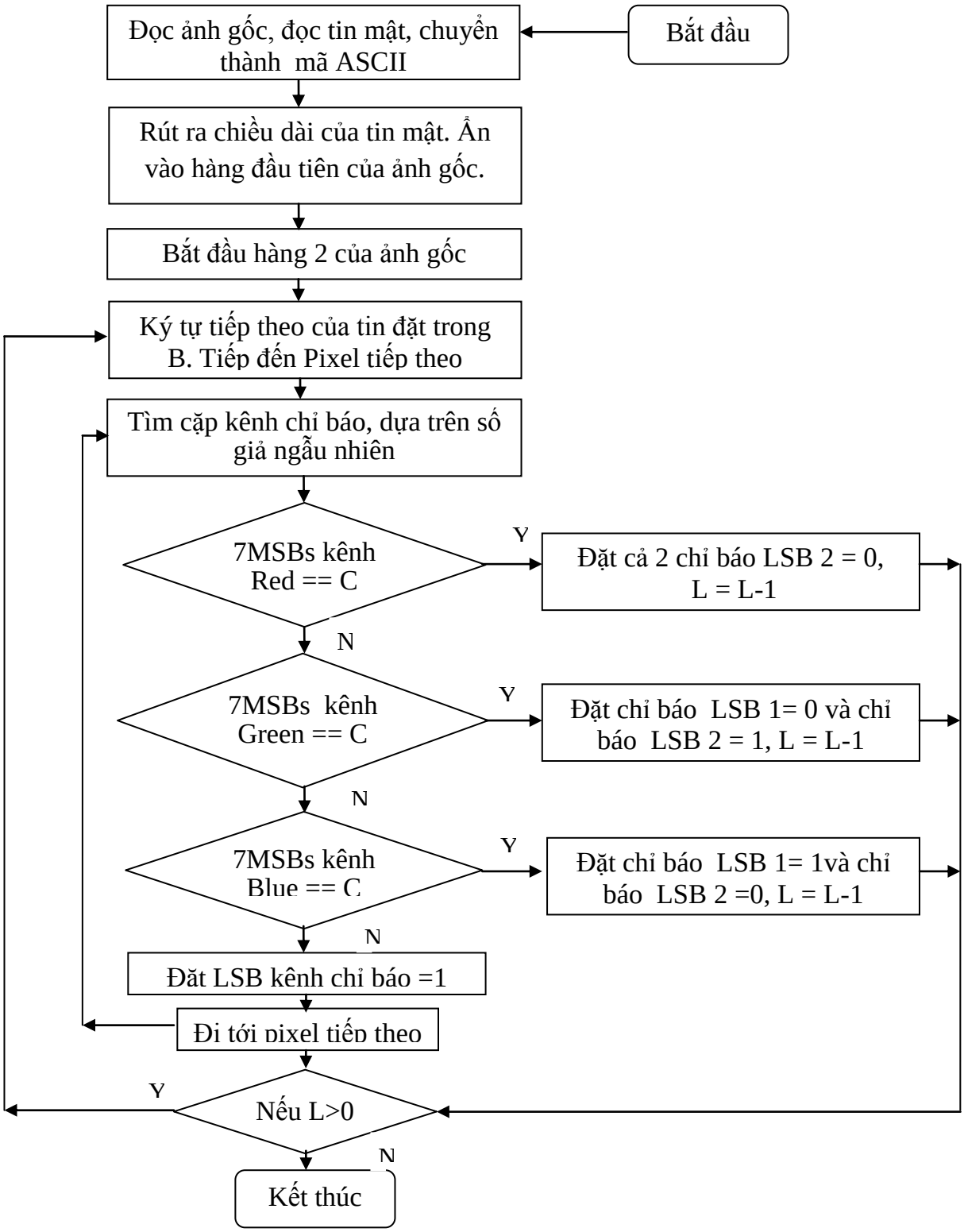
Bước 3: Bắt đầu từ điểm ảnh đầu tiên của hàng thứ hai ảnh gốc.

Bước 4: Ký tự rút ra từ C lưu trữ trong một biến tạm thời B.

Bước 5: Chọn cặp kênh chỉ báo, phụ thuộc vào số giả ngẫu nhiên.

Bước 6: Nếu ký tự tương ứng với 7 MSB của một trong ba kênh. Thiết lập giá trị của các kênh chỉ báo. Đặt L = L-1. Và đi đến điểm ảnh kế tiếp. Ngược lại nếu ký tự không phù hợp với bất kỳ kênh nào, đặt giá trị của các kênh chỉ báo bằng 1. Đi tới điểm ảnh tiếp theo và đi đến bước 5.

Bước 7: Kiểm tra xem L có lớn hơn 0 hay không. Nếu có đi đến bước 4. Ngược lại kết thúc thuật toán.



Hình 3.1: Sơ đồ quá trình giấu tin

3.4. Thuật toán tách văn bản trong ảnh

Đầu vào:

- Ảnh giấu tin.

Đầu ra:

- Ảnh gốc
- Văn bản được giấu.

Quá trình giấu tin sẽ phụ thuộc vào giá trị của hàm tạo lập số giả ngẫu nhiên. Hàm tạo lập số giả ngẫu nhiên sẽ sinh ra các số giống như nó được tạo ra ở quá trình giấu tin. Tùy thuộc vào giá trị số giả ngẫu nhiên mà ta sẽ tìm ra được kênh nào là kênh chỉ báo. Tùy thuộc vào giá trị của các kênh chỉ báo chúng ta sẽ tìm ra được dữ liệu nằm trong kênh màu nào bằng cách sử dụng bảng 3.2.

Các bước thực hiện thuật toán:

Bước 1: Đọc hình ảnh giấu tin. Rút ra các LSB của hàng đầu tiên để tìm ra L.

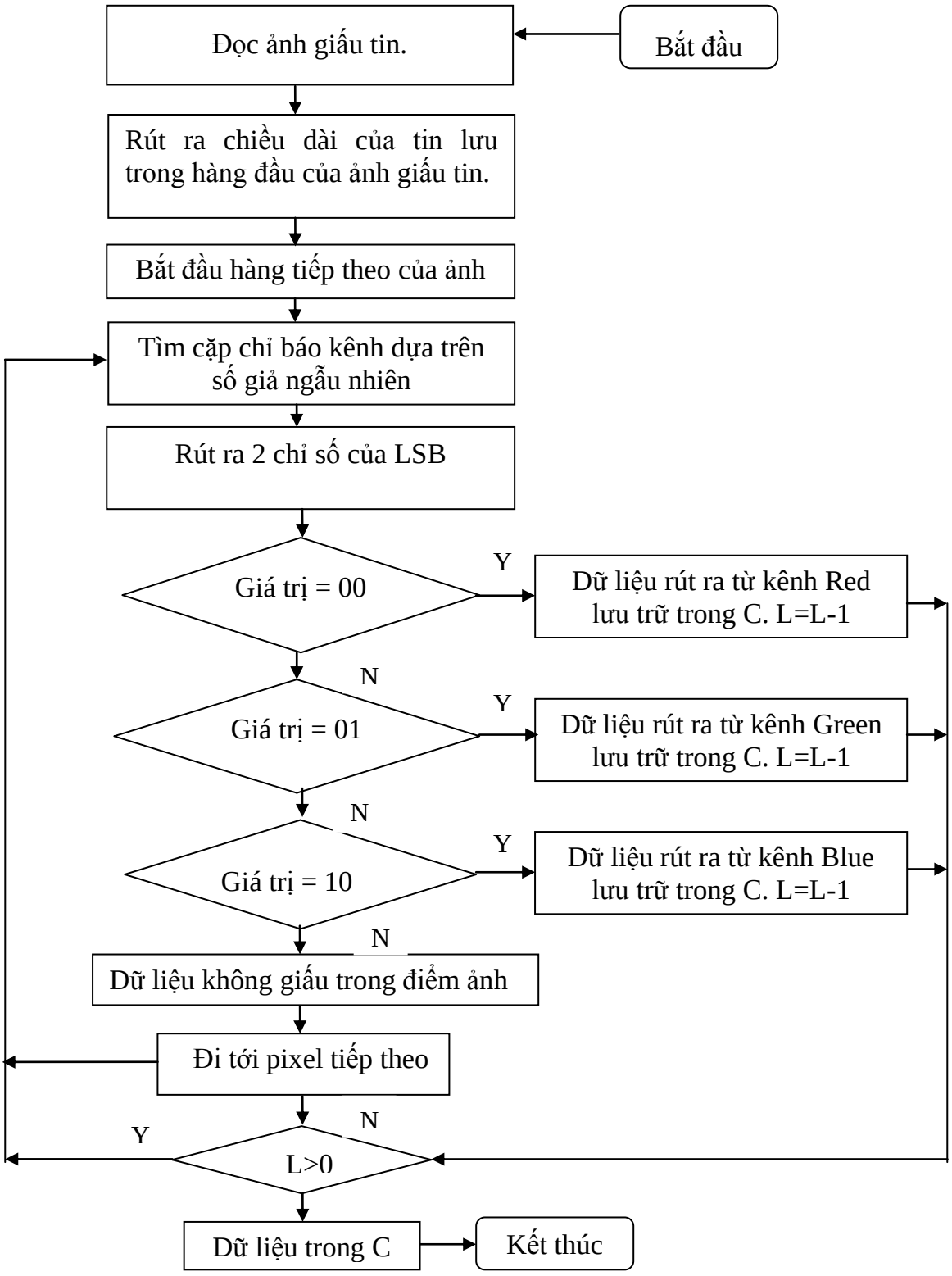
Bước 2: Bắt đầu từ điểm ảnh đầu tiên hàng thứ hai của hình ảnh che giấu.

Bước 3: Tìm cặp kênh chỉ báo, phụ thuộc vào số giả ngẫu nhiên. Và rút ra chỉ báo LSB1 và chỉ báo LSB2.

Bước 4: Nếu giá trị LSB1và LSB2 là 11 thì dữ liệu không tồn tại trong điểm ảnh. Ngược lại tùy thuộc vào giá trị của các kênh chỉ báo, rút ra dữ liệu từ các điểm ảnh lưu vào mảng C. Đặt L = L-1 và đi đến điểm ảnh kế tiếp.

Bước 5: Kiểm tra xem L có lớn hơn 0 hay không. Nếu có đi đến bước 4. Ngược lại thì dừng.

Bước 6: Chuyển đổi các giá trị trong C thành các ký tự tương đương.



Hình 3.2: Sơ đồ quá trình tách tin.

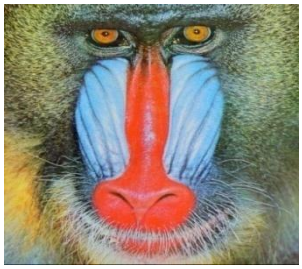
Chương 4. CÀI ĐẶT VÀ THỬ NGHIỆM

4.1. Môi trường cài đặt

- Hệ điều hành Windown XP/Vista/7.
- Cấu hình: bộ vi xử lý Core Duo trở lên, Ram 1Gb..
- Ngôn ngữ cài đặt: ngôn ngữ lập trình Matlab.
- Môi trường soạn thảo: Matlab R2008b.

4.2. Tập dữ liệu thử nghiệm

Tập dữ liệu thử nghiệm gồm các hình ảnh bitmap 24 bit được lấy trên mạng với kích thước khác nhau:



bamboon.bmp

lena.bmp

banh.bmp

banhoa.bmp

hoa1.bmp

hoa2.bmp

vuonhoa.bmp

nen.bmp

conmeo.bmp

bonghoa.bmp

Hình 4.1. Tập hình ảnh thử nghiệm

4.3. Đo độ đánh giá PSNR.

Chất lượng ảnh sau khi giấu tin được đánh giá thông qua giá trị của tỷ số PSNR (Peak Signal to Noise Ratio) tỷ số tín hiệu đỉnh trên nhiễu.

PSNR được định nghĩa thông qua lỗi bình phương (MSE - Mean squared error) cho hai hình ảnh bitmap I và K tương ứng là hình ảnh gốc và hình ảnh giấu tin có kích thước m x n. Do ảnh được sử dụng trong kỹ thuật là ảnh bitmap với 3 kênh màu RGB nên công thức MSE được tính như dưới đây :

$$MSER = (\frac{1}{mR * nR} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [R(i,j) - KR(i,j)]^2)$$

$$MSEG = (\frac{1}{mR * nR} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [G(i,j) - KG(i,j)]^2)$$

$$MSEB = (\frac{1}{mR * nR} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [B(i,j) - KB(i,j)]^2)$$

$$MSE = (MSER + MSEG + MSEB) * \frac{1}{3 * m * n}$$

PSNR được tính như sau :

$$PSNR = 10.\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20.\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

Ở đây, MAX_I là giá trị tối đa của pixel trên ảnh. Khi các pixels được biểu diễn bởi 8 bits, thì giá trị của nó là 255. Trường hợp tổng quát, khi tín hiệu được biểu diễn bởi B bits trên một đơn vị lấy mẫu, MAX_I là $2^B - 1$.

Khi hai hình ảnh giống hệt nhau, MSE sẽ bằng 0 và PSNR đi đến vô hạn

4.4. Một số giao diện của chương trình

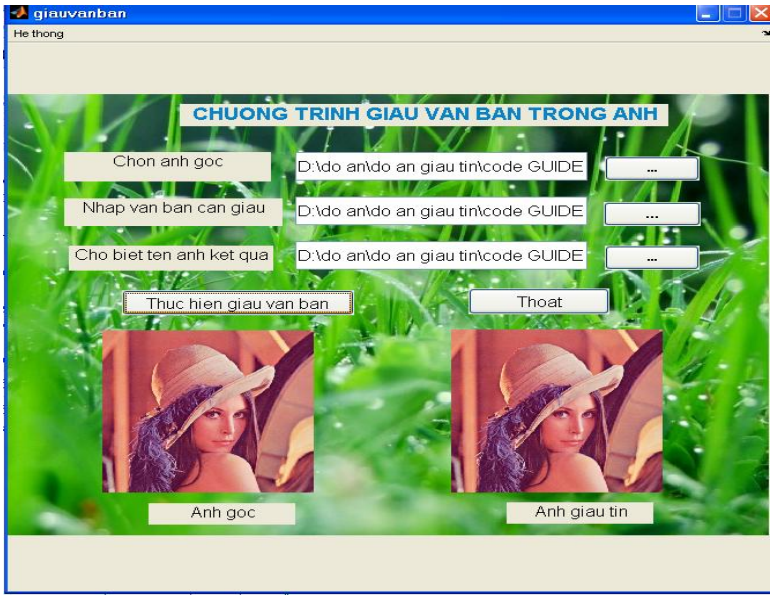
- Giao diện chính

Hình 4.2. Hình ảnh giao diện chính

Giao diện chính gồm các chức năng:

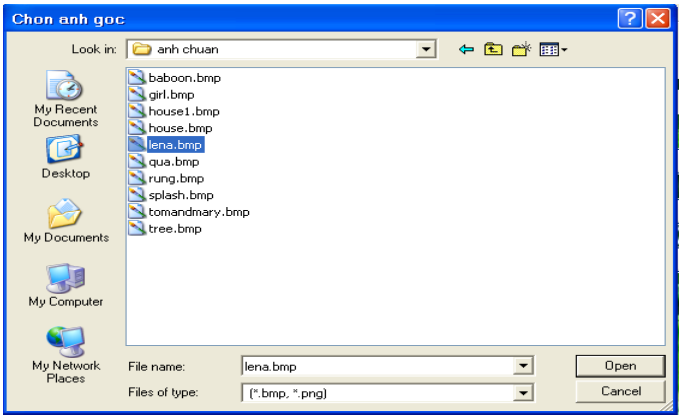
- Hệ thống
- Giấu văn bản
- Tách văn bản
- Kiểm tra PSNR

- Giao diện giấu văn bản:

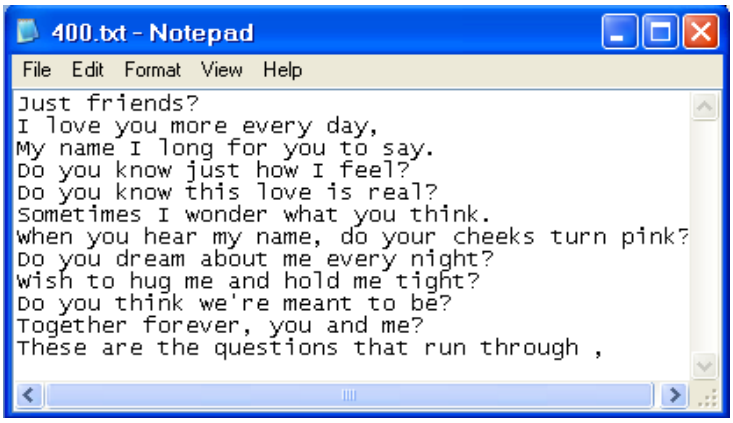


Hình 4.3. Giao diện giấu văn bản trong ảnh

Để thực hiện giấu văn bản ta phải nhập ảnh gốc, tệp văn bản, và chọn nơi lưu trữ ảnh giấu tin và thực hiện giấu tin.

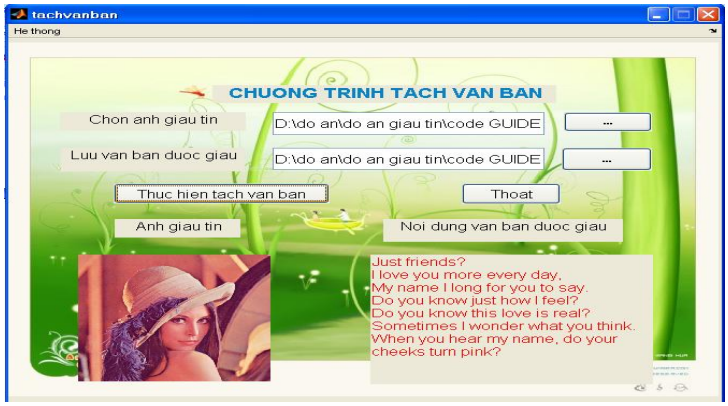


Hình 4.4 Giao diện chọn ảnh gốc



Hình 4.5. Tệp văn bản

- Giao diện tách văn bản



Hình 4.6. Giao diện tách văn bản trong ảnh

Để thực hiện tách văn bản ta chọn ảnh giấu tin và chọn nơi lưu trữ văn bản được tách ra từ ảnh giấu tin.

- Giao diện kiểm tra PSNR

Hình 4.7. Giao diện kiểm tra PSNR

4.5. Kết quả kiểm tra PSNR

Bảng 4.1. Kết quả PSNR khi tăng kích cỡ dữ liệu mật

Tên ảnh	**Kết quả PSNR**		
	Giấu 116 ký tự	Giấu 221 ký tự	Giấu 400 ký tự
Bamboo.bmp	73,2553	72,4826	71,1614
Bonghoa.bmp	86,6199	81,1664	76,9217
Banh.bmp	74,0954	70,5064	68,5626
Conmeo.bmp	77,6444	72,8321	68,8984
Hoa1.bmp	72,0574	69,3474	65,4288
Hoa2.bmp	76,1502	70,489	65,024
Lena.bmp	83,6837	78,1351	74,4447
banhoa.bmp	70,6465	70,0484	68,7673
vuonhoa.bmp	88,0806	83,5877	80,1678
Nen.bmp	76,4615	74,646	72,9206

Nhận xét : Bảng 4.1 cho thấy giá trị của các PSNR khác nhau sau khi giấu văn bản trong các ảnh. Số lượng ký tự càng ít thì PSNR càng cao chứng tỏ chất lượng dữ liệu được khôi phục càng tốt. Các kết quả cho thấy PSNR > 40 chứng tỏ chất lượng hình ảnh sau khi giấu tin là tốt.

KẾT LUẬN

Khóa luận đã thực hiện nhiệm vụ:

1. Trình bày tổng quan kỹ thuật giấu tin trong ảnh.
2. Nghiên cứu cấu trúc ảnh BMP, ảnh PNG.
3. Nghiên cứu một giải pháp giấu văn bản trong ảnh.
4. Xây dựng được chương trình giấu văn bản trong ảnh.

Trong báo cáo này em đã trình bày một kỹ thuật mới để ẩn văn bản bên trong hình ảnh. Mục tiêu chính là để đạt được an ninh chống lại các cuộc tấn công thống kê, trực quan và có thể giấu được nhiều thông tin. Bằng cách sử dụng số giả ngẫu nhiên nên đã tạo 1 lớp bảo mật cho kỹ thuật. Hình ảnh được sử dụng làm hình ảnh gốc là hình ảnh bitmap với lượng thông tin dư thừa là lớn và bằng cách sử dụng 7 bit/pixel để ẩn dữ liệu nên có thể giấu được nhiều thông tin hơn. Kết quả cho thấy kỹ thuật đã thành công trong việc đạt được những mục tiêu trên. Kết quả cho thấy giá trị của PSNR rất tốt có nghĩa là kỹ thuật cho thấy tính không thể nhận thấy tốt hơn.

Sau một thời gian tìm hiểu và nghiên cứu dưới sự hướng dẫn tận tình của cô giáo hướng dẫn Th.S Hồ Thị Hương Thơm em đã hoàn thành báo cáo và chương trình thử nghiệm. Tuy nhiên kỹ thuật giấu tin trong ảnh là một kỹ thuật mới mẻ, với thời gian thực hiện đề tài có hạn cộng với khả năng, kinh nghiệm còn hạn chế nên báo cáo của em còn gặp nhiều thiếu sót. Vì vậy em rất mong nhận được sự đóng góp ý kiến của các thầy giáo cô giáo để bài báo cáo tốt nghiệp của em được hoàn thiện hơn.

Em xin chân thành cảm ơn các thầy các cô!

TÀI LIỆU THAM KHẢO

[1 Luận văn thạc sĩ - Ngô Thái Hà "*Nghiên cứu kỹ thuật bảo vệ bản quyền các sản phẩm đồ họa vectơ"* Khoa Công nghệ thông tin trường Đại Học Thái Nguyên.

[2] Ingemar Cox, Jeffrey Bloom, Matthew Miller, Ton Kalker, Lessica Fridrich *"DigitalWatermarking anh Steganography"* - Morgan Kaufmann, 2008.

[3] Sukhpreet Kaur, Smeet Kaur *"A Novel Approach for Hiding Text Using Image Steganography"* – International Journal of computer science anh information security (IJSCIS), Vol.8, No.7, October 2010

[4] T. Morkel, J.H.P. Eloff and M.S. Olivier "*An overview of image steganography*" - in Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005), Sandton, South Africa, June/July 2005.

[5] Adnan Gutub, Ayed Al-Qahtani, Abdulaziz Tabakh "*Triple-A: secure RGB image steganography based on randomization*" AICCSA, IEEE/ACS International Conference on Computer Systems and Applications, pp. 400-403, 2009.

[6] A. Kaur, R. Dhir, G. Sikka," *A new image steganography based on first component alteration technique*", International Journal of Computer Science and Information Security (IJCSIS), vol. 6, pp. 53-56, 2009.

[7] M.T.Parvez , A. Gutub , "*RGB intensity based variable-bits image steganography*", APSCC 2008 –Proceedings of 3rd IEEE Asia-Pacific Services Computing Conference, Yilan, Taiwan, December 2008.

[8] Nameer N.EL-Emam, "*Hiding a large amount of data with high security using steganography algorithm*" Journal of Computer Science, vol.3 pp.223-232, 2007.

[9] Mohammed A.F. Al-Husainy, "*Image steganography by mapping pixels to letters*" Journal of Computer Science, vol. 5, pp. 33-38, 2009.

[10] A. Ibraheem Abdul-Sada, "*Hiding data using LSB-3*" J.Basrah Researches (Sciences), vol. 33, pp. 81-88, December, 2007.

[11] Website *http://vi.wikipedia.org*